A Statistical Learning Method for Mobile Robot Environment Modeling

Luz A. Torres-Méndez and Gregory Dudek
Center for Intelligent Machines, McGill University
Montreal, QC, Canada
{latorres,dudek}@cim.mcgill.ca

Abstract

This paper presents a statistical learning method for computing range data as an initial solution to the environment modeling problem in the context of mobile robotics. Unlike other methods that are based on a set of geometric primitives, our method computes dense range maps of locations in the environment using only intensity images and very limited amount of range data as an input. This is achieved by exploiting the following assumptions: 1) the observed range and intensity images are correlated and, 2) variations of pixels in the range and intensity images are related to the values elsewhere in the image(s). These variations can be efficiently captured by the neighborhood system of a Markov Random Field (MRF). Experimental results show the feasibility of our method.

1 Introduction

Knowledge of its surrounding environment is crucial for a mobile robot to accomplish even the simplest task. The use of range data for navigation and mapping has become a key methodology, but the acquisition of complete range maps (i.e. volume scans) from a single view point remains prohibitive for many real systems (such as range scans from a single viewpoint are sometimes known as two-and-a-half dimensional representations). Stereo cameras can produce volumetric scans that are economical, but they often require calibration or produce range maps that are either sparse or of limited resolution. Volumetric laser scanners tend to be costly and physically demanding or slow. In robotics, a particular common simplifying assumption is to represent 3D structure as a 2D "slice" through the world. However, in practice this is not sufficient to capture structures of interest. In this paper, we bypass this assumption and propose a novel method to estimate 3D data from a combination of a single video intensity image and a limited amount of observed range data.

By using a limited amount of range data, the complexity of the data acquisition system is reduced, while still being able to make estimates over a full range map (from a single viewpoint at a time). This should allow a robotic system to rapidly collect a small amount of range data and a video image, and then infer the rest of the range map it does not capture directly. 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 line-striping scans across the environment. This paper examines our ability to extrapolate range data given initial data in various configurations. For this study we do not use data collected from our own laboratory, but rather have elected to use widely available ground-truth data from Oak Ridge National Labs. As such, while our target application is mobile robotics, this paper does not explicitly address the issues of navigation and data acquisition.

Our methodology is to statistically learn the relationship between the observed range data and the variations in the intensity image and use this to compute the unknown range data. This can be regarded as a form of shape-from-shading based on statistical learning, although traditional shape-from-shading is quite different from this approach in its technical details. In our approach, 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 from the observed data, without having to hypothesize constraints that might be inapplicable to a particular environment.

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 in 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 efficiently captured by the neighborhood system of a Markov Random Field. Both these assumptions have been considered before [Geman and Geman, 1984; Efros and Leung, 1999; Wei and Levoy, 2000; Efros and Freeman, 2001; Hertzmann *et al.*, 2001], but they have never been exploited in tandem.

This paper is structured as follows. Section 2 briefly consider some of the related prior work. Section 3 describes our

method to infer range data. Section 4 tests the proposed algorithm on several types of experimental data. Finally, in Section 5 we give some conclusions and future directions.

2 Previous work

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

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 physically sensitive. 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.

Much of the previous work on environment modeling uses one of either photometric data or geometric data [Debevec *et al.*, 1996; Hilton, 1996; Fitzgibbon and Zisserman, 1998; Pollefeys *et al.*, 2000] to reconstruct a 3D model of an scene. For example, Fitzgibbon and Zisserman [Fitzgibbon and Zisserman, 1998] 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 trifocal 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. [Pollefeys *et al.*, 2000]; 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 [Horn and Brooks, 1989; Oliensis, 1991] 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 considered the use of both intensity data as well as range measurements. Several authors [Pulli *et al.*, 1997; El-Hakim, 1998; Sequeira *et al.*, 1999; Levoy *et al.*, 2000; Stamos and Allen, 2000] have obtained promising results. Pulli et al. [Pulli *et al.*, 1997] 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 [Sequeira *et al.*, 1999] 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 range data and synthetically estimating the areas of missing range 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.

3 Methodology

As noted above our objective is to compute range values where only intensity is known. In the current presentation, we assume that the resolution of the intensity and range data is the same and that they are already registered.

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 to kinds of information: knowledge of when the surface is smooth, and 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.

3.1 Markov Random Fields (MRF) for range synthesis

We focus on our development of a set of **augmented voxels $\mathbf{V}$** that contain intensity and range information (where the range is initially unknown for some of them). Thus, $\mathbf{V} = (\mathrm{I}, \mathrm{R})$, where I is the matrix of known pixel intensities 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 $m \times m$ 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 $\mathbf{V}$ 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* $\mathcal{N} = \{\mathcal{N}_{x,y} \in Z_m\}$, where $\mathcal{N}_{x,y} \subset Z_m$ 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),

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

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 both, intensity and range information 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 efficiently 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 $R_{x,y}$ deterministically by selecting the range value $R_{k,l}$ from the augmented voxel whose neighborhood most resembles the region being filled in, i.e.,

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

where $\mathcal{A} = \{\mathcal{A}_{k,l} \subset \mathcal{N}\}$ is the set of local neighborhoods, 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 $\parallel \, . \, \parallel$ is described over the partial data about locations (x,y) and (k,l) and is calculated as follows,

$$\sum_{\vec{v} \in \mathcal{N}^*} G(\sigma, \vec{v} - \vec{v}_0)[(I_{\vec{v}} - I'_{\vec{v}})^2 + (R_{\vec{v}} - R'_{\vec{v}})^2], \quad (3)$$

where $\vec{v}_0$ is the augmented voxel located at the center of the neighborhood $\mathcal{N}^*$, $\vec{v}$ is a neighboring voxel of $\vec{v}_0$. I and R are the intensity and range values of the neighboring augmented voxels of the depth value $R_{x,y} \in \vec{v}_0$ to synthesize, and I' and R' are the intensity and range values to be compared with and in which, the center voxel $\vec{v}_0$ has already assigned a depth value. G is a 2-D Gaussian kernel that gives more weight to those voxels near the center than those at the edge of the window.

In our algorithm we synthesize one depth value $R(x,y)$ at a time. The order in which we choose the next depth value to synthesize will influence the final result. In our first experiments, depth values are assigned in a spiral-scan ordering, either growing inwards or outwards, depending on the shape of the area to synthesize.

4 Experimental Results

Experiments were conducted on data acquired in a real-world environment. The real intensity (reflectance) and range images of indoor scenes were acquired by an Odetics laser range finder mounted on a mobile platform. Images are 128×128 pixels and encompass a $60^o \times 60^o$ field of view. 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.

4.1 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 1 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.

Input intensity Ground truth range

Figure 1: 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.

Five cases of subsampling are shown in Figure 2. The initial range data, shown in the left column, goes from dense to very sparse. The width of the input stripes r_w and the width of the area with missing range data x_w are indicated below each image. For the first 4 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 first 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 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 first convert the range images to the Cartesian coordinate system (range units) by using the equations in [Storjohann, 1990] and then we convert the range units to centimeters. Table 1 shows the MAR

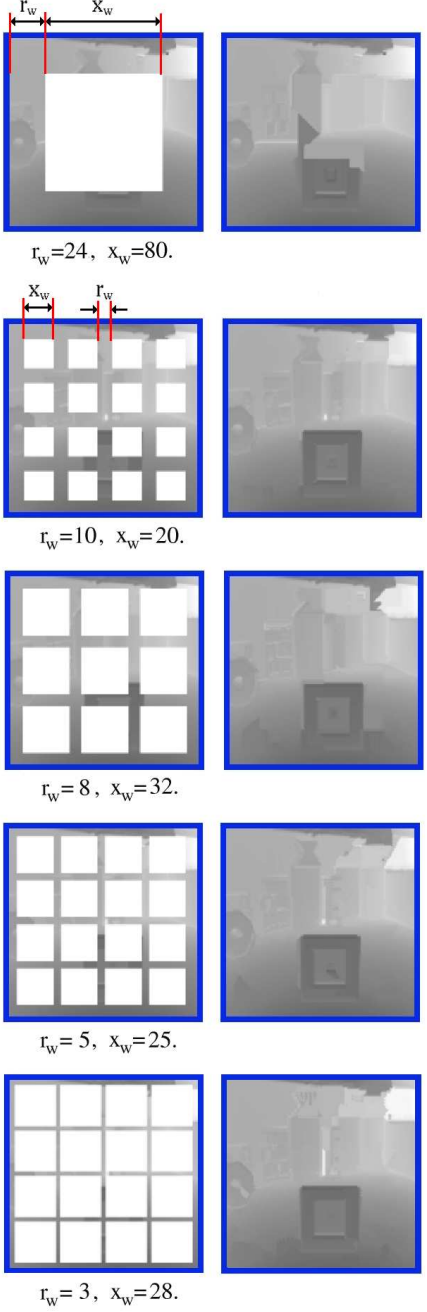


Figure 2: Results on real data. The left column shows the initial range data and to their right is the synthesized result (the white squares represent unknown data to be estimated).

errors (calculated only on the unknown areas) of the examples shown in Figure 2. The approximated depth size of the input scene is 550 centimeters.

For each case, we show the histogram of the pixels based on the absolute residual errors in Figure 3. 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 first case) there is a high concentration of pixels with residual errors ≤ 10.98 centimeters.

Input range s_w	Input range x_w	% of area with missing range	MAR Error (in centimeters)
24	80	39	36.36
10	20	39	5.76
8	32	56	12.07
5	25	61	8.86
3	28	76	9.99

Table 1: MAR errors for the cases shown in Figure 2.

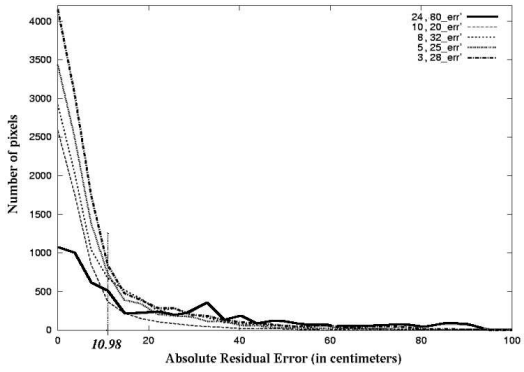


Figure 3: Histograms of pixels based on the absolute residual errors for the cases shown in Fig. 2. Note that the concentration of pixels is with residual errors between ≤ 10.98 cms.

In general, the results are surprinsingly good in all cases, except for the first. Our algorithm was capable of recovering the whole range of the image. For the case with $r_w = 5$ and $x_w = 25$, Figure 4 displays two different views using the synthesized range and ground truth for comparison purposes.

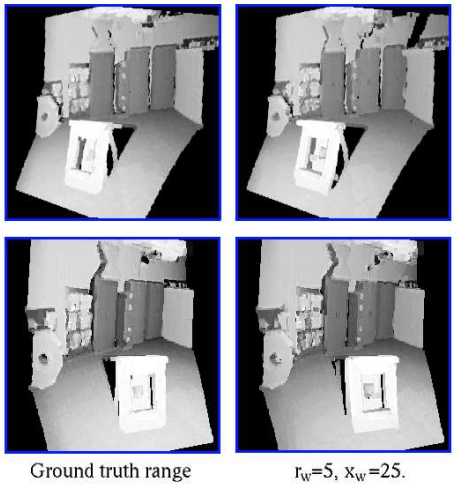


Figure 4: Results in 3D. Two views of the real range (left column) and the synthesized results (right column) of case where $r_w = 5$ and $x_w = 25$.

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

From the previous experiments, the case of subsampling with $r_w = 5$ and $x_w = 25$, is of our interest. The input range measurements are very sparse and the obtained results were very satisfactory. We conducted experiments on 30 images of common scenes found in a general indoor man-made environment using this case of subsampling. 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 2. The approximated depth size of each scene and the correlation coefficient are also given. We can normalize the MAR error by dividing the average residual error by the depth size of the scene.

Figure 5: Examples on real data for the case of subsampling with $r_w = 5$ and $x_w = 25$. The first and second columns are the input intensity and range data, respectively. White regions in the input data are unknown data to be inferred by the algorithm. The synthesized results are shown in the third column and, the real range images are displayed in the last column for visual comparison.

It can be seen that the synthesized range images are very similar to the real range images. However, it can be noted also (especially in the last example) that our algorithm performs poorly when high variations in depth are not captured by the intensity-range combinations. These variations can be captured by incorporating edge information to the Markov Random Field model. In the next section, we explain in more detail how we solve this problem.

MAR Error (in cms)	Approx. depth size (in cms)	MAR/Depth	Correlation coefficient
10.40	600	0.017	0.47
16.58	800	0.021	0.63
12.16	500	0.024	0.32
19.17	400	0.048	0.62

Table 2: The input information and MAR errors of the cases shown in Figure 5.

4.2 Using Edge Information

From the experimental results previously shown, we noted that our algorithm sometimes is not effective for inferring depth near object edge, where high discontinuities exist. Edge detection from intensity images is useful to solve this problem. Strong edges from the object or grey levels boundaries act to prevent the range estimation beyond that point. Thus, our algorithm synthesizes first depth in voxels located *"inside"* of a detected edge. The estimation of those points having an edge are estimated at the end, when all other points with unknown range are already synthesized.

We use the Canny edge detector [Canny, 1986] for extracting the edges from the intensity images. The incorporation of edge information to our Markov Random Field model is very simple, we just add that information such that each augmented voxel now contains intensity, range (if known), and edge information (1 is there exist an edge, 0 otherwise).

We also noticed that the order in which we choose the next depth value to synthesize will reflect the final result. With the spiral-scan ordering there is a strong dependence from the previous assigned voxel. A better scan ordering would be to synthesize first those voxels having the maximum number of neighbors voxels with available range and intensity information. For future reference, we call this ordering the mnn-scan ordering. In sum, the following outlines the mnn-scan ordering process including the edge information:

Step 1: For each augmented voxel with unknown range, calculate the number of neighbors with already assigned intensity and range except, for those where an edge exists.

Step 2: Add the location of each augmented voxel to the corresponding list according to its number of neighbors.

Step 3: From the list of augmented voxels with the maximum number of neighbors, randomly choose the voxel to be synthesized and erase it from the list.

Step 4: Add 1 to the number of neighbors of each neighboring voxel of the voxel being synthesized. If an edge exists in any of the neighboring pixels, do nothing.

Step 5: Go to **Step 3** until there are no more voxels in the list to synthesize.

Step 6: Select the next list with the maximum number of neighbors, and go to **Step 3** until all lists are selected.

Step 7: Randomly synthesize all the augmented voxels having an edge.

We implemented the mnn-scan ordering and incorporate the edge information to our method. Our algorithm was tested again on the 30 images. The smoothing parameter for edge detection was set to 0.8 in all examples. For purpose of comparison, Figure 6 shows the results obtained (upper row) for the input images in Figure 5. The lower row shows the corresponding edge information. The MAR errors are now, from left ot right, 8.58, 13.48, 11.39 and 7.12, respectively [1].

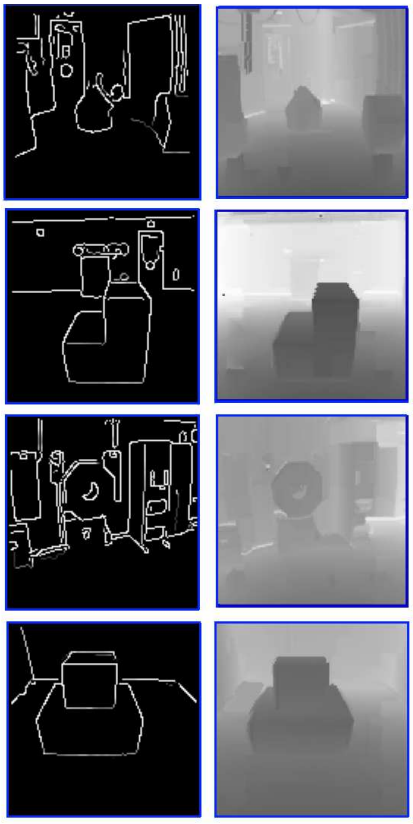


Figure 6: Range synthesis using the mnn-scan ordering and edge information. The left column shows the edges detected in the input intensity image and to their right is the synthesized result. For purposes of comparison, the input images are the same shown in the first two columns of Figure 5.

4.3 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 6.

[1]Computation time for these results using non-optimized code, is on the order of minutes on generic PC's.

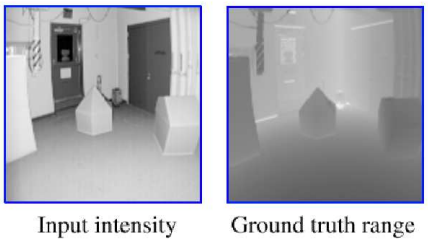


Figure 7: The input intensity image and the associated ground truth range. Since the unknown data are withheld from genuine ground truth data, we can estimate our performance.

Figure 8 shows three experiments. The initial range data is shown in the left column. The width of the input stripes r_w and the width of the area with missing range data x_w are indicated below each image. The right column shows the synthesized range data obtained after running our algorithm.

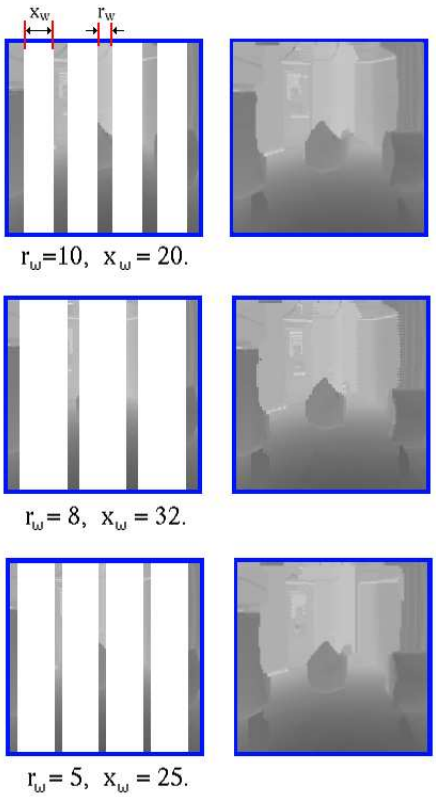


Figure 8: Results on real data. The left column shows the initial range data and to their right is the synthesized result (the white squares represent unknown data to be estimated).

The MAR errors for the experiments are shown in Table 3. The approximated depth size of the scene is 600 centimeters. It can be seen that the MAR errors are a bit high. However, the histograms of the pixels based on the residual errors for each case show that the most frequent residual error (the mode) is 7.32 cms.

Input range s_w	Input range x_w	% of area with missing range	MAR Error (in centimeters)
10	20	62.5	20.72
8	32	75	18.98
5	25	78	20.23

Table 3: MAR errors for the cases shown in Figure 8.

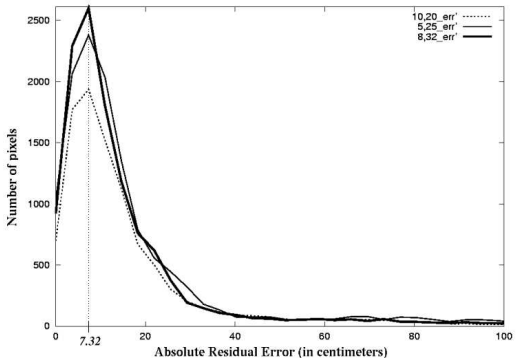


Figure 9: Histograms of pixels based on the absolute residual errors for the cases shown in Figure 8. Note that the most frequent residual error (the mode) is 7.32 centimeters.

In general, the synthesized range images are good in all cases. The very limited amount of input range was enough to capture the underlying structure of the scene. In Figure 10, 4 more examples are given for the case of subsampling with $r_w = 8$ and $x_w = 32$. Table 4 displays their respective MAR errors (from top to bottom).

MAR Error (in cms)	Approx. depth size (in cms)	MAR/Depth	Correlation coefficient
17.22	600	0.029	0.57
16.16	650	0.025	0.09
17.17	440	0.039	0.46
20.25	670	0.030	0.54

Table 4: The input information and MAR errors of the cases shown in Figure 10.

It is important to note, that depending on the scene, 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 difficult. However, the results presented here demonstrate that this is a viable option to facilitate environment modeling.

5 Conclusions and Future Work

We have presented a novel method for inferring range data given an intensity image with little associated range data. The method is proposed as an initial solution for the environment modeling problem in the context of mobile robotics.

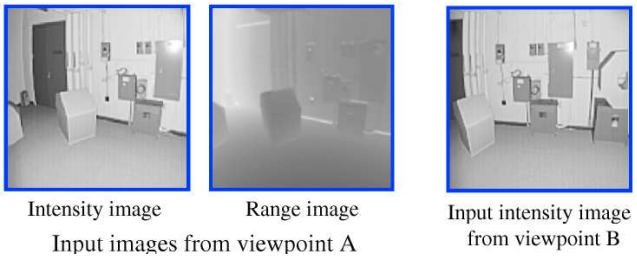
Our approach is based on Markov Random Fields to model the relationship between the observed range data and the variations in the intensity image, and use this to extrapolate/interpolate new range values. This approach was tested using data from a real environment with promising results.

Figure 10: Examples on real data for the case of subsampling with $r_w = 8$ and $x_w = 32$. The first two columns show the input intensity and range data, respectively. The last two columns show the synthesized results and the real range images for visual comparison.

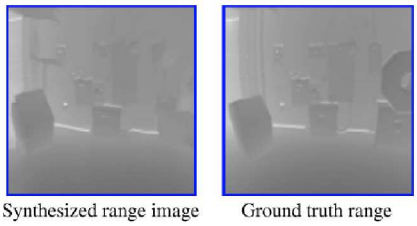
There are a number of parameters that can greatly influence the quality of the results: the size of the neighborhood used in computing correlations, the amount of initial range and the characteristics captured in that initial range. The characterization of how these parameters effect the results is the subject of ongoing work.

Our approach as described in this paper exploits the statistically observed relationship between the intensities in a neighborhood and range data to interpolate (or extrapolate) the range. While this formalism can explicitly capture local differential geometry, we do not explicitly compute local surface properties, nor does this approach make substantive assumptions regarding surface reflectance functions of surface geometry such as smoothness. The approach does assume that the relationship between intensity and range can be expressed by a stationary distribution; an assumption that could be relaxed. While avoiding strong assumptions about the surfaces in the scene allows greater generality, it also means we do not exploit potentially useful constraint information. In ongoing work, we are examining the incorporation of more elaborate priors and geometric inferences.

An interesting problem we are currently working on, is as follows: Having as an input the intensity image I_A and its associated range map R_A, taken from viewpoint A. And a second intensity image I_B, taken from viewpoint B, such that viewpoints A and B are spatially close to each other. Our objective is to infer the complete range map R_B associated



(a)



(b)

Figure 11: The input data to our algorithm is shown in (a). (b) shows the synthesized range from viewpoint B and the ground truth range for comparison purposes.

to I_B. In Figure 11a, we show the input data and Figure 11b shows the obtained result. In this context, we are also examining the use of multiple sensors with different resolutions and viewpoints.

Acknowledgments

We would like to thank the CESAR lab at Oak Ridge National Laboratory in Tennessee for making their range image database available through the website of the University of South Florida *http://marathon.csee.usf.edu/range/Database.html*.

The first author gratefully acknowledges CONACyT for providing financial support to pursue her Ph.D. studies at McGill University.

We would like to thank also the Federal Centers of Excellence (IRIS) for ongoing funding.

References

[Canny, 1986] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[Debevec *et al.*, 1996] P.E. Debevec, C.J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry and image-based approach. In *SIGGRAPH*, pages 11–20, 1996.

[Efros and Freeman, 2001] A. Efros and W.T. Freeman. Image quilting for texture synthesis and transfer. In *SIGGRAPH*, pages 1033–1038, August 2001.

[Efros and Leung, 1999] A. Efros and T.K. Leung. Texture synthesis by non-parametric sampling. In *ICCV (2)*, pages 1033–1038, September 1999.

[El-Hakim, 1998] S.F. El-Hakim. A multi-sensor approach to creating accurate virtual environments. *Journal of Photogrammetry and Remote Sensing*, 53(6):379–391, December 1998.

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

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

[Hertzmann *et al.*, 2001] A. Hertzmann, C.E. Jacobs, N. Oliver, B. Curless, and D.H. Salesin. Images analogies. In *SIGGRAPH*, August 2001.

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

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

[Levoy *et al.*, 2000] 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.

[Oliensis, 1991] J. Oliensis. Uniqueness in shape from shading. *Int. Journal of Computer Vision*, 6(2):75–104, 1991.

[Pollefeys *et al.*, 2000] M. Pollefeys, R. Koch, M. Vergauwen, and L. Van Gool. Automated reconstruction of 3d scenes from sequences of images. *ISPRS Journal Of Photogrammetry And Remote Sensing*, 55(4):251–267, 2000.

[Pulli *et al.*, 1997] 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*, pages 385–397, September 1997.

[Sequeira *et al.*, 1999] V. Sequeira, K. Ng, E. Wolfart, J.G.M. Goncalves, and D.C. Hogg. Automated reconstruction of 3d models from real environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54:1–22, February 1999.

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

[Storjohann, 1990] K. Storjohann. Laser range camera modeling. Technical report, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1990.

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