RGBD-Fusion: Real-Time High Precision Depth recovery

Ankith Bheemanakone Venkatagiri

Department of Informatics - Technische Universität München

Abstract

The introduction of consumer based RGBD scanners set off a major boost in 3D computer vision research. Yet, the precision of existing depth scanners is not accurate enough to recover fine details of a scanned object. The paper presents a method to fuse intensity data and depth information obtained from RGBD scanner under natural illumination, in order to enhance depth and recover fine details in the reconstructed image. The high precision depth is obtained from - "shape from shading technique", without finding and integrating the surface normals[5].

1 Introduction

The advent of low cost RGBD scanners triggered the development of new algorithms to exploit the associated intensity image to improve the lack of accuracy of the scanners. The goal is to fuse intensity data and depth map obtained from RGBD scanner in order to compensate for measurement errors inherent in the depth scanners and to obtain fine details in the reconstructed image. The image is reconstructed using a photometric stereo technique which estimates the surface normals of the object by observing the object under different lighting conditions. It is based on the fact that the amount of light reflected by a surface is dependent on the orientation of the surface in relation to the light source and the observer. By measuring the amount of light reflected into a camera, the space of possible surface orientation is limited. Given enough light sources from different angles, the surface orientation may be constrained to a single orientation or even overconstrained. Therefore, a specialized photometric stereo technique called shape from shading is used to reconstruct image from a data of a single image. But shapes recovered from shape from shading suffer from ambiguities since there can be several other possible surface to explain a given image, especially under natural illumination. These ambiguities can be eliminated by combining data from depth sensor with shape from shading.

Apart from shape from shading, to account for various lighting conditions and various local lighting effects such as specularities and interreflections, a lighting model is used. This lighting model is assumed to account for all lambertian and non-lambertian reflections. Lighting model uses surface normals that are estimated from the input depth map making the algorithm less sensitive to the calibration of the RGBD scanner. Lighting model estimate is expressed in terms of surface gradient and its normals. This eliminates the need for finding and integrating surface normals like in traditional shape from shading technique. But expressing lighting model estimate in terms of surface gradients, makes the model non-linear. In order to achieve fast convergence, the variational model of lighting estimate is relinearized.

The main contributions of the paper are:

1. A method to enhance depth accuracy in the reconstructed image by efficiently fusing single scene RGBD inputs.

2. Proposes a non-traditional way of estimating shape from - "shape from shaping technique", without integrating the surface normals.

3. Proposes a real-time depth enhancement method which operates under natural illumination and with multiple albedo objects.

2 Shape Refinement Framework

The intensity image and depth map of a naturally illuminated scene is obtained from a stationary RGBD scanner. The intrinsic matrices and extrinsic parameters of the depth and color sensors is assumed to be known. However, due to measurement inaccuracies and inherent errors present in the depth sensor, the input depth profile is fairly noisy. Hence, a bilateral filter is applied to obtain a smooth estimate of the depth input and surface normals corresponding to smoothed surface is evaluated in the preprocessing stage. The estimated surface normals eliminate the need for pre-calibrating the system lighting and can handle dynamic lighting environments.

Next, a lighting estimate is modeled and its corresponding parameters are evaluated using the intensity image, smoothed depth profile and estimated surface normals. Once the lighting model is determined, high quality surface is reconstructed by refining the lighting model with the depth profile and intensity image1.

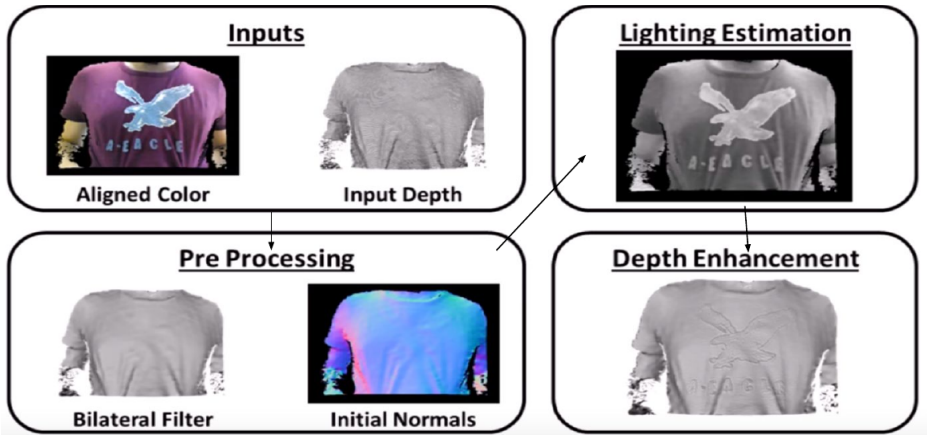


Figure 1: Pictorial Representation of the proposed framework

3 Lighting Estimate

The image decomposition aims to retrieve intrinsic properties of the image like shading and reflectance. Since, the image is taken under natural illumination where there is no single point light source, thus, the Lambertian model cannot be used to recover correct lighting scene. Hence, the image is decomposed into three components as introduced in Grosse *et al*[2]: Lambertian shading, reflectance and specularities. Thus,

$$L(i,j,\vec{n}) = \rho(i,j)S(\vec{n}) + \beta(i,j) \tag{1}$$

where $L(i,j,\vec{n})$ is the image lighting at each pixel, $S(\vec{n})$ is the shading, $\rho(i,j)$ accounts for multiple albedos and $\beta(i,j)$ accounts for local lighting variation.

The components of lighting model are recovered using single input image and depth profile starting with shading. Setting the specularity term to zero, lighting estimate corresponds to

Lambertian scene. According to Basri and Jacobs[4] and Ramamoorthi and Hanrahan[3] the irradiance of diffuse object under natural illumination is described by low order spherical harmonics. Spherical Harmonics is analogous to Fourier analysis, but on the surface of the sphere. To model the way diffuse surfaces turn light into an image, the amount of light reflected as a function of the surface normal (assuming unit albedo) is determined, for each lighting condition. These reflectance functions are produced through the analog of a convolution of the lighting function using a kernel that represents Lambert's reflection. This kernel acts as a low-pass filter with 87.5 percent of its energy in the first four components (the zero and first order harmonics). Thus, a smooth function of spherical harmonics which are linear polynomial for surface normals and independent of surface location are used to model shading:

$$S(\vec{n}) = \vec{m}^T \tilde{n} \tag{2}$$

where $\vec{m}$ is a vector of four first order spherical harmonics and $\tilde{n} = (\vec{n}, 1)^T$corresponds to the surface normals.

The intensity image I is used to determine first four spherical harmonic coefficients. Since every valid pixel is used to recover shading, an overdetermined least squares problem is used to recover the coefficients. The least square process is insensitive to high frequency changes which correspond to abrupt change in geometry.

$$\vec{m} = \arg\min_{\vec{m}} \|\vec{m}^T \tilde{n} - I\|_2^2 \tag{3}$$

Once shading $\vec{S}$ is computed, move on to recover albedos ρ by freezing the shading term. Set the fidelity term to minimize the ℓ_2 error between the proposed algorithm and input image. To prevent overfitting, a penalty term is added for the minimization problem:

$$\min_{\rho} \|\rho S(\vec{n}) - I\|_2^2 + \lambda_\rho \| \sum_{k \epsilon \mathbb{N}} \omega_k^c \omega_k^d (\rho - \rho_k)\|_2^2 \tag{4}$$

where $\mathbb{N}$ is the neighbourhood of the pixel, intensity weighting term (ω_k^c) and depth weighting term (ω_k^d) as:

$$\omega_k^c = \begin{cases} 0, & \|I_k - I\|_2^2 > \tau \\ exp\left(-\frac{\|I_k - I(i,j)\|_2^2}{2\sigma_c^2}\right), & otherwise \end{cases} \tag{5}$$

$$\omega_k^d = exp\left(-\frac{\|z_k - z(i,j)\|_2^2}{2\sigma_d^2}\right) \tag{6}$$

I and z correspond to intensity and depth respectively. σ_c and $\sigma_d$correspond to discontinuties in intensity and depth respectively.

The penalty term performs three dimensional segmentation of the image into piecewise smooth parts. Therefore, material and albedo changes are accounted for but subtle changes in the surface are smoothed. Once albedo is recovered, lighting variation β is also recovered using similar formulation to albedo. But there is an extra penalty term to limit the energy of β in order to make it consistent with the shading model.

$$\min_{\beta} \|\beta - (I - \rho S(\vec{n}))\|_2^2 + \lambda_\beta^1 \| \sum_{k \epsilon \mathbb{N}} \omega_k^c \omega_k^d (\beta - \beta_k)\|_2^2 + \lambda_\beta^2 \|\beta\|_2^2 \tag{7}$$

4 Refining the Surface

Now that the parameters of lighting model1 are evaluated, the fine details of the geometry need to be restored. The lighting model now is expressed in terms of depth - z, using the relation between surface normals ($\vec{n}$) and surface gradient (∇z).

$$\vec{n} = \frac{\left(\frac{dz}{dx}, \frac{dz}{dy}, -1\right)}{\sqrt{1 + \|\nabla z\|^2}} \tag{8}$$

Expressing the lighting model in terms of depth forces the surface to change only in the viewing direction, thus limiting the surface distortion and increasing the robustness. By fixing the lighting model parameters and allowing the surface gradient to vary, subtle details in the geometry can be recovered. The objective function $f(z)$ minimizes the difference between input intensity image and estimated shading model with two regularization terms. The first regularization term corresponds to the simple fidelity and the second regularization term ($\triangle z$)corresponds to smoothness in shading:

$$f(z) = \|L(\nabla z) - I\|_2^2 + \lambda_z^1 \|z - z_0\|_2^2 + \lambda_z^2 \|\triangle z\|_2^2 \tag{9}$$

where z_0 is the initial depth map.

The nonlinear terms in the least square is frozen and the numerical scheme1 iterates over the linear terms and, at the end of the iteration the nonlinear terms are updated. The process is repeated as long as the objective function $f(z)$ decreases.

Algorithm 1 Accelerated Surface Enhancement

Input: z_0, $\vec{m}$, ρ, β - initial surface, lighting parameters
1. **while** $f(z^{k-1})$ - $f(z^k)$>0 **do**
2. | Update $\tilde{n}^k$=$\left(\vec{n}^k, 1\right)^T$
3. | Update $L(\nabla z^k) = \rho(\vec{m}^T \tilde{n}^k) + \beta$
4. | Update z^k to be the minimizer of $f(z^k)$
5. **end**

5 Results

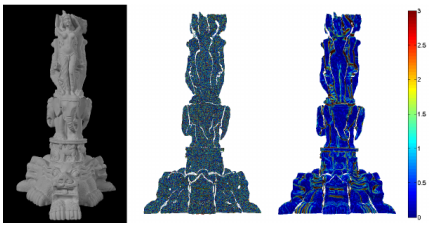
The proposed algorithm was tested on synthetic data and real data. Synthetic data was obtained from Stanford 3D repository, Blendswap repository and Smithsonian 3D archive, with the lighting environment simulated using Blender®. Each model was used to test different scenario and, Gaussian noise with zero mean and standard deviation of 1.5 to the depth map were added to the simulated models. The algorithm parameters were set to $\lambda_\rho = 0.1$, $\lambda_\beta^1 = 1$, $\lambda_\beta^2 = 1$, $\tau = 0.05$, $\sigma_c = \sqrt{0.05}$, $\sigma_d = \sqrt{50}$, $\lambda_z^1 = 0.004$, $\lambda_z^1 = 0.0075$. "Thai statue"2 tests for Lambertian object in three point lighting environment with minimal shadows. "Lincoln"3a was a Lambertian object in a complex lighting environment with multiple shadows. "C3PO"3b was a non-Lambertian object with a point light source, while "Cheese Burger"4a was a non-Lambertian multiple albedo object with three point lighting. The results from the synthetic data were compared with algorithms proposed by - Han *et al.*[6] and Wu *et al.*[1], and the corresponding median of the depth error and 90^{th} percentile of the depth error compared to ground truth were tabulated4b.

The proposed algorithm was tested on real data captured by Intel's Real-Sense RGBD sensor and the algorithm's behaviour towards the world shapes with multiple albedo objects were observed. From the figure5, it was observed that the algorithm successfully reveals the letter and

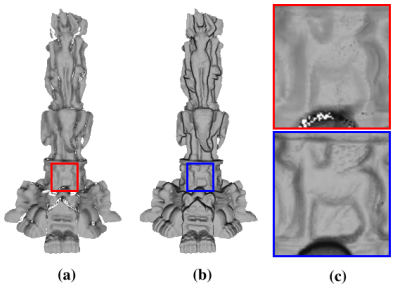
the bird on the shirt alongwith the “SF” logo, “champions” and even the stitches on the baseball cap. But, the algorithm was slightly confused by the grey “N” which was printed on the cap (does not have any thickness) while in the reconstructed image, “N” was observed to have some thickness. This texture copy artifact can be mitigated by reducing the regularization constant for albedo.

Finally, an unoptimized implementation of the algorithm was tested on Intel i7, 3.4 GHz processor with 16GB RAM and Nvidia GeForce GTX TITAN GPU. The implementation on this particular setup was observed to process10 frames per second for 640 x 480 depth profiles. The time breakdown for the entire process was tabulated.

Hence, it was observed that the proposed algorithm was significantly more robust than the previously known state-of-art methods and was able to handle real time data at 10 frames per second.

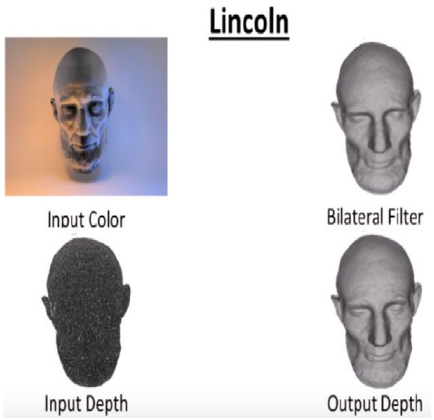


(a) Thai Statue error analysis. From left to right: Input color image. Error image of the raw depth map. Error image of the final result. Initial surface errors are reduced.

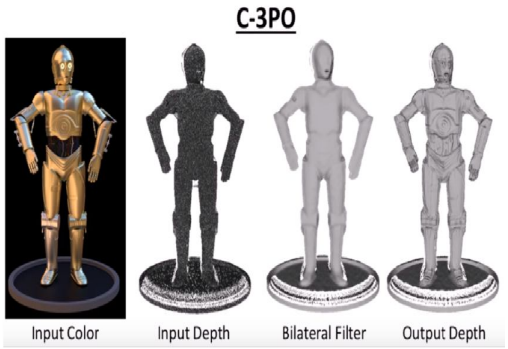


(b) Left to right: HLK reconstruction with the entire depth map normals (a)Our method reconstruction with the entire depth map normals (b)Magnification of the results is presented in (c)The proposed method yield accurate reconstruction despite the distorted shading.

Figure 2: Thai Statue

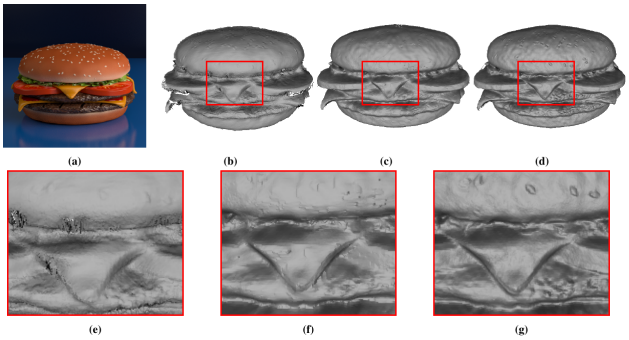


(a) Lincoln



(b) C3PO

Figure 3

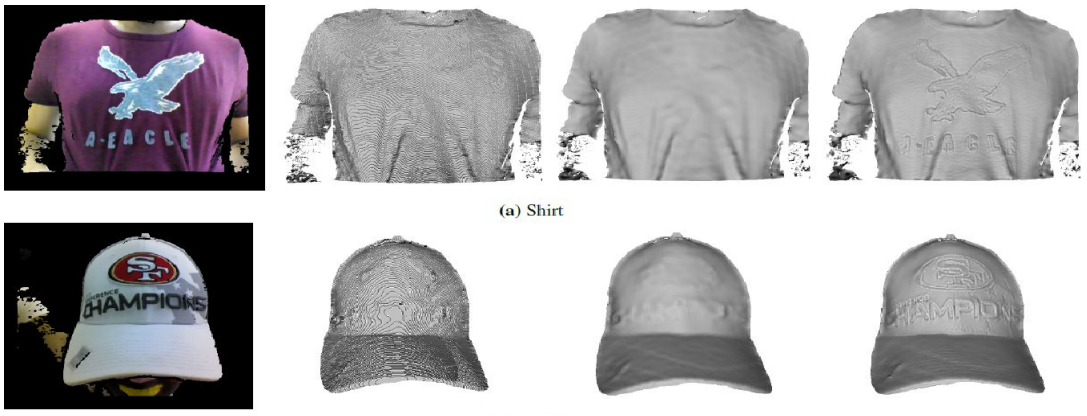


(a) Cheese Burger: (a) Color Image.Reconstruction. (e) - (g) Magnifications of HLK, WZNSIT and Ourdistinguish albedo changes. (b) HLK Reconstruction. (c) WZNSIT Reconstruction. (d) Our Method respectively. Note how the proposed framework sharply.

Section	Time
Bilateral Filter	3.8ms
Image Alignment	31.1ms
Normal Estimation	5.3ms
Lighting Recovery	40.3ms
Surface Refinement	22.6ms
Total Runtime	103.1ms

(b) Algorithm's Profiling

Figure 4



(a) Shirt

(b) Baseball Cap

Figure 5: Results of shape enhancement of real world multipleFiltering and the Proposed Method. Note how surface wrinkles and smallalbedo objects. Left to right: Color Image, Raw Depth, Bilateral surface protrusions are now visible.

References

[1] M. Niener M. Stamminger S. Izadi C. Wu, M. Zollhfer and C. Theobalt. Real-time shading-based refinement for consumer depth cameras. in acm transactions on graphics (proceedings of siggraph asia 2014), volume 33. 2014. 5

[2] E. H. Adelson R. Grosse, M. K. Johnson and W. T. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. in international conference on computer vision, pages 2335–2342. 2009. 3

[3] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. in proceedings of the 28th annual conference on computer graphics and interactive techniques pages 497–500. 2001. 3

[4] R.Basri and D.W. Jacobs. Lambertian reflectance and linear subspaces. ieee transactions on pattern analysis and machine intelligence, 25(2):218–233. 2003. 3

[5] Aaron Wetzler Ron Kimmel Alfred M. Bruckstein Roy Or El, Guy Rosman. Rgbd-fusion: Real-time high precision depth recovery. 2015. (document)

[6] J. Y. Lee Y. Han and I. S. Kweon. High quality shape from a single rgb-d image under uncalibrated natural illumina- tion. in ieee international conference on computer vision, pages 1617–1624. 2013. 5