Noise vs Feature:
Probabilistic Denoising of Time-of-Flight Range Data

Derek Chan
CS229 Final Project Report
ddc@stanford.edu

Abstract

Advances in active 3D range sensors have enabled the recording of depth maps at video frame rates. Unfortunately, the captured depth data is often noticeably contaminated with noise. We present a series of statistical analyses and denoising techniques to improve the raw output of depth sensors. Unifying our investigations is the concept that in the presence of high sensor noise, the ability to distinguish effectively between noise corruption and real depth features is needed to reconstruct the original true geometry realistically.

1 Introduction

Range sensors, such as time-of-flight (TOF) cameras, have enabled the reconstruction of 3D geometry independent of texture required for stereo vision methods. Unfortunately, range sensors which are able to produce a 2D array of depth values at real-time speeds generally produce data heavily corrupted by noise. In this paper we elect to focus on a particular TOF camera, the Mesa SwissRanger SR3000. The SR3000 is capable of capturing 176x144 resolution depth maps at approximately 30 frames per second.

Although depth sensors are increasingly being use for a variety of commercial and academic applications, the high level of noise makes the effective use of these sensors challenging; thus, we propose a technique for estimating the true original geometry of a scene based upon an input depth map. Following certain previous investigations (noted in Sect. 2) in the area of denoising we have elected to use a Bayesian model for our approaches. In Section 3 we present a series of techniques which illustrate some of the issues that depth reconstruction techniques should be aware of to enable the better recovery of true geometry by denoising of range data. In particular we endeavor to tackle raw data that is noisier than previous works have used. The high noise to distinguishable feature ratio in the raw data makes naïve denoising techniques problematic. We seek to sharpen edges and blur noise; however, this requires being able to distinguish the categories. In Section 3.2 we argue how such distinctions allow for data dependent denoising choices. In Section 4, we conclude with further suggestions as to further applications of such a model.

2 Related Works

There exist present techniques of obtaining less noisy depth maps from these sensors. The simplest method is to integrate over time by averaging many frames together; however, this is obviously not a viable option in real time scenarios. We seek to provide a method which could be applied to individual depth frames since we believe such a method would be general and allow for use in a greater set of scenarios including dynamic scenes. Most related to our work is that by [Diebel06] where a Bayesian model was also used to perform probable surface reconstruction based upon slightly noisy depth data. A primary difference in our work is in the type of sensor used. The previous work focused on synthetic data or used data obtain from a laser line scanner. These sensors capture at much slower rates than TOF sensors, but they also provide much stabler measurements. In comparison, our sensors have a much higher noise to size of feature ratio. However like the aforementioned work, our technique will ultimately be a Bayesian method which will estimate the true depths under statistically learned assumptions. In addition the previous work primarily focused on using simply the depth measurements from their sensor while we hope to also use other data and statistics such as the gray scale intensity image provided by our TOF camera.

Also relevant to our work are investigations into upsampling depth data. Some of these techniques are robust for at least a small amount of noise in the input depth maps, and in the process of increasing resolution, also effectively do some noise reduction. In particular upsampling does smooth out some peppering noise. [Schuon08] used the concept of multiple slightly shifted captures of the same scene to provide increased information on a scene to upsample it. However this technique also relies upon multiple images just as time averaging frames does. The use of a high resolution color image as a guide to upsample low resolution depth [Diebel05, Kopf07, Chan08]. Since the color image is comparatively noise free, these techniques ultimately also guide the result towards being comparatively noise free. This technique relies upon a fairly accurate alignment between the color and depth images. Since no sensor presently exists with a unified lens set, imperfect calibration procedures generally results in some misalignment artifacts in these setups when used on real world scenes.

Finally the task of super resolution for standard color images is also related to our goal. Towards this endeavor, learning techniques exist [Sun 2003, Freeman 2002] which ultimately manufacture plausible high frequency details for a low resolution image based upon training sets of low resolution vs high resolution pairs. Previous works use a number of ideas on different statistical priors of real images which will provide us with ideas for construction of potential priors for our scenario where we train upon low noise vs and high noise data pairs of the same scene. Where as these previous works sought to improve visual upsampling in a relatively noise free RGB space, our analysis is in the realm of noisy depth images. This poses the possibility for investigating some of the characteristics unique to depth sensors. Most noticeably the data we have provided gives depth and intensity values. This allows us derive further features such as patch orientation for a given pixel.

3.0 Denoising Investigations

3.1 Probabilistic Denoising Framework

Our framework is admittedly similar to that denoted in [Diebel06]. In our scenario, we are provided with depth measurements m, which are some noise corrupted approximation of the true scene distance x. We wish to find the probable value for x given the measurements based upon some statistical assumptions. Thus, we wish to optimize over the following objective function.

$$p(x|m)=p(m|x)\,p(x)/p(m) \qquad (1)$$

where p(m) is a constant with respect to our optimization over the x's. The x, which minimizes this objective, will provide our best guess for the true scene distances.

$$\hat{x}=argmin_x(-\log p(m|x)-\log p(x)) \qquad (2)$$

Here, the first term (log p(z|x)) will henceforth be denoted the measurement potential, *M*, because it is a probability of the noise in the measurements given the real data. The second term (log(p(x))) is our surface smoothness prior, *S*, which is a prior probability of the smoothness of surfaces in the world. These two terms counteract each other during the optimization process.

As mentioned, one method for denoising depth maps is to fuse the data with aligned unnoisy color data under the general assumption that that areas of similar color are likely to have similar similar depth. We start with a reimplementation of the Markov Random Field (MRF) objective from [Diebel05] to demonstrate some of the problems with existing approaches. With this heuristic, our measurement potential simply checks the squared distance between the estimated real depth *x* with the actual measurement *m*.

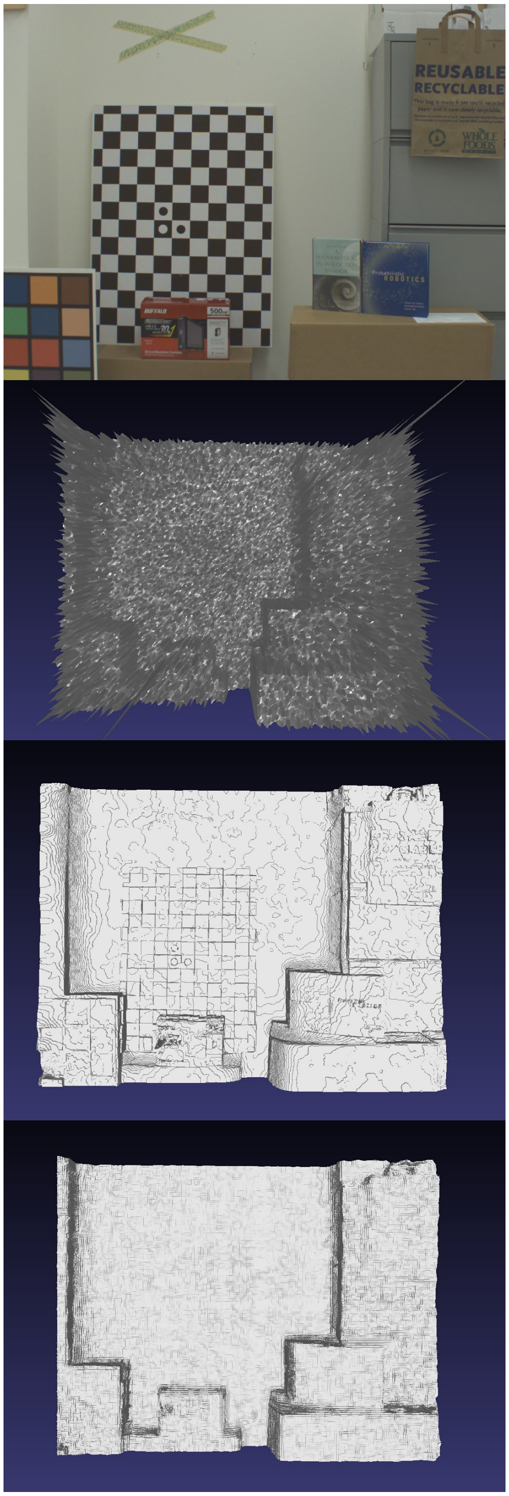

Figure 1: Scene (top) with raw depth map (top middle) is denoised using the single mode MRF method (bottom middle) using both the depth map and color image as inputs. Note the texture embossing from the colour image onto the depth map for the checkerboard, book cover and shopping bag text. In comparison when a 'noise vs feature' decision is made (bottom) the texture copying is removed.

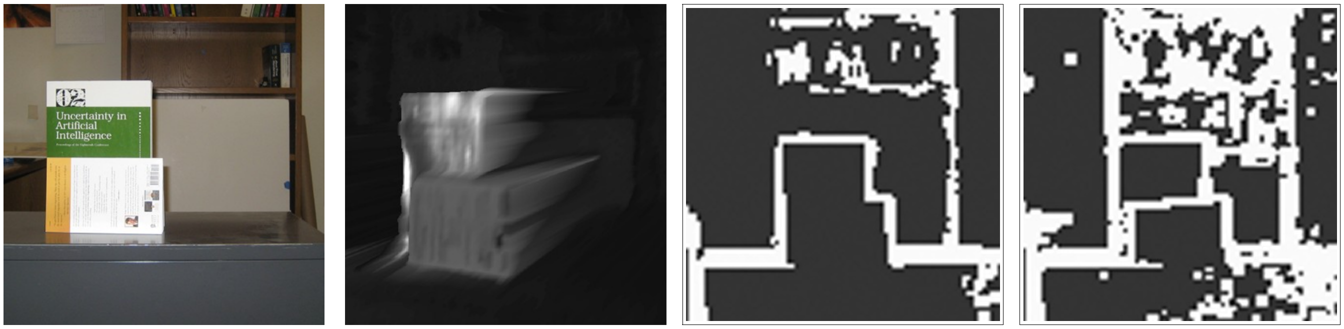

Figure 2: The scene (left) of two separated books has a 3D reconstruction (middle left) from raw data. Note the depth separation of the books is evident in the 3D scene. However, when we use a simple heuristic (Sect. 3.2) for distinguishing features from noise, this depth discontinuity can not be effectively determined. Large thresholds (middle left) fail to mark the separation between the books. Small thresholds (left) cause the general depth map to be contaminated with noise.

$$M=\sum_{i=1..n} k*(x_i-m_i)^2 \quad \textit{// k is a weighting constant}$$

S tries to maintain depths for neighboring pixels when their visible colors, denoted *I*, are similar.

$$S=\sum_{i}\sum_{j\in\Omega_i} e^{-c*\|I_i-I_j\|^2}(x_i-x_j)^2$$

By taking the derivatives of this objective function, we can set up this optimization to be solved using a conjugate gradient optimizer. A result from this process is shown in Figure 1. Although the output depths are cleaner, this naïve approach also noticeably suffer from texture embossing since noise along color edges are enhanced to appear incorrectly as depth discontinuity.

3.2 Dual Mode Denoising

Alternatively, we design an augmented bilateral filter [Tomasi98] kernel. Within this closed formed solution, we distinguish between noise and feature with a simple metric removing texture copying as shown in Figure 1. The filter is:

$$x_p=\frac{1}{k_p}\sum_{q\in\Omega} I_q\, f(\|p-q\|)\, d(\Omega,\|I_p-I_q\|,\|m_p-m_q\|)$$

Here, *q* is an index for a nearby value from the set of surrounding data points omega. *f* is a Gaussian over spatial differences; thus without function *d*, we would have a Gaussian blur.

$$d=\alpha(\Omega)\, g(\|I_p-I_q\|)+(1-\alpha(\Omega))\, h(\|m_p-m_q\|)$$

d provides a blend between two different Gaussian functions. *g* operates upon differences in visible color while *h* operates upon differences in depth. When noise is present without features, we need not consult the color image, and the input depths can simply be blurred. This decision is made by function *alpha* which takes in the neighborhood of depths and appropriately samples from the sigmoid function.

In the naïve MRF technique color features are erroneously visible in the depth map. The objective function does not differentiate between noise and features; thus, noise along color edges is enhanced. The bilateral filter method effectively has a dual mode measurement potential where the function alpha makes a decision between feature or noise and either chooses to edge enhance from the color image or blur from the depth map respectively.

In the augmented bilateral filter technique, the decision function, *alpha*, between noise vs feature is made using a simple heuristic of comparing the maximum and minimum values within a neighborhood of a blurred copy of the input. However, this metric is not entirely robust (Fig 2). When small thresholds are shown foreground features (such as the separation of the two books) are visible but much background noise is marked as feature. When large thresholds are chosen, foreground edges are not appropriately marked. A thresholding problem exists with the current metric; thus, it seems as if a distance dependent threshold could be learned from data to combat this problem. With a noise model, we could for example mark small discontinuities in the foreground as feature while seeing the same depth difference more likely as noise if the measurement is faraway.

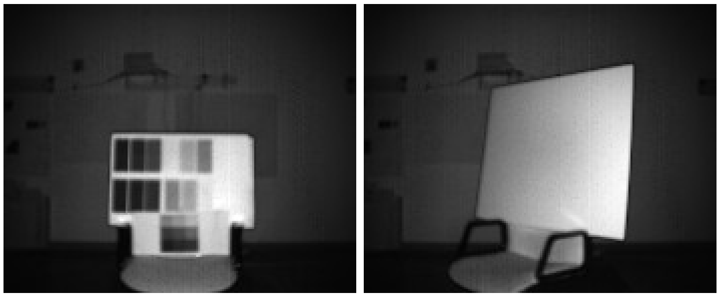

Figure 3: Intensity images provided by the depth sensor. To gather data for our noise model, we collected data of a Macbeth check at different distances and of a board at different orientations and distances.

3.3 Noise Statistics and Model

Based upon the results from our dual mode design, we seek to learn a noise model to better guide the noise vs feature decision. For TOF range sensors, the random error in measurement lies primarily along the ray from the sensor. From recorded tests we derive that the noise can be approximated by a Gaussian. The noise for measurements is seemingly dependent on a number of parameters. In our current investigation, we have focused on parameters based upon the inherent measurement principle of the SR3000 which relies on the reflection off surfaces of infrared rays from the camera.

Choosing parameters likely to change from this measurement principle we decided upon the following:
(1) distance from the sensor
(2) incident angle between the infrared (IR) measurement rays and the surface being measured
(3) infrared absorption of the material being measured

We collect data of a board at a variety of different orientations and distances (Fig. 3). In addition recordings or a Macbeth check at different distances is taken to measure noise due to intensity differences. Some generalizations of the results are shown in Figure 4. As expected, the variance over time of the measurement of a surface increases as distance increases, incident angle increases, and IR absorption increases. In addition, the greatest variance is seen for low intensity values (ie high IR absorption) as expected since in this case no measurement rays are reflected back towards the sensor.

Using this captured data we can run a gradient descent algorithm over a parameter space specified over a polynomial for each of the above variables. Thus we learn a polynomial regression for the variance values dependent on our parameters. This model is used in the formulation of a measurement potential. We next note an example objective for denoising using just a single depth sensor by itself.

3.4 An Application Upon Single Perspective Denoising

The denoising of a single depth frame can be done with the following proposed objective. To evaluate the similarity between the measured depths and range values obtained during conjugate gradient optimization, we use the Mahalanobis Distance between the recorded data and optimized values as our measurement potential.

$$M=\sum_{i=1..n}(m_i-x_i)^T(.5\,\Sigma_i^{-1})(m_i-x_i)$$

where *sigma* is the covariance matrix of our measured data. It is specified as:

$$\Sigma=R^T\begin{pmatrix}\sigma_x^2 & 0 & 0\\ 0 & \sigma_y^2 & 0\\ 0 & 0 & \sigma_z^2\end{pmatrix}R$$

In our case, *sigma z* is the variance in the ray direction where we can use the results of our model. The other variances are provided as small constants. *R* provides a rotation matrix into the ray direction for a given pixel in our depth map. For our smoothness prior we use the Laplacian surface prior below where *omega* is once again the surrounding neighborhood. This metric is passed through function *h* which is a Huber function. The Huber function returns an L1 norm for large inputs otherwise it returns an L2 norm. The L1 norm for large distances makes this theoretically less negatively influenced by sudden large peppering noise spikes.

$$S=\sum_{i=1..n}h(x_i-\frac{1}{|\Omega_i|}\sum_{q\in\Omega_i}q)$$

Application of this optimization to a raw depth map results in a noticeable cleaner range output (Fig. 5)

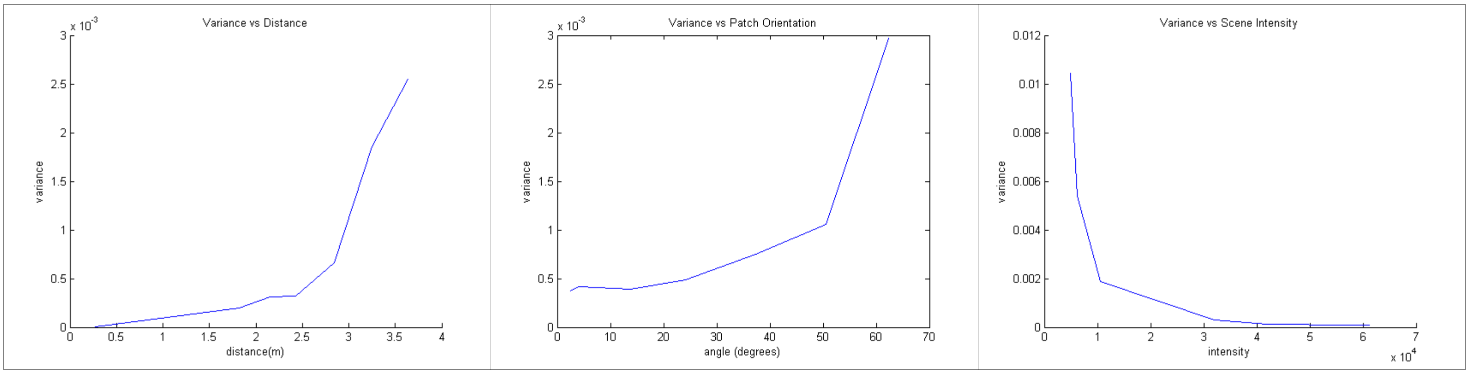

Figure 4: From the collected data, example plots of variance compared with each of our feature statistics when the other parameters are held to some constant. The calculated variance is based off of measurements made in meters. The units of scene intensity are from the raw output of the depth sensor which provides a 12-bit gray scale intensity reading along with the depth measurements.

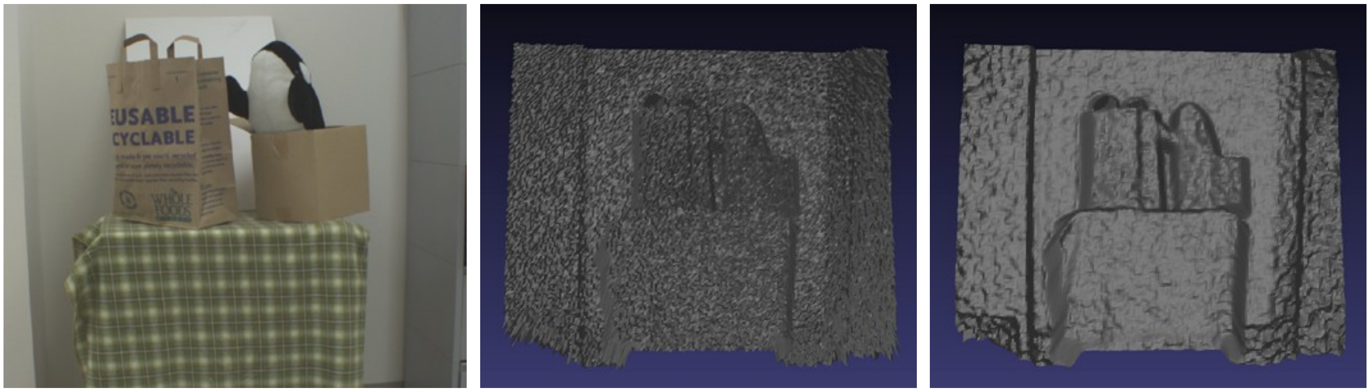

Figure 5: The scene (left) with raw 3D reconstruction (middle) is passed through our single sensor optimization to produce the result (right). Note that the input has a fair amount of noise throughout the reconstruction. In the output, details such as the folds in the green blanket and paper bag are more evident.

4 Conclusion and Future Work

We have presented a set of different techniques for denoising range data. In particular we demonstrated the need for a dual mode objective which differentiates between noise and feature to perform separate actions as necessary. For this goal we have designed a data model for the noise variance based upon recorded measurements. Further quantitative analysis of our model should be performed. In addition , present ground truth for our data model is obtained through temporal averaging of frames from the SR3000. This unfortunately does not take into account the slight systematic bias of the sensor. Comparisons with measurements from a more accurate device such as a laser range finder could increase accuracy.

Insertion of a noise model into an objective with a dual mode surface prior term which smooths or edge enhances as is appropriate seems the next logical step. In particular the gradient profile prior (Sun08) seems a good candidate for enhancing known features. Further testing with a variety of surface priors should also be performed.

5 Acknowledgments

I would like to thank Professor Christian Theobalt, Professor Sebastian Thrun, and James Diebel for their advice through my TOF data investigations as well as providing me with the sensor and computing equipment. The denoising investigations presented in this work are part of a larger set of investigations into TOF camera data reconstruction and data fusion with color cameras that started previous to this quarter and will be continuing past this class. Some of the bilateral filter investigations discussed culminated in mid-October of this year; however, the work is still ongoing. I am also grateful to Sebastian Schuon for the recent setup of a motor mounted laser range finder that will allow me to continue with these investigations using a more reliable basis for ground truth measurements.

6 References

D. Anderson, H. Herman, and A. Kelly. Experimental characterization of commercial flash ladar devices. International Conference of Sensing and Technology 2005.

D. Chan, H. Buisman, C. Theobalt, and S. Thrun, A Noise-Aware Filter for Depth Upsampling. ECCV Workshop 2008.

J. Diebel, S. Thrun, and M. Breunig. A Bayesian Method for Probable Surface Reconstruction and Decimation. ACM Transactions on Graphics 2006.

J. Diebel and S. Thrun. An Application of Markov Random Fields to Range Sensing. NIPS 2005.

W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based superresolution. IEEE Computer Graphics. 2002.

J. Kopf, M. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. ACM Transactions on Graphics 2007.

A. Levin, A. Zomet, and Y. Weiss. Learning to perceive transparency from the statistics of natural scenes. NIPS pages 1247–1254 2002.

C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In ICCV, pages839–846 1998.

S. Schuon, C. Theobalt, J. Davis, and S. Thrun. High-quality scanning using time-of-flight depth superresolution. CVPR Workshops 2008.

J. Sun, Z. Xu, and H. Shum. Image super-resolution using gradient profile prior. CVPR 2008.

J. Sun, N. N. Zheng, H. Tao, and H. Y. Shum. Image hallucination with primal sketch priors. CVPR 2003.

Q. Yang, R. Yang, J. Davis, D. Nistér, Spatial-Depth Super Resolution for Range Images, CVPR 2007.