MRF and DP Based Specular Surface Reconstruction

Ravindra Redddy K
Center for Visual Information Technology
International Institute of Information Technology
Hyderabad, India - 500032
Email: rrk.ravindra@gmail.com

Anoop Namboodiri
Center for Visual Information Technology
International Institute of Information Technology
Hyderabad, India - 500032
Email: anoop@iiit.ac.in

***Abstract*—This paper addresses the problem of reconstruction of specular surfaces using a combination of Dynamic Programming and Markov Random Fields formulation. Unlike traditional methods that require the exact position of environment points to be known, our method requires only the relative position of the environment points to be known for computing approximate normals and infer shape from them. We present an approach which estimates the depth from dynamic programming routine and MRF stereo matching and use MRF optimization to fuse the results to get the robust estimate of shape. We used smooth color gradient image as our environment texture so that shape can be recovered using just a single shot. We evaluate our method using synthetic experiments on 3D models like Stanford bunny and show the real experiment results on golden statue and silver coated statue.**

I. Introduction

Reconstruction of the specular (mirror-like) surfaces is a challenging problem and drawn considerable attention in recent years. As, the observed images of a specular object is a function of the object shape as well as the exact nature of environment that surrounds the object. Traditional methods of specular object reconstruction follows one of the two approaches: i) Exert complete control over the environment (coded environments) and recover the shape of the object with very less restrictions on the object shape except self reflections, and ii) Assume that the environment is highly unconstrained and use assumed object properties (integrability and smoothness) to disambiguate possible hypotheses that arise from the reconstruction. We explore a method that assumes far lesser control of the environment while allowing arbitrary object shapes (except self reflections).

In this paper, we attempt to reconstruct the surface using optimization framework on matching two stereo images. This method mainly comprises of three steps, Firstly, the images which are captured in a artificial setup with very minimal constraints on the perfect calibration of the environment and camera capturing mode and given as input for pre-processing. The pre-processing stage involves, extraction of approximate normals at each imaged object points. Secondly we apply two schemes for extracting depth from stereo, one being the dynamic programming approach and the second by formulating the problem into optimization problem and use Loopy Belief Propagation (LBP). Finally, a simple LBP approach to fuse the results of two algorithms to get the optimized result. We evaluate our method using synthetic experiments of the rigid objects and show our results of reconstruction of ganesha statue captured in a controlled environment.

This paper contains two contributions. First we develop an experimental setup which can be used to extract object features and thus facilitating us adapt the methods of lambertian surface reconstruction for specular object reconstruction. Secondly, we propose an MRF based integration framework which takes the depth estimates from various methods and gives the robust estimate of depth of the object.

The remainder of the paper was organized as follows, literature review is discussed in Sec. II. In Sec. III we present our reconstruction algorithm. Sec. IV contains discussion on our method and finally, we present our results in Sec. V for both synthetic and real objects.

II. Related Work

Over the past few years much progress has been made towards solving the specular object reconstruction problem. Early works include study of specularities [1]–[3] and use of calibrated patterns [4], [5] to estimate the normals. Tirani *et al.* [5], integrated the normals around a seed point and using a global self-coherence measure to estimate the correct depth for the seed point.

Geometrical methods like multi-view, stereopsis and single view techniques gives dense reconstruction but it requires precise knowledge of environment points. In [6], Nehab *et al.* used a calibrated setup to find the specularities and in turn normals at each point They used a dynamic programming for matching sequences. In [7], using light-path triangulation Kutulakos *et al.* showed that it is practically impossible to reconstruct the surface if the object self reflects more than once (reflects more than twice).

Higher order differential geometry of the surface is explored by [8], [9]. Adato *et al.* [10] exploited the specular flow and formulated the reconstruction into linear PDEs. Aswin C. *et al.* [11], [12] proposed a method of finding image invariants for smooth specular objects and in turn sparse reflection correspondences. With a finite motion of object, fixed camera and uncalibrated environment.

Estimating stereo correspondences using Markov Random Fields (MRF) optimization for finding disparities has been extensively used for lambertian surfaces. Most modern approaches frame the problem as inference on a Markov random field and utilize global optimization techniques to estimate the depth/ disparity at each object pixel in the image. Various optimization techniques like, Iterative Conditional Methods, graph-cuts, loopy belief propagation using sum-product and max-product were tried in [13]–[16] and compared [14] by various

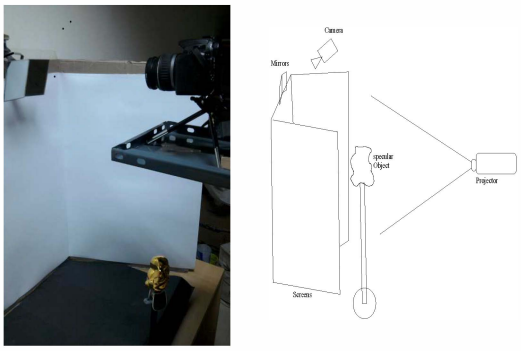


Fig. 1. Physical Setup

researches. Efficient way of giving costs has been discussed by [17], [18]. Birchfield and Tomasi [19] proposed a measure which is insensitive to image sampling. This measure was used for stereo matching in [14]. Jian Sun *et al.* [20] use belief propagation to solve dense stereo correspondence problem. Hershmuller *et al.* [21] proposed semi global matching which extends polynomial time 1D scan-line methods to propagate information along 16 orientations. This reduces streaking artifacts and improve accuracy compared to traditional methods.

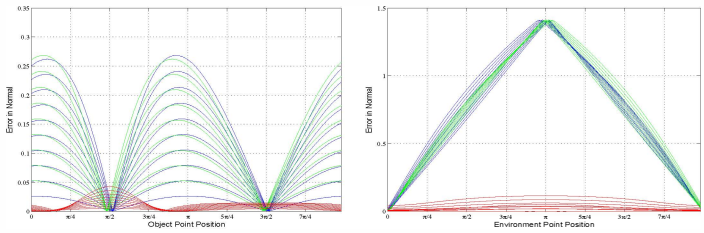
III. Reconstruction Method

The primary goal of our approach is to relax the constraints on the environment without having to impose any additional constraints on the object shape. Multiview geometry provides us with the additional information that is required to achieve this. To avoid explicit and accurate calibration of the environment, we assume that the normals that are computed for the object surface are approximate in nature. We introduce the concept of using approximate normals as the features of the specular object.

A. Experimental Setup and Approximate Normals Computation

Lambertian surface reconstruction techniques extract features of the object based on surface textures. For specular surfaces these features were not directly visible as we see the environment in the object instead of the object texture. Hence we need to extract texture independent object features that carry shape information. We use approximate normals as the surface feature.

1) Experimental Setup: A typical stereo setup consists of a stereo pair of cameras observing a specular object, placed at the center of a controlled environment. We used three adjacent panels placed at an angle of 120° as our environment screen (See Fig. 1). Two adjustable mirrors were mounted on top of our environment panels. These mirrors were adjusted such that, the camera placed in front of the mirrors observes two reflections (one in each mirror) of the object. The split mirror and camera system behave as a typical stereo camera system. This kind of setup avoids synchronization problem between camera while capturing moving objects. Finally a smoothly varying color pattern is projected on the environmental panel while capturing the stereo image data. In the rest of the paper we treat the virtual stereo cameras as regular stereo camera pair.



(a) (b)

Fig. 2. (2a) Net error in normal estimation due to object point approximation. (2b) Net error in normal estimation due to environment point approximation.

2) Approximate Normal Computation: Consider an environment point P_e is reflected by the object point P_o and imaged by the camera with optical center P_c. Then normal is calculated as $\hat{n}_{actual} = \frac{\hat{i}+\hat{r}}{\|\hat{i}+\hat{r}\|}$, where $\hat{i}$ is an unit incident ray pointing towards the camera and $\hat{r}$ is an unit reflection vector in the direction of $P_e - P_o$. In *approximate normal* estimation, environment center is considered as P_o. We normalize the color observed on the object surface and consider P_e to be a point on the normalized color sphere, which is scaled by a value equal to the distance of the panels from the center. Generally the observed color is influenced surrounding light and tint of the object. Hence P_e is a rigid transform of the actual environment point. In the overall transformation of the environment, the relative positions of the environment points were preserved, and similar transformations happen for both stereo images. As a result, net error in normal estimation is less. This enables us to use these *approximate normals* as object features for stereo matching.

Each curve in Fig. 2a correspond to the euclidean error in normal computation, when P_o is approximated as P'_o a distance rho and at an angle θ to the horizontal. Error curves for 10 ρ values were plotted for each θ. Green, blue and red curves represent normal estimation error for left, right stereo images and net difference respectively. Similarly Fig. 2b illustrates normal estimation error for environment point approximation. Here θ and ρ correspond to rotation and translation of estimated environment point. Note that even though the error in normal estimation is high individually, due to the consistancy in errors net error is minimal.

As a convention, normal is considered as the angular bisector of incident and reflected rays. Rotation of environment point induces normal ambiguity for the normals close to 90° to any of the camera's optical lines. Due to rotation of P_e, for one camera estimated normal points outwards while for the other it points inside. To overcome this ambiguity we choose between, computed reflected ray and its reflection based on the minimum euclidean error between stereo pair.

B. Dynamic Programming approach

The approximate normals of the surface were chosen as the features of the surface. Even though the computed normals were not accurate, the continuous pattern of normals carry shape information. we can obtain the shape of the surface by matching the left and right sequences as the variation of normals in both the sequences have similar information. Also, the stereo setup gives the flexibility of having consistency of

errors in the left and right sequences which reduces the overall error while matching. Thus using dynamic programming we try to match the pattern present in the normal sequences to obtain the shape of the object.

Given rectified left and right images, the dynamic programming approach exploits the ordering constraint of the images to obtain the disparity and in turn depths of the surfaces. The normal sequences corresponding to the left and right scan lines were matched using the traditional dynamic programming method. If $\Gamma_1 = \{(n_i, p_i); i \in [1, m_1]\}$ and $\Gamma_2 = \{(n_j, p_j); j \in [1, m_2]\}$, n_i, n_j are the two normal sequences, the cost of matching x_i of first image matching with x_j of the second image is defined as follows

$$C(x_i, x_j) = cos^{-1}(n_i \cdot n_j) + \alpha(g(n_i) - g(n_j))^2 \quad (1)$$

$$where, g(n_i) = cos^{-1}(n_{i-1} \cdot n_i) - cos^{-1}(n_i \cdot n_{i+1}) \quad (2)$$

and $\alpha \in [0, 1]$. Eq. 1 has two terms, the first term is simply angle between two matching normals, and the second term signifies matching of change in normals. It is similar to matching curvature at a point. α is a weighing factor which controls how strongly we want the curvature to be matched.

As we know there are three factors we need to consider while formulating dynamic programming based matching: 1) A pixel or normal finds a match in the left image. 2) A normal is visible in left while it is occluded in the right image. during matching process we need to skip left image pixel (SLP) and 3) Similarity, if a normal is occluded in left image, we need to skip right image pixel (SRP). The costs of matching for SLP and SRP were chosen to be the maximum allowed matching cost.

C. Loopy Belief Propagation

As the image is considered as the graph defined by four-connected image grid. Stereo matching is seen as a problem of finding disparities at each point on the left (reference) image. Given two rectified stereo pair of images, each pixel of the left image is given as label and corresponding costs are defined . Then the stereo correspondence problem is nothing but finding set of labels which yields to minimum energy in the energy formulation.

For each pixel p a label l_p has to be assigned. The collection of all pixel-label assignments is denoted by l, the number of pixels is N and the number of labels is M. The energy function E, which can also be viewed as the log likelihood of the posterior distribution of an MRF, is composed of a data energy E_d and a smoothness energy E_s. *i.e.*

$$E = E_d + \lambda E_s. \quad (3)$$

The data energy E_d is simply the sum of a set of per-pixel data costs $d_p(l)$ $E_d = \sum_p d_p(l_p)$. where we consider distance between two normals as data costs. *i.e.* when a p_i of left image is matched with p_j of the right image, and the corresponding estimated normals at these points are n_i and n_j then $|\vec{n_i} - \vec{n_j}|$ is the data cost at p_i for the label $|i - j|$.

Smoothness cost at each p in the 2D image grid can also be written in terms of its coordinates $p = (i, j)$. We use the standard 4-connected neighborhood system, so that the smoothness energy is the sum of spatially varying horizontal and vertical nearest neighbor smoothness costs, $V_{pq}(l_p, l_q)$, where if $p = (i, j)$ and $q = (s, t)$, then $|i - s| + |j - t| = 1$. If we let $\mathcal{N}$ denote the set of of all such neighboring pixel pairs, the smoothness energy is $E_s = \sum_{\{p,q\} \in \mathcal{N}} w_{pq} \cdot V_{pq}(l_p, l_q)$. Note that the notation $\{p, q\}$ stands for an unordered set, that is, the sum is over unordered pairs of neighboring pixels. $w_{p,q}$ is a spatially varying weights to handle discontinuities caused either due to visibility of only portion of the specular surfaces, self reflections or discontinuities in the object itself. $V(l_p, l_q)$ is a non-decreasing function of the label difference $V(l_p, l_q) = \min(|l_p - l_q|^k, V_{\max})$.

The convergence of the algorithm my not occur in all cases. The energy function might reach a point where it fluctuates around. To handle such cases we keep track of lowest energy so far and if it does not changes over few iterations, the iterations can be stopped and use the disparity map at that stage as our result.

D. Disparity Fusion

Disparity/Range information can be obtained using many different routines. It is often necessary to combine these results to obtain a robust estimate of range data. In our case we used only two approaches MRF and DP. We use one more LBP formulation for fusing the disparity measures, from two approaches. This method is similar to regularization of data. In regularization, the data term is only a function per-pixel difference of label and data of single image/result, where as in this method, the data term is a weighted sum of per-pixel difference of label and data of each result. Mathematically, the per-pixel data cost is defined as

$$V_d = |l_p - l_{mrf}|^2 + \alpha |l_p - l_{dp}|^2 \quad (4)$$

where, l_{mrf} and l_{dp} are the labels at p from results of MRF and DP routines respectively and α is a weighing factor which controls how much weight has to be given for each method. The smoothness prior is a typical smoothening function used in vision algorithms. *i.e.* the sum of the absolute difference with its neighbors. $V_s = w(p, q) * \sum_{p \in N(q)} |p - q|$ where $w(p, q)$ is a variable per-pairing weights which is there to handle discontinuities and boundary conditions.

Similar to the basic MRF formulation (Eq. 3) the λ is used as the weighted factor for balancing data term and smoothness term. To avoid instability of the function due to noisy inputs, we apply clipping of the smoothness term. The result of this output is the robust estimate of disparities at each pixel

The primary advantage of our algorithm is that, even though one optimization routine settles at local minimum, the information from the other method pushes the solution towards the actual solution.

IV. Discussion

Dynamic programming approach exploits the ordering constraint. It does not see the depth estimates of neighboring pixels. As a result there can be inconsistency of depth estimation among scan lines. Where as, MRF does not enforce hard ordering constraint but ensures that the smoothness along the scan lines as well as between the scan lines. MRF algorithms has the issue of leaking at the edges, settling a local minimum

TABLE I. MEAN SQUARE ERROR OF VARIOUS SYNTHETIC OBJECTS

Object	 Bunny	 Budha	 Venus	 Hebe
MSE DP	0.005489	0.003424	0.003543	0.005624
MSE MRF	0.006512	0.004387	0.003942	0.005406
MSE Fused	0.003916	0.003172	0.003362	0.004694

etc. The issues in one approach may not be present in other. Hence, combining these two approaches has the advantage of obtaining robust estimate.

The error in depth estimation is less at the regions with high curvatures, as the matching of normals suffer less in error due to sub-pixel matching. Where as at flatter regions, DP suffers from continuous matching problem and in MRF smoothness prior dominates resulting a flat estimation. Unlike other methods which fail at higher curvature and self reflection, our method favors them. Since, similar normal estimation is done in both the stereo images at self reflection regions, we can match them and get depth information. This is evident from real results fig. 4e, where, the region under the belly of ganesha is reconstructed properly. MRF locks on the regions of high curvatures and self reflection regions and propagate the information to flat regions to get good depth estimation. Hence our method is highly suitable for surfaces with both high and low curvatures like Stanford Bunny.

We will now look into qualitative and quantitative results of our experiments on images from synthetic and real objects.

V. RESULTS

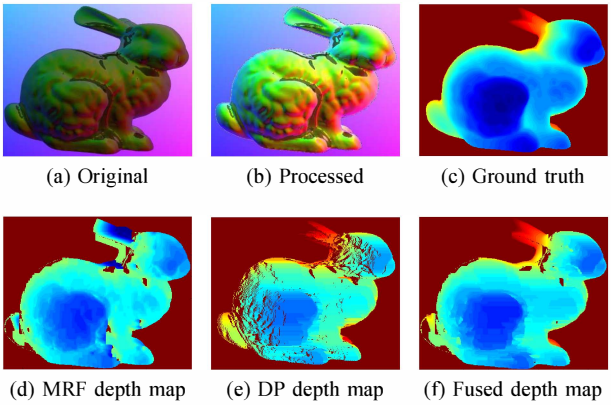
For synthetic images we used 3D polygonal models as objects. We show the results of reconstruction of golden idol captured using a relatively simple setup as shown in Fig. 1.

A. Synthetic Results

In the synthetic experimental setup we used a cube as the environment with each of it faces pasted with the sides of the RGB color cube. The specular object is placed at the center, which was texture mapped with a polished bronze metal texture. The center of the cube is considered as our origin. The two stereo cameras sit at a distance capture the image of their object. This setup was synthesized in POV-Ray 3.7.

The environment is considered as sphere of approximately same distance and the colors are mapped on to the sphere. These are considered as the environment points for estimating approximate normals. The object is taken to be a point at the object and estimates the normal as described in III-A. The approximate normal field thus generated is used for estimation of disparity using the proposed approach.

We have conducted our experiment on 3D models of Stanform Bunny, laughing Budha, Venus de Milo and Hebe (a Greek godess). The object is scaled to one unit and is placed in an environment cube with dimensions $10 \times 10 \times 10$ unit^3. The two stereo cameras were placed at a distance of 3 units away



(a) Original (b) Processed (c) Ground truth

(d) MRF depth map (e) DP depth map (f) Fused depth map

Fig. 3. Synthetic Results of Stanford Bunny: (3a) original left image, (3b) pre-processing image, (3c) ground truth depthmap and (3d), (3e) and (3f) are depth maps of MRF, Dynamic Programming and Fused algorithms.

from the object. The ground truth depthmaps were extracted using POV-Raytracer. Table I shows the Mean squared errors (MSE) between depth maps of DP, MRF and Fused algorithms with the ground truth depth maps. MSE of Fused results is lesser than MSEs of DP and MRF results.

Figure (3e), (3d) and (3f) shows the depth maps of MRF, DP and integration algorithms. One can observe that the fused results does not have the flattening problem of DP results and also corrected the wrong estimate of MRF depths near the ears of bunny model. Quantitative results show the improvement in depth estimate in integrated algorithm.

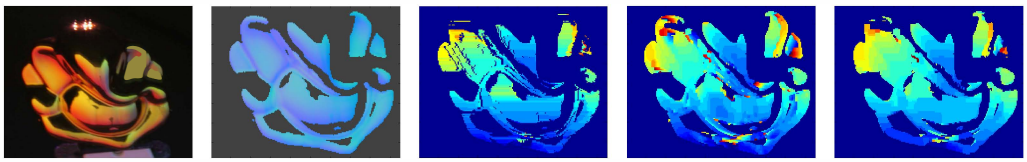
B. Real Results

In our setup, we calibrate camera-mirror setup and camera-projector system. As explained in III-A2, the camera mirror setup simulates two virtual stereo camera system. We calibrate this stereo camera setup using checkerboard patterns and rectify two stereo images. In order to correct the differences in color projected and color observed, we adopt camera - projector calibration similar to Tirani *et al.* [5].

The image captured contains colors that are influenced by factors like surrounding lighting conditions, object's reflectance properties and the tint of the surface. To minimize interferences by other light sources, we conduct the experiment in a dark room. Since the capturing device in general adds noise to the image, the captured image is filtered using median filter. As the input pattern is a uniform gradient, smoothing ideally will not change the non-noise pixel.

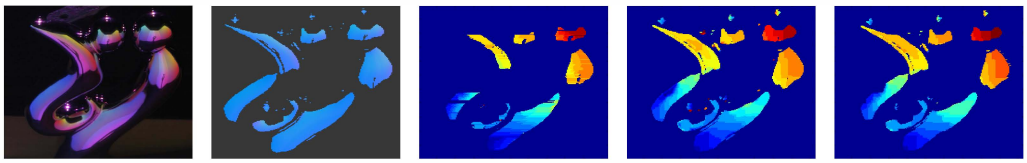
We use the texture of three sides of the RGB color cube as our projected pattern. We adjust the color positions such that the colors are uniformly varying with resepect to the object. The pattern projected on the three panel screen must be smooth gradient with respect to the object. In general, while projecting on the screen, the brightness on the center panel will be high compared to the side panels. We adjust the brightness of the pattern by projecting single color such that with respect to the object all the panels has equal brightness.

We conducted experiment on two real objects, (i) a polished Ganesha (Hindu god) idol with golden color coated on it and (ii) a steel coated statue. The object was placed on small



(a) Original (b) Normal Map (c) DP depth map (d) MRF depth map (e) Fused depth map

Fig. 4. Real Experiment on Ganesh Statue: (4a) Golden tinted ganesha statue in the three panel environment setup, (4b) is estimated approximate normal map. (4c), (4d) and (4e) are the depth estimates using Dynamic Programming, MRF and Fused algorithms respectively.



(a) Original (b) Normal Map (c) DP depth map (d) MRF depth map (e) Fused depth map

Fig. 5. Real Experiment on steal coated statue: (5a) Steel tinted statue in three panel environment setup, (5b) is estimated approximate normal map, (5c), (5d) and (5e) are the depth estimates using Dynamic Programming, MRF and Fused algorithms respectively.

platform which is raised to the center of the environment panels. We used a Canon 350D DSLR camera which is mounted above the environment panel facing the split mirrors as shown in Fig. 1.

Figure 4 and 5 illustrates the behavior of our method on real object surfaces. The object is slightly tilted backwards for a better view as it is seen from top. Range map is plotted using matlab with jet colormap (here, red is far and blue is near). In fig. 4c and fig. 5c, we can observe at flat regions, the dynamic programming method has the problem of matching continuous lines and result in flat regions. Also we can observe the discontinuities between the scan lines. In fig. 4e and 5e these discontinuities were reduced to a good extent.

VI. Conclusion

In this paper we presented a simple experimental setup which can used for estimating a dense normalmap of approximate normals. And our method has very less constraints on the accuracy of calibration. A two stage loopy belief propagation along with dynamic programming based stereo matching can recover shape of the object. This method is more accurate on surfaces with larger curvatures. Future work involves incorporating other methods like normal integration for finding range data into out current frame work.

References

[1] A. Blake and G. Brelstaff, "Geometry from specularities," in *ICCV88*, 1988, pp. 394–403.

[2] A. Zisserman, P. Giblin, and A. Blake, "The information available to a moving observer from specularities," *IVC*, vol. 7, no. 1, pp. 38–42, 1989.

[3] J. Zheng and A. Murata, "Acquiring a complete 3d model from specular motion under the illumination of circular-shaped light sources," *PAMI*, vol. 22, no. 8, pp. 913–920, 2000.

[4] T. Binford and G. Healey, "Local shape from specularity," in *In Proc. ICCV*, 1987, pp. 151–160.

[5] M. Tarini, H. P. A. Lensch, M. Goesele, and H. peter Seidel, "3d acquisition of mirroring objects," Graphical Models, Tech. Rep., 2003.

[6] D. Nehab, T. Weyrich, and S. Rusinkiewicz, "Dense 3d reconstruction from specularity consistency," in *CVPR08*, 2008, pp. 1–8.

[7] K. N. Kutulakos and E. Steger, "A theory of refractive and specular 3d shape by light-path triangulation," *IJCV*, vol. 76, no. 1, pp. 13–29, 2008.

[8] S. Savarese and P. Perona, "Local analysis for 3d reconstruction of specular surfaces," in *CVPR01*, 2001, pp. II:738–745.

[9] Y. Ding, J. Yu, and P. Sturm, "Recovering specular surfaces using curved line images," 2009, pp. 2326–2333.

[10] G. Canas, Y. Vasilyev, Y. Adato, T. Zickler, S. Gortler, and O. Ben Shahar, "A linear formulation of shape from specular flow," in *ICCV09*, 2009, pp. 191–198.

[11] A. C. Sankaranarayanan, A. Veeraraghavan, O. Tuzel, and A. Agrawal, "Image invariants for smooth reflective surfaces," in *In Proc. ECCV'10: Part II*, 2010, pp. 237–250.

[12] A. C. Sankaranarayanan, A. Veeraraghavan, O. Tuzel, and A. K. Agrawal, "Specular surface reconstruction from sparse reflection correspondences." in *CVPR*. IEEE, 2010, pp. 1245–1252.

[13] T.-P. Wu, K.-L. Tang, C.-K. Tang, and T.-T. Wong, "Dense photometric stereo: A markov random field approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1830–1846, nov 2006.

[14] R. Szeliski and e. a. Zabih, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, jun 2008.

[15] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *Int. J. Comput. Vision*, vol. 70, no. 1, pp. 41–54, oct 2006.

[16] M. Lhuillier and L. Quan, "Surface reconstruction by integrating 3d and 2d data of multiple views," in *In Proc. ICCV '03 - Volume 2*, 2003, pp. 1313–.

[17] R. R. Paulsen, J. A. Baerentzen, and R. Larsen, "Markov random field surface reconstruction," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 4, pp. 636–646, jul 2010.

[18] P. Li, R. Klein Gunnewiek, and P. H. With, "Scene reconstruction using mrf optimization with image content adaptive energy functions," in *In Proc. ACIVS '08*, 2008, pp. 872–882.

[19] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 4, pp. 401–406, apr 1998.

[20] J. Sun, N.-N. Zheng, and H.-Y. Shum, "Stereo matching using belief propagation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 7, pp. 787–800, jul 2003.

[21] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *In Proc. CVPR '05, volume 2*, 2005, pp. 807–814.