Statistical Inference and Synthesis in the Image Domain for Mobile Robot Environment Modeling

Luz A. Torres-Méndez and Gregory Dudek
Center for Intelligent Machines, McGill University
Montreal, Quebec H3A 2A7, CA
Emails: {latorres,dudek}@cim.mcgill.ca

Abstract— We address the problem of computing dense range maps of indoor locations using only intensity images and partial depth. We allow a mobile robot to navigate the environment, take some pictures and few range data. Our method is based on interpolating the existing range data using statistical inferences learned from the available intensity image and from those (sparse) regions where both range and intensity information is present. The spatial relationships between the variations in intensity and range can be ef£ciently captured by the neighborhood system of a Markov Random Field (MRF). In contrast to classical approaches to depth recovery (i.e. stereo, shape from shading), we can afford to make only weak assumptions regarding speci£c surface geometries or surface re¤ectance functions since we compute the relationship between existing range data and the images we start with. Experimental results show the feasibility of our method.

I. INTRODUCTION

In robotics, the use of range data combined with visual information information, has become a key methodology for navigation and mapping. However, it is often hampered by the fact that range sensors that provide complete (2-1/2D) depth maps with a resolution akin to that of a camera, are prohibitively costly. Stereo cameras can produce volumetric scans that are economical, but they often require calibration or produce range maps that are either incomplete or of limited resolution. A particular common simplifying assumption is to represent 3D structure as a 2D "slice" through the world. However, in practice this is not suf£cient to capture structures of interest.

Surface depth recovery is one of the most standard vision problems, both because of its scienti£c and its pragmatic value. Many standard "shape-from" methods, however, are based on strong assumptions regarding scene structure or re¤ectance functions. While several elegant algorithms for depth recovery have been developed, the use of laser range data in many applications has become commonplace due to their simplicity and reliability (but not their elegance, cost or physical robustness).

In this paper, we extend our prior work [1] by presenting an improved algorithm for depth map estimation from a combination of color (or achromatic) intensity data and a limited amount of known depth data. We seek to reconstruct suitable 3D models from sparse range data sets while simultaneously facilitating the data acquisition process. It has been shown in [2] that although there are clear differences between optical and range images, they do have similar second-order statistics and scaling properties. Our motivation is to exploit this fact and also that both video imaging and *limited* range sensing are ubiquitous readily-available technologies while complete volume scanning remains prohibitive on most mobile platforms. It is important to highlight that we are not simply inferring a few missing pixels, but synthesizing a complete range map from as little as few laser scans across the environment.

Our methodology is to learn a statistical model of the (local) relationship between the observed range data and the variations in the intensity image and use this to compute the unknown range data. We approximate the *composite* of range and intensity at each point as a Markov process. Unknown range data is then inferred by using the statistics of the observed range data to determine the behavior of the Markov process. The presence of intensity where range data is being inferred is crucial since intensity data provides knowledge of surface smoothness and variations in depth. Our approach learns that knowledge directly from the observed data, without having to hypothesize constraints that might be inapplicable to a particular environment.

In the following section we consider relevant prior work. Section II describes our method to infer range data. Section III tests the proposed algorithm on different con£gurations of experimental data. Finally, in Section IV we give some conclusions and future directions.

A. Previous Work

We base our range estimation process on the assumption that the pixels constituting both the range and intensity images acquired in an environment, can be regarded as the results of pseudo-random processes, but that these random processes exhibit useful structure. In particular, we exploit the assumption that range and intensity images are correlated, albeit potentially complicated ways. Secondly, we assume that the variations of pixels in the range and intensity images are related to the values elsewhere in the image(s) and that these variations can be ef£ciently captured by the neighborhood system of a Markov Random Field. Both these assumptions have been considered before [3]–[7], but they have never been exploited in tandem.

Digital inpainting [8]–[10] is quite similar to our problem, although our domain and approach are quite different. Baker and Kanade [11] used a learned representation of pixel variation for perform resolution enhancement of face images. The processes employed to interpolate new high-resolution pixel data is quite similar in spirit to what we

describe here, although the application and technical details differ significantly. The work by Freeman [12], [13] on learning the relationships between intrinsic images is also related.

In prior work [1], [14], we performed reconstruction by inferring depth values using predetermined schedule over space, essentially walking a spiral from the boundary of a region towards the center. We have observed that reconstruction across depth discontinuities is often problematic as there is comparatively little constraint for probabilistic inference at these locations. Further, such locations are often identified with edges in both the range and intensity maps. Based on these observations, we have developed two important algorithmic refinements that have a dramatic effect of the quality of the results. In the present work, we modify the reconstruction sequence to first recover the values of those augmented voxels for which we can make the most reliable inferences. This leads to two critical algorithmic refinements: (1) as we recover augmented pixels we defer the reconstruction of augmented voxels close to intensity or depth discontinuities as much as possible, and (2) as we reconstruct we select those voxels for reconstruction that have the largest degree of boundary constraint (aside from those deferred by condition (1)).

The inference of 3D models of a scene is a problem that subsumes a large part of computer vision research over the last 30 years. In the context of this paper we will consider only a few representative solutions.

In general, the process of building a full 3D model of a real environment can be divided onto two processes: acquisition of measurements in 3D and synthesis of useful geometric models from measurements. In some cases, for example when models are generated manually, these steps may be combined. In other cases the processes of collecting sets of 3D points (often referred to as range scans), combining them onto surfaces and then generating suitable models for graphics applications entail distinct computations. In this paper we focus only on the processes of obtaining 3D data.

Over the last decade laser rangefinders have become affordable and available but their application to building full 3D environment models, even from a single viewpoint, remains costly or difficult in practice. In particular, while laser line scanners based on either triangulation and/or time-of-flight are ubiquitous, full volume scanners tend to be much more complicated and error-prone. As a result, the acquisition of *dense, complete* 3D range maps is still a pragmatic challenge even if the availability of laser range scanners is presupposed.

Most of prior work on synthesis of 3D environment models uses one of either photometric data or geometric data [15]–[18] to reconstruct a 3D model of an scene. For example, Fitzgibbon and Zisserman [17] proposed a method that sequentially retrieves the projective calibration of a complete image sequence based on tracking corner and/or line features over two or more images, and reconstructs each feature independently in 3D. Their method solves the feature correspondence problem based on the fundamental matrix and tri-focal tensor, which encode precisely the geometric constraints available from two or more images of the same scene from different viewpoints. Related work includes that of Pollefeys et. al. [18]; they obtain a 3D model of an scene from image sequences acquired from a freely moving camera. The camera motion and its settings are unknown and there is no prior knowledge about the scene. Their method is based on a combination of the projective reconstruction, self calibration and dense depth estimation techniques. In general, these methods derive the epipolar geometry and the trifocal tensor from point correspondences. However, they assume that it is possible to run an interest operator such as a corner detector to extract from one of the images a sufficiently large number of points that can then be reliably matched in the other images.

Shape-from-shading is related in spirit to what we are doing, but is based on a rather different set of assumptions and methodologies. Such method [19], [20] reconstruct a 3D scene by inferring depth from a 2D image; in general, this task is difficult, requiring strong assumptions regarding surface smoothness and surface reflectance properties.

Recent work has focus on combining information from the intensity and range data for 3d model reconstruction. Several authors [21]–[25] have obtained promising results. Pulli et al. [21] address the problem of surface reconstruction by measuring both color and geometry of real objects and displaying realistic images of objects from arbitrary viewpoints. They use a stereo camera system with active lighting to obtain range and intensity images as visible from one point of view. The integration of the range data into a surface model is done by using a robust hierarchical space carving method. The integration of intensity data with range data has been proposed [23] to help define the boundaries of surfaces extracted from the 3D data, and then a set of heuristics are used to decide what surfaces should be joined. For this application, it becomes necessary to develop algorithms that can hypothesize the existence of surface continuity and intersections among surfaces, and the formation of composite features from the surfaces.

However, one of the main issues in using the above configurations is that the acquisition process is very expensive because dense and complete intensity and range data are needed in order to obtain a good 3D model. As far as we know, there is no method that bases its reconstruction process on having a small amount of intensity and/or range data and synthetically estimating the areas of missing information by using the current available data. In particular, such a method is feasible in man-made environments, which, in general, have inherent geometric constraints, such as planar surfaces.

II. METHODOLOGY

To restate our objective, we wish to infer a dense range map from an intensity image and a limited amount of initial range data. At the outset, we assume that resolution of the intensity and range data is the same and that they are already registered (in practice this registration

could be computed as a £rst step, but we omit this in the current presentation.) Note that while the process of inferring distances from intensity super£cially resembles shape-from-shading, we do not depend on prior knowledge of re¤ectance or on surface smoothness or even on surface integrability (which is a technical precondition for most shape-from-shading methods, even where not explicity stated).

We solve the range data inference problem as an extrapolation problem by approximating the *composite* of range and intensity at each point as a Markov process. Unknown range data is then inferred by using the statistics of the observed range data to determine the behavior of the Markov process. Critical to the processes is the presence of intensity data at each point where range is being inferred. Intuitively, this intensity data provides at least two kinds of information: (1) knowledge of when the surface is smooth, and (2) knowledge of when there is a high probability of a variation in depth. Our approach learns that information from the observed data, without having to fabricate or hypothesize constraints that might be inapplicable to a particular environment.

A. Algorithmic re£nements

In our algorithm we synthesize one depth value $R(x, y)$ at a time. From previous experiments we know that reconstruction sequence (the order in we choose the next depth value to synthesize) highly in¤uences the quality of the £nal result. One of the problem with the spiral-scan ordering (also known as onion-peel ordering) was the strong dependence from the previous assigned voxel. In this work, our reconstruction sequence is to £rst recover the values of those augmented voxels for which we can make the most reliable inferences, so that as we reconstruct we select those voxels for reconstruction that have the largest degree of boundary constraint.

We have also observed that reconstruction across depth discontinuities is often problematic as there is comparatively little constraint for probabilistic inference at these locations. Further, such locations are often identi£ed with edges in both the range and intensity maps. In this work we have incorporated edge information and, as we recover augmented voxels, we defer the reconstruction of augmented voxels close to intensity or depth discontinuities as much as possible. We use the Canny edge detector [26] for extracting the edges from the intensity images.

B. The MRF model for range synthesis

Markov Random Fields (MRF) are used here as a model to synthesize range. We focus on our development of a set of **augmented voxels** **V** that contain intensity (either from grayscale or color images), edge (from the intensity image) and range information (where the range is initially unknown for some of them). Thus, $\mathbf{V} = (\mathrm{I}, \mathrm{E}, \mathrm{R})$, where I is the matrix of known pixel intensities, E is a binary matrix (1 if an edge exists and 0 otherwise) and R denotes the matrix of incomplete pixel depths. We are interested only in a set of such augmented voxels

such that one augmented voxel lies on each ray that intersects each pixel of the input image I, thus giving us a registered range image R and intensity image I. Let $Z_m = (x, y) : 1 \leq x, y \leq m$ denote the m integer lattice (over which the images are described); then $\mathrm{I} = \{I_{x,y}\}$, $(x, y) \in Z_m$, denotes the gray levels of the input image, and $\mathrm{R} = \{R_{x,y}\}$, $(x, y) \in Z_m$ denotes the depth values. We model **V** as an MRF. Thus, we regard I and R as a random variables. For example, $\{\mathrm{R} = r\}$ stands for $\{R_{x,y} = r_{x,y}, (x, y) \in Z_m\}$. Given a *neighborhood system* $\mathcal{N} = \{\mathcal{N}_{x,y} \subset Z_m\}$, where $\mathcal{N}_{x,y}$ denotes the neighbors of (x, y), such that, (1) $(x, y) \notin \mathcal{N}_{x,y}$, and (2) $(x, y) \in \mathcal{N}_{k,l} \Longleftrightarrow (k, l) \in \mathcal{N}_{x,y}$. An MRF over $(Z_m, \mathcal{N})$ is a stochastic process indexed by Z_m for which, for every (x, y) and every $v = (i, r)$ (i.e. each augmented voxel depends only on its immediate neighbors),

$$P(V_{x,y} = v_{x,y} \mid V_{k,l} = v_{k,l}, (k, l) \neq (x, y))$$
$$= P(V_{x,y} = v_{x,y} \mid V_{k,l} = v_{k,l}, (k, l) \in \mathcal{N}_{x,y}), \quad (1)$$

The choice of $\mathcal{N}$ together with the conditional probability distribution of $P(\mathrm{I} = i)$ and $P(\mathrm{R} = r)$, provides a powerful mechanism for modeling spatial continuity and other scene features. On one hand, we choose to model a neighborhood $\mathcal{N}_{x,y}$ as a square mask of size $n \times n$ centered at the augmented voxel location (x, y). This neighborhood is causal, meaning that only those augmented voxels already containing information (either intensity, range or both) are considered for the synthesis process. On the other hand, calculating the conditional probabilities in an explicit form is an infeasible task since we cannot ef£ciently represent or determine all the possible combinations between augmented voxels with its associated neighborhoods. Therefore, we avoid the usual computational expense of sampling from a probability distribution (Gibbs sampling, for example), and synthesize a depth value from the augmented voxel $V_{x,y}$ with neighborhood $\mathcal{N}_{x,y}$, by selecting the range value from the augmented voxel whose neighborhood $\mathcal{N}_{k,l}$ most resembles the region being £lled in, i.e.,

$$\mathcal{N}_{best} = \underset{(k,l) \in \mathcal{A}}{\mathbf{argmin}} \parallel \mathcal{N}_{x,y} - \mathcal{N}_{k,l} \parallel, \quad (2)$$

where $\mathcal{A} = \{\mathcal{A}_{k,l} \subset \mathcal{N}\}$ is the set of local neighborhoods, in which the center voxel has already assigned a depth value, such that $1 \leq \sqrt{(k - x)^2 + (l - y)^2} \leq d$. For each successive augmented voxel this approximates the maximum a posteriori estimate; $R(k, l)$ is then used to specify $R(x, y)$. The similarity measure $\parallel . \parallel$ between two generic neighborhoods $\mathcal{N}_a$ and $\mathcal{N}_b$ is de£ned as the weighted sum of squared differences (WSSD) over the partial data in the two neighborhoods. The "weighted" part refers to applying a 2-D Gaussian kernel to each neighborhood, such that those voxels near the center are given more weight than those at the edge of the window.

C. Range Synthesis Ordering

We based our reconstruction sequence on the amount of reliable information surrounding the augmented voxel

whose depth value is to be estimated, and also on the edge information. Let V_p be an augmented voxel with unknown range and $\mathcal{N}_p$ be a 3×3 square window centered at V_p (i.e. we are considering just the 8-closest neighbors). Then, for each augmented voxel V_p, we count the number of neighbor voxels with already assigned range and intensity. We start by synthesizing those augmented voxels with the maximum number of £lled neighbors, leaving to the end those with an edge passing through them. After a depth value is estimated, we update each of its neighbors by adding 1 to their own neighbor counters. We then proceed to the next group of augmented voxels to synthesize until no more augmented voxels exist.

III. Experimental Results

In this section we show experimental results conducted on data acquired in a real-world environment. We use ground truth data from two widely available databases. The £rst database provides real intensity (re¤ectance) and range images of indoor scenes acquired by an Odetics laser range £nder mounted on a mobile platform. The second database [27] provides color images with complex geometry and pixel-accurate ground-truth disparity data. We also show preliminary results on data collected by our mobile robot, which has a video camera and a laser range £nder mounted on it. We start with the complete range data set as ground truth and then hold back most of the data to simulate the sparse sample of a real scanner and to provide input to our algorithm. This allows us to compare the quality of our reconstruction with what is actually in the scene. In the following, we will consider several strategies for subsampling the range data.
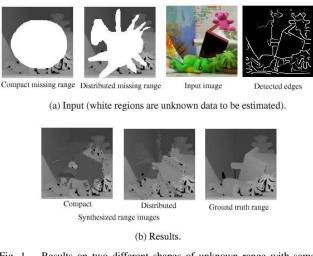

Compact missing range   Distributed missing range   Input image   Detected edges

(a) Input (white regions are unknown data to be estimated).

Compact   Distributed   Ground truth range

Synthesized range images

(b) Results.

Fig. 1. Results on two different shapes of unknown range with same area 6800 pixels.

A. Arbitrary shape of unknown range data

The £rst type of experiment involves the range synthesis when the unknown range data is of arbitrary shape. In particular, we show how the shape that contains the unknown range in¤uences their estimation. In Fig. 1a, two input range images (the left and middle images), and

input intensity, with its corresponding edge information, are given. The number of pixels in each unknown area (shown in white) of both range images is 6800. The perimeters however are different. Fig. 1b shows the synthesized range images (left and middle) and the ground truth range image for comparison purposes. It can be seen that when synthesizing big areas of unknown range data, our algorithm performs better if the area is not compact, since combinations of already known range and intensity give more information about the geometry of the scene.

B. Range measurements with variable width along the $x-$ and $y-$ axis

This type of experiment involves the range synthesis when the initial range data is a set of stripes with variable width along the $x-$ and $y-$axis of the intensity image. In the following cases, we tested our algorithm with the same intensity image in order to compare the results. Figure 2 shows the input intensity image (left) of size 128×128 and for purpose of comparison we show the ground truth range image (right) from where we hold back the data to simulate the samples.

Four cases of subsampling are shown in Figure 3. The initial range data, shown in the left column, goes from dense to very sparse. The percentage of missing range data is indicated below each image. For the £rst three cases the size of the neighborhood is set to be 5×5 pixels and for the last case 3×3. The right column shows the synthesized range data obtained after running our algorithm.

The £rst two cases have the same amount of missing range, however the synthesized range for the second case is much better. Intuitively, this is because the sample spans a broader distribution of range-intensity combinations. The Odetics LRF uses perspective projection, thus the range image coordinate system is spherical. The absolute value of each error is taken and the mean of those values is computed to arrive at the mean absolute residual (MAR) error. To calculate the absolute residual errors, we £rst convert the range images to the Cartesian coordinate system (range units) by using the equations in [28] and then we convert the range units to centimeters.
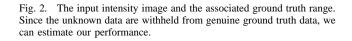
Table I shows the MAR errors (calculated only on the unknown areas) of the examples shown in Figure 3. The approximated size of the input scene is 550 centimeters. For each case, we show in Figure 4, the histogram of the pixels based on the absolute residual errors. Each class in the histogram covers a range of 3.66 centimeters. We do this because the MAR error does not accurately represents the performance of our algorithm in cases where there are very few pixels (it may be only one) with high absolute residual error. From the histograms we can see that (except for the £rst case) there is a high concentration of pixels with residual errors ≤ 10.98 centimeters.
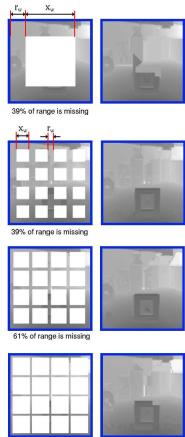
In general, the results are good in all cases, except for the £rst. Our algorithm was capable of recovering the whole range of the image. It is important to note that we do not assume that the range and intensity images are correlated (i.e. dark regions tend to be further). In the

Input intensity   Ground truth range

Fig. 2.   The input intensity image and the associated ground truth range. Since the unknown data are withheld from genuine ground truth data, we can estimate our performance.

39% of range is missing

39% of range is missing

61% of range is missing

76.5% of range is missing

Fig. 3.   Results on real data. The left column shows the initial range data and to their right is the synthesized result (the white squares represent unknown data to be estimated).

% of area with missing range	MAR Error (in centimeters)
39	36.36
39	5.76
61	8.86
76.5	9.99

TABLE I

MAR ERRORS FOR THE CASES SHOWN IN FIGURE 3.

previous example, the correlation coef£cient is 0.64. We will show examples where this coef£cient is low and still good results are obtained.
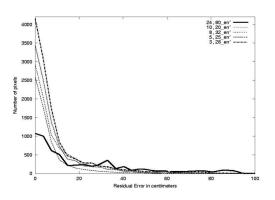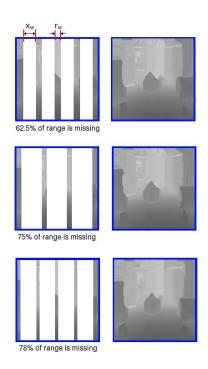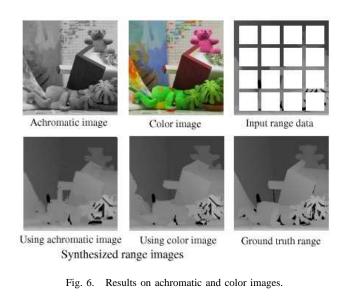

Fig. 4.   Histograms of pixels based on the absolute residual errors for the cases shown in Fig. 3. Note that the concentration of pixels is with residual errors between ≤ 10.98 cms.

We conducted experiments on 32 images of common scenes found in a general indoor man-made environment. The smoothing parameter for edge detection was set to 0.8 in all examples. Due to space limitations, we are only showing 4 more examples in Figure 5. The MAR errors from top to bottom are shown in Table II. The approximated size of each scene and the correlation coef£cient are also given. We normalize the MAR error by dividing it by the scene size. This normalized MAR measure is a better indication of how large the error is according to the scene size. Computation time for the results using non-optimized code, is on the order of minutes on generic PC's.

Fig. 5.   Examples on real data. The £rst two columns show the input intensity and range images, respectively. For all cases, 61% of the range is unknown (shown in white). The third column shows the edges detected in the input intensity. The last two columns show the synthesized results and the ground truth range for visual comparison.

C. Using color images

We now show how color information may improve the range synthesis. Figure 6 displays in the £rst row, the grayscale and color images of the same scene, and to their right the input range data. The synthesized results after running our algorithm is shown below together with the ground truth data for comparison purposes.

MAR Error (in cms)	Scene size size (in cms)	MAR/Size	Correlation coefficient
8.58	600	0.017	0.47
13.48	800	0.021	0.63
11.39	500	0.024	0.32
7.12	400	0.048	0.62

TABLE II

MAR ERRORS OF THE CASES SHOWN IN FIG. 5.

It can be seen that there are some regions where color information may help in the synthesis process. For example, the chimney in the center of the image is separated from the background since they have different colors. This is hardly notice in the grayscale image.

Fig. 6. Results on achromatic and color images.

D. Range measurements with variable width along the $x-$axis.

We now show experimental results where the initial range data is a set of stripes only along the $x-$axis. This type of experiment is interesting since it resembles what is obtained by sweeping a one-dimensional LIDAR sensor. We have selected the same intensity image shown at the top of Figure 5 in order to compare the results. Figure 7 displays this input intensity image(left) and the ground truth range image from where we hold back the data to simulate the samples. The edge information used is shown at the top of Figure 5.

Figure 8 shows three experiments. The initial range images are shown in the left column. The percentage of missing range is indicated below each of them. The right column shows the synthesized range data obtained after running our algorithm.

The MAR errors for the experiments are 20.72, 18.98 and 20.23, respectively. The approximated size of the scene is 600 centimeters. In general, our algorithm was able to capture the underlying structure of the scene. However, it can be seen that the reconstruction was not good in regions containing surfaces sloping away (for example walls and

Fig. 7. The input intensity image and the associated ground truth range. Since the unknown data are withheld from genuine ground truth data, we can estimate our performance.

Fig. 8. Results on real data. The left column shows the initial range data and to their right is the synthesized result (the white squares represent unknown data to be estimated).

¤oor). This is due to the fact that the very limited amount of input range does not cover much of the changes in depth, and our algorithm fails by assigning already identical depth values instead of different depths at each point.

Therefore, the initial range data given as an input is crucial to the quality of the synthesis, that is, if no interesting changes exist in the range and intensity, then the task becomes dif£cult. One solution to this problem is to use the surface normals to generate new depth values instead of using range values directly. We are currently working on this problem. However, the results presented here demonstrate that this is a viable option to facilitate environment modeling.

We now show some preliminary results on data collected in our own building. We use a mobile robot with a video camera and a laser range £nder mounted on it, to navigate the environment. For our application, the laser range £nder was set to scan a 180 degrees £eld of view horizontally and 90 degrees vertically. Figure 9 shows a picture of

our mobile robot. As it was mentioned previously, the input intensity and available range data needs to be already registered. Range and intensity are different type of data, their sampling resolution are not the same. We achieved the registration of the intensity and range data in a semi-automatic way, by using crosscorrelation on the video frames and then manually selecting those corresponding regions from the range image. Details about this registration step is not in the scope of this paper. We are currently seeking to have a fully automatic way of registering both type of data.

Fig. 9.   Our mobile robot used to acquired the data.

Figure 10 shows one experimental result for a case where the input range data is a set of stripes along the x- and y-axis. The input intensity and the ground truth range data (for comparison purposes) are shown on the £rst row. The second row displays the input range image (left) with 62% of unknown range and the synthesized range data (right) after running our algorithm. It can be seen that our algorithm was capable of recovering the whole depth map of the scene.

IV. Summary and Conclusions

In this paper we have presented an approach to depth recovery that allows good quality scene reconstruction to be achieved in real environments. The method requires both an intensity images and set of partial range measurements input. In fact, the input range measurements are most effective if they are provided in the form of clusters of measurements scattered over the image. This form of sampling is best since it allows local statistics to be computed, but also provides boundary conditions at various locations in the image. While clumps per set are not available from most laser range scanners, swaths of data can, in fact, be readily and ef£ciently extracted using laser scanners.

When we use color images in the reconstruction process, it appears that the £delity of the reconstruction is somewhat improved over achromatic images. This appears to be due to the fact that the color data provides tighter constraint over where and how the interpolation process should be applied. At the same time, the higher dimensionality of the Markov Random Field model for color images may make the reconstruction problem more dif£cult in some cases. Although we have not observed it in our test cases, we expect reconstructions to fail in some cases with color

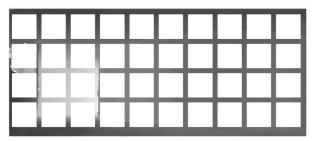data whereas they might succeed with achromatic input images.

References

[1] L. Torres-Méndez and G. Dudek, "Range synthesis for 3d environment modeling," in *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*. Las Vegas, NV: IEEE Press, October 2003, p. 8.

[2] A. Lee, K. Pedersen, and D. Mumford, "The complex statistics of high-contrast patches in natural images," 2001, private correspondence. [Online]. Available: citeseer.nj.nec.com/lee01complex.html

[3] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.

[4] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *ICCV (2)*, September 1999, pp. 1033–1038. [Online]. Available: citeseer.nj.nec.com/efros99texture.html

[5] L. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," in *SIGGRAPH*, July 2000, pp. 479–488.

[6] A. Efros and W. Freeman, "Image quilting for texture synthesis and transfer," in *SIGGRAPH*, August 2001, pp. 1033–1038.

[7] A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, and D. Salesin, "Images analogies," in *SIGGRAPH*, August 2001.

[8] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. ACM Conf. Computer Graphics (SIGGRAPH)*, July 2000, pp. 417–424.

[9] G. S. M. Bertalmio, L. Vese and S. Osher, "Simultaneous structure and texture image inpainting," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2003.

[10] P. P. A. Criminisi and K. Toyama, "Object removal by examplar-based inpainting," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2003.

[11] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[12] A. Torralba and W. Freeman, "Properties and applications of shape recipes," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2003.

[13] E. P. W.T. Freeman and O. Carmichael, "Shape recipes: scene representations that refer to the image," *Vision Sciences Society Annual Meeting*, pp. 25–47, 2003.

[14] L. Torres-Méndez and G. Dudek, "Range synthesis for 3d environment modeling," in *IEEE Workshop on Applications of Computer Vision*, 2002, pp. 231–236.

[15] P. Debevec, C. Taylor, and J. Malik, "Modeling and rendering architecture from photographs: A hybrid geometry and image-based approach," in *SIGGRAPH*, 1996, pp. 11–20.

[16] A. Hilton, "Reliable surface reconstruction from multiple range images," in *ECCV*, 1996.

[17] A. Fitzgibbon and A. Zisserman, "Automatic 3d model acquisition and generation of new images from video sequences," in *Proceedings of European Signal Processing Conference*, 1998, pp. 1261–1269.

Input intensity image

Ground truth range

Initial range data. r_w=10, x_w=y_w=50.

Synthesized range image

Fig. 10. Results on real data collected from our mobile robot.

[18] M. Pollefeys, R. Koch, M. Vergauwen, and L. V. Gool, "Automated reconstruction of 3d scenes from sequences of images," *ISPRS Journal Of Photogrammetry And Remote Sensing*, vol. 55, no. 4, pp. 251–267, 2000.

[19] B. Horn and M. Brooks, *Shape from Shading*. MIT Press, Cambridge Mass., 1989.

[20] J. Oliensis, "Uniqueness in shape from shading," *Int. Journal of Computer Vision*, vol. 6, no. 2, pp. 75–104, 1991.

[21] K. Pulli, M. Cohen, T. Duchamp, H. Hoppe, J. McDonald, L. Shapiro, and W. Stuetzle, "Surface modeling and display from range and color data," *Lecture Notes in Computer Science 1310*, pp. 385–397, September 1997.

[22] S. El-Hakim, "A multi-sensor approach to creating accurate virtual environments," *Journal of Photogrammetry and Remote Sensing*, vol. 53, no. 6, pp. 379–391, December 1998.

[23] V. Sequeira, K. Ng, E. Wolfart, J. Goncalves, and D. Hogg, "Automated reconstruction of 3d models from real environments," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 54, pp. 1–22, February 1999.

[24] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk, "The digital michelangelo project: 3d scanning of large statues," in *SIGGRAPH*, July 2000.

[25] I. Stamos and P. Allen, "3d model construction using range and image data," in *CVPR*, June 2000.

[26] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[27] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2003, pp. 195–202.

[28] K. Storjohann, "Laser range camera modeling," Oak Ridge National Laboratory, Oak Ridge, Tennessee, Tech. Rep., 1990.