EUROGRAPHICS 2002 / I. Navazo Alvaro and Ph. Slusallek
(Guest Editors)

Automated Enhancement of 3D Models

L.A. Torres-Méndez and G. Dudek

Center for Intelligent Machines, McGill University, Montreal, Quebec, Canada

Abstract

The acquisition of a 3D model of a real environment can be accomplished using range sensors. In practice, suitable sensors to densely cover a large environment are often impractical. This paper presents ongoing work on the synthesis of 3D environment models from as little as one intensity image and sparse range data. Our method is based on interpolating the available range data using statistical inferences learned from the available intensity image and from those (sparse) regions where both range and intensity information is available. Since we compute the relationship between extisting range data and the images we start with, we do not need to make any strong assumptions about the kind of surfaces in the world (for example we do not need to assume the world exhibits only diffuse reflectance). Experimental results show the feasibility of our method.

1. Introduction

While Image-Based Rendering (IBR) has enjoyed great success, most rendering methods depend on 3D models. While there has been enormous progress on the automated acquisition of such models, the process of obtaining 3D data for real environments is often costly or labor intensive in practice. 3D scanners that produce range data can commonly be used for building models of individual objects, but even these often need to be refined using manual intervention [?].

More importantly, while volume scanners for measuring the 3D layout of a large environment exist, they often suffer from either high cost, non-uniform coverage of the environment, drop-outs or other complications. Laser line-scanners are attractive but only produce measurements along a "slice" through the environment and covering a volume with such slices entails significant complications.

In this paper we look at how one can combine a monochromatic intensity image with *very sparse* 3D data to interpolate the missing 3D measurements: this can be referred to as Image-Based Modeling (sadly, the acronym of choice is already taken by a small New York company). There has been work on inferring 3D structure from intensity data using "shape-from-shading" methods but this tends to be inaccurate, dependent of strong knowledge of reflectance functions, and highly error-prone. Our approach, in contrast, is based on computing the statistical relationship between the (limited) observed range measurements and the intensity images and using statistical inference to fill in the missing range based on the appearance of the available intensity data.

The ability to reconstruct a 3D model of an object or scene greatly depends on the type, quality and amount of information available. We are interested in the class of environment modeling which involves the recovery and representation of photometric and 3D geometric information of indoor scenes. This particular problem is challenging because if a perfect 3D model is desired, huge amounts of data from several different viewpoints of the object or scene are needed. However, this is both impractical and computationally costly and demanding. This being the case, the method described here was designed to overcome this problem; it is aimed at reconstructing good 3D models while using a relatively small amount of information and facilitating the acquisition process.

Our approach is based on assumptions regarding the coherence of surfaces in the world and their causal interrelationship. These are formalized in the form of a Markov Random Field (MRF) model of how both pixels and voxels inter-relate. Rather than performing our computations in only the pixel domain, or the voxel domain, we make our inferences in a compound higher-dimensional space that combines reflectance data and spatial occupancy. It should be noted that using Markov Random Fields for image completion has previously proven successful in texture synthesis [?, ?]. In the context of texture synthesis, MRF methods model a texture based on its local stationary properties. A new texture is generated pixel by pixel in such a way that these two properties are preserved in a small set of spatially neighboring pixels which characterizes every pixel on the texture image. Instead of modeling textures, we model common characteristics between range and intensity images.

The organization of this paper is as follows. In the next section, we review some related work. In Section 3, we describe our range synthesis method. Some experimental results are shown in Section **??**, and, finally, in Section **??** we give some conclusions and suggestions for future research.

2. Previous Work

The process of producing a 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 a range scans), combining them onto surfaces and then generating suitable models for graphics applications entail distinct computations (this is quite typical in automated measurement applications) [?, ?, ?]. In this paper we focus only on the processes of obtaining 3D data.

Most prior work on synthesis of 3D environment models uses one of either photometric data (intensity) or geometric data (range) [?, ?, ?] to reconstruct a 3D model of an scene. Fitzgibbon and Zisserman [?] 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 solve the feature correspondence problem using methods 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. A similar work is that of Pollefeys et. al. [?], they obtain a 3D model of an object from image sequences acquired from a freely moving camera. The camera motion and its settings are unknown. 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. It appears that if one uses information of only one type, the reconstruction task becomes very difficult and works well only under narrow constraints. For example, shape-from-shading methods [?, ?] reconstruct a 3D scene by inferring depth from a 2D image; in general, this task is difficult, requiring strong assumptions about reflectance and bounds surface variability. More recently several researchers have considered using two or more types of data. Specifically, the combination of intensity information and range data appears to be particularly promising due to the obvious relationship between many types of singularity in these two domains [?, ?, ?, ?, ?]. Pulli et al. [?] address the problem of surface reconstruction by scanning both the intensity in different colors channels and the 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 accomplished by using a hierarchical space carving method. The integration of intensity data with range data has been proposed [?] to help define the edges of surfaces extracted from 3D data since 3D measurements are often somewhat sparse and do not explicitly define surface boundaries. In this work is was necessary to hypothesize the existence of surface continuity and intersections among surfaces, and the formation of composite features from the surfaces.

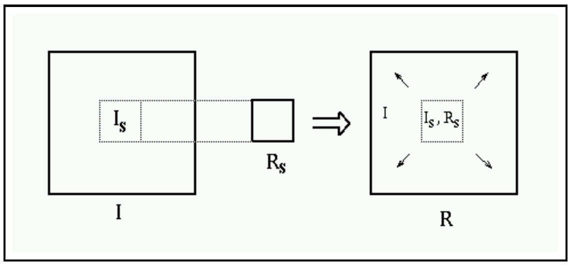
However, one of the main issues in using the above configurations is that the acquisition process is very expensive because dense and complete intensity and range data are needed in order to obtain a good 3D model. As far as we know, there is no method that bases its reconstruction process on having 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.

3. The Algorithm

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 make two key assumptions: (1) the initial range data and intensity data is registered and (2) the range data is clumped into at least some sets of mutually-adjacent voxels (as opposed to scattered measurements far from one-another). In this paper we are only interpolating data across a single range scan (i.e. a graph surface of the form $z(x,y)$). Note that while the process of inferring distances from intensity superficially resembles shape-from-shading, we do not depend on prior knowledge of reflectance or on surface smoothness or even on surface integrability (which is an technical precondition for most shape-from-shading methods, even where not explicitly stated).

We use a Markov Random Fields (MRF) to express that relationship between local neighborhoods of range and intensity data. Although intensity alone does not constrain the surface depth, intensity information can be viewed as a probabilistic bias on the extrapolation of range data. As in MRF-based methods, we assume that the intensity and range value at each pixel depends only on its immediate spatial neighborhood. While this assumption is not strictly valid in real scenes, the local coherence of matter makes this an acceptable approximation since it only needs to hold in between the original 3D estimates. Contrary to most MRF based algorithms, our approach is completely deterministic, meaning that no explicit probability distribution needs to be constructed.

It is important to highlight that we are not just dealing with intensity images that represent textures, but arbitrary images which may contain one or more objects. Since range data of the form $z(x,y)$ is represented in the same form as the intensity images, the implementation is straightforward.



(a)

(b)

Figure 1: *Algorithm overview.*

3.1. Synthesizing range

We start by considering an illustrative example with a simplified (special-case) geometry. Let I be an intensity image of size $m \times n$ and R_s be a range image of size $r \times s$, such that $m \times n \gg r \times s$, and the area covered by R_s is also covered by I. Let I_s be the intensity image of size $r \times s$ where R_s overlaps I. The complete range image R of size $m \times n$ is to be synthesized from the combined information of R_s and I_s and from intensity data available around the pixel to be synthesized (see Figure **??**a).

The algorithm starts growing R_s pixel by pixel incrementally until its size equals the size of R. The range value of pixel $\vec{p}$ with position (x,y) in R is determined by comparing its intensity and range spatial neighborhoods $n_I(\vec{p})$ and $n_R(\vec{p})$, respectively, against all current available intensity and range neighborhoods $n'_I(\vec{p})$ and $n'_R(\vec{p})$. The range value of the pixel with the most similar range and intensity neighborhoods is assigned to $\vec{p}$. Similarity between spatial neighborhoods is calculated as follows,

$$\sum_{\vec{q}\varepsilon\vec{N}_i} G(\sigma,\vec{q}-\vec{q}_o)[(n_I(\vec{q})-n'_I(\vec{q}))^2+(n_R(\vec{q})-n'_R(\vec{q}))^2]$$

where $\vec{N}_i$ is the neighborhood, $\vec{q}\varepsilon(x,y)$ is a pixel position, n_I and n_R represents the intensity and range spatial neighborhoods respectively, around the pixel to be synthesized, and n'_I and n'_R are the intensity and range neighborhoods to be compared with. G is a 2-D Gaussian kernel that gives more weight to those pixels near the center than those at the edge of the window. Figure **??**b gives a graphical illustration of the range synthesis process.

The shape and dimension of the intensity and range spatial neighborhoods will directly determine the quality of R; we chose to model the spatial neighborhood of a pixel as a square window around that pixel, where only those pixels with already assigned intensity and/or range values are considered in the synthesis process. Thus, the neighborhood is in fact, of an arbitrary shape depending on the current available information on each of its pixels. The size of the neighborhood is a parameter that can be defined by the user.

4. Experimental Results

We present illustrative results of our approach based on trials with both a very simple synthetic data set as well as real range data from a volume scanner. In the cases shown here, we start with a complete range scan (real or synthetic) that can be used as ground truth to estimate correctness of our solution, and then remove some portion of that data to use as input to our process. Of course in practice, the complete ground-truth data set would not be available.

The synthetic data were generated with the 3D rendering package PovRay. In the left side of Figure **??**a a synthetic intensity image of an empty bookshelf is shown, the subwindow in the right side is the associated range image taken from the position indicated by the red rectangle in the intensity image, these images are given as an input to our algorithm. For this case, the size of spatial neighborhoods n_I and n_R, is set to be 9×9 pixels. The left side of Figure **??**b shows the range synthesis results, and to the right the complete synthetic range data is shown. It can be seen that the 3D

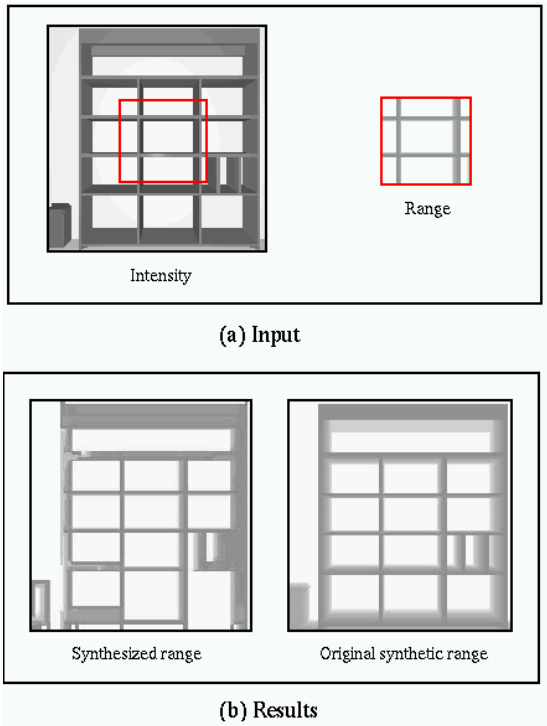


Figure 2: *Results on synthetic data.*

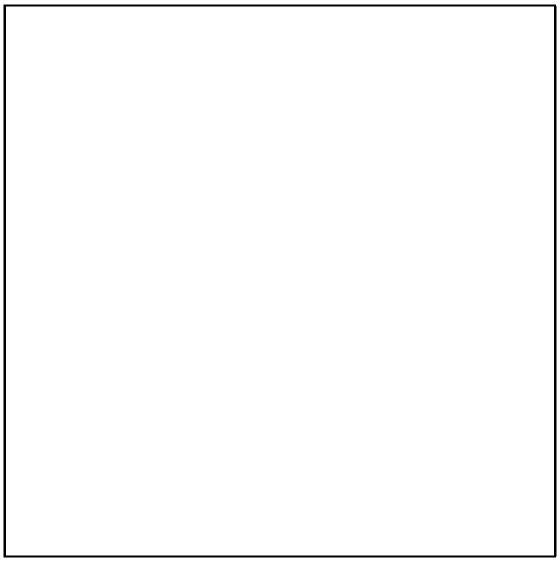
recovery process captures most of the 3D structure indicated by the intensity image. Note that while this synthetic example is very simple, it would present an essentially impossible challenge to traditional shape-from-shading methods.

Representative results on range data from a real environment are shown in Figures **??** and **??**. The real intensity and range images of indoor scenes were obtained from the USF range image database [†]. As with the synthetic data, 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 two strategies for subsampling the range data analogous to the kinds of data returned by two key classes of real range scanner.

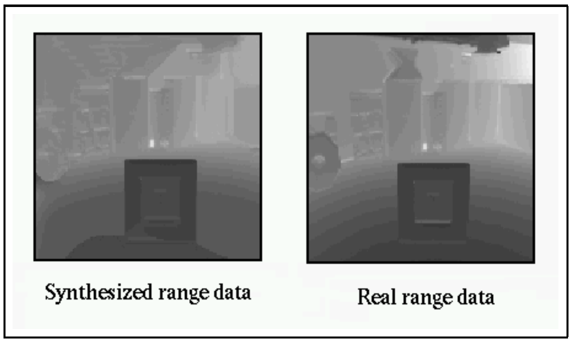
4.1. Limited dense range

The first type of experiment involves the range synthesis when the initial range is acquired in rectangular regions of size $m \times n$ and at position (r_x, r_y) on the intensity image. Figure **??**a shows the intensity image (left) of size 128×128

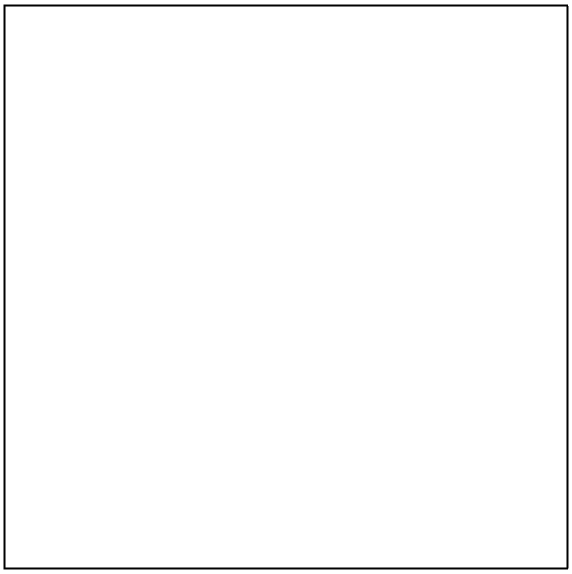
[†] http://marathon.csee.usf.edu/range/DataBase.html.



(a)



(b)



(c)

Figure 3: *Results on real data. (a) Input. (b) Results. (c) Results in 3D.*

and the initial range (right), a window of size 64×64, i.e. only the 25% of the total range is known. Here, the size of the spatial neighborhoods is 5×5 pixels. The synthesized range data obtained after running our algorithm is shown in the left side of Figure **??**b; for purposes of comparison, we show the complete real range data (right side). Figure **??**c displays two different views using our synthesized range and the real range. It can be seen that the synthesized range is very similar to the real range. The percentage of the residual error is 13.06%.

4.2. Sparse range measurements

In the second type of experiment, the initial range data is a set of stripes with variable width along the $x-$ and $y-$axis of the intensity image. We tested with the same intensity image used in the previous section in order to compare both results. Two experiments are shown in Figure **??**. The initial range images are shown in the left column (the white rectangles are the regions to be estimated), and to their right are synthesized results. In Figure **??**a, the width of the stripes s_w, is 5 pixels, and the area with missing range data ($x_w \times x_w$) is 25×25, i.e., 39% of the range image is known. The size of the spatial neighborhoods is 5×5 pixels. For Figure **??**b, the values are $s_w = 3$, $x_w = 28$, in this case, only 23% of the total range is known. Here, the size of the spatial neighborhoods is 3×3. The percentages of the residual errors for the reconstruction are 4.85% and 6.67%, respectively. Figure **??** shows the 3D views for both cases. The 3D views for the real range are shown in the right column of figure **??**c.

Figure **??** shows the density of pixels at different depth values in the original and synthesized range of Figure **??**a.

The results are surprisingly good in both cases. Our algorithm was capable of recovering the whole range of the image. We note, however, that results of experiments using stripes are much better than those using a window as the initial range data. Intuitively, this is because the sample spans a broader distribution of range-intensity combinations than in the local window case.

Our algorithm was tested on images of common (non-simple) scenes found in a general indoor man-made environment. It is important to note, that the initial range data given as an input is crucial to the quality of the synthesis, that is, if no correlated changes exist in the range and intensity, then the task becomes difficult. Also, existing discontinuities in the depth map that are not captured in the initial given range and intensity, will be difficult to detect and therefore the estimation will be wrong. However, since we are dealing with smooth surfaces mostly existing in indoor environments, the results presented here demonstrate that this is a viable option to facilitate environment modeling.

Figure 4: *Results on real data. The left column shows the initial range data and to their right are synthesized results. (a) $r_w = 5$, $x_w = 25$ (b) $r_w = 3$, $x_w = 28$.*

5. Conclusions and Future Work

We have presented an algorithm for recovering 3D geometric data given an intensity image with little associated range information. The approach uses Markov Random Field as a statistical model for how range and intensity inter-relate. There are a number of parameters that can greatly influence the quality of the results — size of the spatial neighborhoods, the amount of initial range and the characteristics captured in that initial range. In ongoing work we are examining the inter-relationship between these parameters.

In this work we exploit statistical relationships between range data and intensity to interpolate the range map. Doing so in the absence of assumptions regarding surface reflectance or geometry gives us a great degree of generality. On the other hand, where such information is available we are neglecting potentially powerful sources of information that might allow us to interpolate range data over much wider regions (i.e. with fewer original measurements). In particular, we are considering extensions of this work that explicitly consider using stronger assumptions or biases regarding the nature of the scenes we are observing. This seems to be a natural direction since such assumptions are used in subsequent triangulation processes that are a requisite part of the graphics pipeline.

Acknowledgements

We would like to thank the CESAR lab at Oak Ridge National Laboratory (Tennessee) for making their range image database available for research purposes.

Figure 5: *3D views from the results of Figure* **??**. *The vantage points are the same as those shown in Figure* **??***c for comparison purposes. (a)* $r_w = 5$, $x_w = 25$ *(b)* $r_w = 3$, $x_w = 28$.

Figure 6: *Density of pixels at different depth values of the real and synthesized range of Figure* **??***a.*

References

1. B. Curless and M. Levoy. A Volumetric Method for Building Complex Models from Range Images. *Proceeding of SIGGRAPH'96*, pp. 303-312, August 1996.

2. A. Efros and T. Leung. Texture synthesis by non-parametric sampling. In *Proc. of IEEE ICCV'99*, Greece, pp. 1033–1038, September 1999.

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

4. A.W. Fitzgibbon and A. Zisserman. Automatic 3D model acquisition and generation of new images from video sequences. In *Proc. of European Signal Processing Conference*, Greece, pp. 1261–1269, 1998.

5. A. Hilton. Reliable Surface Reconstruction from Multiple Range Images. In *Proc. of ECCV'96*, April 1996.

6. H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle. Surface Reconstruction from Unorganized Points, In *Proceedings of SIGGRAPH '92*, **26**(2):71-78, July 1992.

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

8. 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*, 2000.

9. J. Oliensis. Uniqueness in Shape From Shading. In the *Int. Journal of Computer Vision*, **6**(2):75–104, 1991.

10. M. Pollefeys, R. Koch, M. Vergauwen, L. Van Gool. Metric 3D Surface Reconstruction from Uncalibrated Images Sequences. *SMILE Workshop*, pp. 139–154, 1998.

11. 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, Springer-Verlag, pp. 385–397, Italy, September 1997.

12. V. Sequeira, K. Ng, E. Wolfart, J.G.M. Goncalves, D.C. Hogg. Automated Reconstruction of 3D Models from Real Environments. In *ISPRS Journal of Photogrammetry and Remote Sensing* (Elsevier), **54**:1–22, February 1999.

13. I. Stamos and P.K. Allen. 3D Model Construction using range and image data. In *CVPR 2000*, South Caroline, June 2000.

14. G. Turk and M. Levoy. Zippered polygon meshes from range images. In *Proceedings of SIGGRAPH'94*, pp. 311-318, Orlando, Fla., July 1994.

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