Sparse Modeling of Shape from Structured Light

Guy Rosman   Anastasia Dubrovina   Ron Kimmel

Dept. of Computer Science

Technion - IIT

Haifa 32000 Israel

rosman@cs.technion.ac.il

Abstract

Structured light depth reconstruction is among the most commonly used methods for 3D data acquisition. Yet, in most structured light methods, modeling of the acquired scene is partial, and executed separately from the decoding phase. Here, we bridge this gap by viewing the reconstruction process via a probabilistic model combining illumination and shape. Specifically, an alternating minimization algorithm for structured light reconstruction is presented, incorporating a sparsity-based prior for the local surface model. Integrating the 3D surface prior into a probabilistic view of the reconstruction phase results in a robust estimation of the scene depth.

We formulate and minimize reconstruction error and demonstrate performance of the algorithm on data from a structured light scanner. The results demonstrate the robustness of our algorithm to scanning artifacts under low SNR conditions and object motion.

1. Introduction

Structured light and active illumination range scanners have become an important tool for scene understanding [12, 19, 16], robotics [21, 10, 26], object modeling [8, 2], indoor scene mapping [22], and human computer interaction [32], among other tasks. The scanner usually consists of a calibrated camera-projector pair; where coded light patterns emitted by the projector are acquired by the camera and allow robust triangulation and depth reconstruction. For a review of existing structured light techniques see, for example, [29].

Many of the techniques used to reconstruct 3D depth via structured light incorporate ad-hoc assumptions on the scene structure and the 3D imaging process. These include, for instance, smoothness of the acquired surface [40, 17], or temporal objects behavior [11, 40, 17]. Yet, modeling these assumptions in a more complete way is crucial when the captured illumination patterns are of low SNR, due to long

scanning range and short camera exposure times. Furthermore, such assumptions can help when dealing with motion artifacts, where some of the captured images are subject to abrupt intensity changes due to motion of depth discontinuities or albedo boundaries. Failing to model the imaging process in a realistic manner may lead to outliers in the reconstructed depth image, as is often observed in structured light scanners.

Here, we improve upon results obtained by structured light based scanners [25, 29], especially in face of challenging illumination conditions, by providing strong priors for the imaging model and surface shape. Moreover, while strong shape priors are utilized for range image correction, i.e., surface denoising and completion, the approach we suggest incorporates shape and illumination priors into the reconstruction itself, giving us a principled approach of combining a strong surface prior and probabilistic understanding of the acquisition process. Here we introduce a patch-based image similarity prior, similar to those successfully utilized for image and surface processing [1, 7, 35, 31, 39].

2. Regularized Structured Light Model

In shape from structured light, one is attempting to reconstruct the geometric structure of the scene, by illuminating the scene with a set of projected patterns $\{I_P^{(i)}\}_{i=1}^N$, where N is the number of patterns, and taking a set of images $\{I_C^{(i)}\}_{i=1}^N$ of the scene using a camera. We denote the optical centers of the camera and projector by points $\mathcal{C}$ and $\mathcal{P}$ respectively. The overall setup is shown in Figure 1. In our formulation, we denote the estimated range image as $z(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^2$ denotes the (two-dimensional) camera image coordinates.

We assume a Lambertian surface model for objects in the scene, and a projector emitting directional light in a temporal sequence of patterns. The main source of imaging noise is assumed to be the sensor. The lighting conditions we deal with are such that the photon count per image sensor pixel is

high enough so that the image noise model is approximately Gaussian, yet the signal is weak enough so that correctly decoding the coded light patterns poses a challenge. This is the typical scenario in real structured light systems with temporal multiplexed code, aimed for example at capturing motion, thus requiring short exposure intervals.
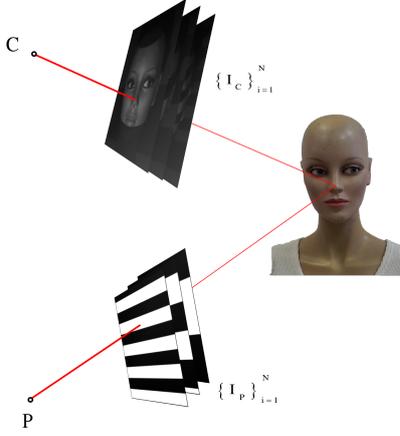
Figure 1. An example of a structured light system setup.

Assuming a global illumination component and a projector illumination component, we can model every pixel's intensity at each frame i as

$$I_C^{(i)}(\mathbf{x}) = a(\mathbf{x}) I_P\left(\Pi_{z,i}(\mathbf{x})\right) + b(\mathbf{x}) + n(\mathbf{x}), \quad (1)$$
$$n(\mathbf{x}) \sim N\left(0, \sigma_I^2\right).$$

$a(\mathbf{x})$ and $b(\mathbf{x})$ are pixel-wise coefficients that depend on the global illumination of the scene, the surface properties, object albedo, projector properties, and so forth. $\Pi_{z,i}(\mathbf{x})$ denotes the depth-dependent transformation from pixel $\mathbf{x}$ to a corresponding pixel on the projector image. It is obtained by backprojecting the camera ray to depth z and projecting the point into the projector optical center. $n(\mathbf{x})$ is the pixel noise, assumed to be additive white Gaussian noise, independent and identically distributed (i.i.d.).

We wish to formulate and maximize a probability function of the depth given the known camera images and projected textures. By applying Bayes' rule, and removing constant factors we obtain

$$\begin{aligned}
z &= \underset{z,a,b}{\arg\max} \, P\left(z, a, b | I_P, I_C\right) \\
&= \underset{z,a,b}{\arg\max} \, \frac{P\left(z, a, b, I_P, I_C\right)}{P\left(I_P, I_C\right)} \\
&= \underset{z,a,b}{\arg\max} \, \frac{P\left(I_P, I_C, a, b | z\right) P\left(z\right)}{P\left(I_P, I_C\right)} \\
&= \underset{z,a,b}{\arg\max} \, P\left(I_P, I_C, a, b | z\right) P\left(z\right) \\
&= \underset{z,a,b}{\arg\min} \, -\log P\left(I_P, I_C, a, b | z\right) - \log P\left(z\right). \quad (2)
\end{aligned}$$

We incorporate the maximum-likelihood choice of a, b into $P(I_P^{(i)}, I_C^{(i)} | z)$, minimizing the negative log-probability over a and b,

$$\min_{z,a,b} \left[-\log\left(P\left(I_P, I_C, a, b | z\right)\right)\right] = \quad (3)$$
$$\min_z \left(\min_{a,b} \left[\sum \frac{(a(\mathbf{x}) I_P\left(\Pi_{z,i}(\mathbf{x})\right) + b(\mathbf{x}) - I_C\left(\mathbf{x}\right))^2}{\sigma_I^2}\right]\right).$$

The optimal values of a and b for this least-squares fitting problem are given in analytical form by solving the normal equations using I_C, I_P at points $\mathbf{x}, \Pi_{z,i}(\mathbf{x})$ respectively,

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \mu_{PP} & \mu_P \\ \mu_P & N \end{pmatrix}^{-1} \begin{pmatrix} \mu_{CP} \\ \mu_C \end{pmatrix}, \quad (4)$$
$$\mu_P = \sum I_P\left(\Pi_{z,i}(\mathbf{x})\right), \; \mu_C = \sum I_C\left(\mathbf{x}\right),$$
$$\mu_{CP} = \sum I_C\left(\mathbf{x}\right) I_P\left(\Pi_{z,i}(\mathbf{x})\right),$$
$$\mu_{PP} = \sum I_P\left(\Pi_{z,i}(\mathbf{x})\right)^2.$$

In order to obtain an efficient algorithm for computing and optimizing photoconsistency in the structured light case, we note that we can incorporate the computation of the maximum-likelihood expressions for a, b into a plane-sweep operation [5] when seeking the optimum value of z.

Inserting the optimal a, b as a function of z and noting the conditional independence (given z) of neighboring pixel values $I_C(x)$, $I_P\left(\Pi_{z,i}(\mathbf{x})\right)$ provides us with a functional to minimize with respect to $z(\mathbf{x})$, similar to [34],

$$\underset{z}{\arg\min} \int_{\mathbf{x}} \min_{a,b} \left(-\log\left(P\left(I_P, I_C, a, b | z\right)\right)\right) d\mathbf{x} + \psi\left(z\right) =$$
$$\underset{z}{\arg\min} \int_{\mathbf{x}} \rho_{SL}\left(z; I_C, I_P, \mathbf{x}\right) dx + \psi\left(z\right). \quad (5)$$

The expression $\rho_{SL}\left(z; I_C, I_P, \mathbf{x}\right)$ denotes a penalty for the photoconsistency assumption. This term is often optimized per pixel by several steps, including binarization of the code letters, decoding of the code, and depth reconstruction. These separate steps, however (for any specific code) are sub-optimal, even if efficient to compute.

The term $\psi\left(z\right)$ denotes our choice for approximating the negative log-probability prior for the surface shape, $-\log P(z)$. There are several possible choices of surface shape priors. These can incorporate either smoothness assumptions and more elaborate geometric priors, assumptions on local shape of patches on surfaces, or reasoning on natural depth image statistics [38]. We now describe two such possible regularization priors for depth images.

Total-Variation regularization The minimum area [4] and *total-variation* [28] (TV) priors, and related smoothness measures have been suggested in several forms for regularization of range images [23] and surface reconstruction [15, 33]. TV regularization for structured light can be ex-

pressed as

$$\operatorname*{argmin}_{z} \int_{\mathbf{x}} \rho_{SL}\left(z; I_C, I_P, \mathbf{x}\right) + \tilde{c}\|\nabla z\|dx, \qquad (6)$$

where $\|\nabla z\|$ is the total variation of the range image. This form of regularization is strongly related to MRF-based structured light [34].

Patch-based Priors for Structured Light Another possibility for modeling range images involves assuming a local model for each patch of the surface. Regularizing the surface then expresses itself via the parameters of this model. This includes modelling via polynomials or similar functions, leading to the moving-least-squares [18] approach, or expressing the patch via a functional basis with sparse coefficients, leading to sparsity-based regularization. Priors for depth images based on patch-estimators are described, for example, in [31, 13, 20, 35].

In our case, we assume that the depth image can be locally viewed as a sparse combination of basis functions. We note by $\tilde{\psi}(\cdot)$ our prior for surface patches. This leads to a patch-based regularizer of the reconstruction,

$$\operatorname*{argmin}_{z} \int_{\mathbf{x}} \rho_{SL}\left(z; I_C, I_P, \mathbf{x}\right) d\mathbf{x} + \tilde{c}_1 \sum_{j} \tilde{\psi}\left(P_j z\right), \quad (7)$$

where $P_i z$ denotes extraction of a small neighborhood i from the surface z. For example, for an L_1-sparse representation prior, Equation 5 becomes

$$\operatorname*{argmin}_{z, \alpha_j} \int_{\mathbf{x}} \rho_{SL}\left(z; I_C, I_P, \mathbf{x}\right) + \qquad (8)$$
$$c_1 \sum_{j} \|P_i z - D\alpha_j\|^2 + \lambda\|\alpha_j\|_1,$$

where D denotes a dictionary for depth image patches, where P_j denotes a matrix extracting block j from the image in column-stacked notation, α_j denotes the representation of patch $P_j z$ in that dictionary. The vector α concatenates all local representations α_j of all patches in the depth image.

3. Alternating Minimization Algorithm for Regularized Structured Light

We assume the coded light pattern can be reconstructed by minimizing per-pixel the decoding error function $\rho_{SL}(\mathbf{x}, I_C, I_P; z)$. While this reconstruction is usually obtained by binarization and decoding of the time-multiplexed code, we view it as a photoconsistency term between the structured light patterns and the resulting camera image intensities [24], while estimating the illumination conditions. Note that this function depends only on the depth value and camera intensities per pixel. In order to regularize the solution we suggest to use an alternating minimization, adding an auxiliary variable to model each patch.

Algorithm 1 Alternating Minimization Sparse Structured Light

1: **for** $k = 1, 2, \ldots$, until convergence **do**
2:     Update $\alpha_j^k(\mathbf{x})$ for all j, according to Equation (10).
3:     Update $z^k(\mathbf{x})$, according to Equation (9).
4: **end for**

We decouple the problems of regularization and structured light decoding, minimizing the functional in Equation 5, which is of a half-quadratic form [9],. Minimization with respect to α_j given z results in a per-patch denoising algorithm of $P_j z$, similar to the approach taken in [14]. We now describe the different steps of the algorithm, which is given as Algorithm 1.

Solving for z The update of z depends on the structured light patterns, and may not even be continuous. Since the similarity term relating z and $D\alpha_j$ is quadratic, we can rewrite the term for each pixel $\mathbf{x}$ in z as the sum of a photoconsistency measure and a sum of squared distances from versions of $z(\mathbf{x})$ in all of the patches containing this pixel,

$$z^{n+1}(\mathbf{x}) = \operatorname*{argmin}_{z} \rho_{SL}(z) + c_1 w(\mathbf{x})\|z - \tilde{z}(\mathbf{x})\|^2. \quad (9)$$

A solution can be obtained by sweeping the set of possible z values, similar to stereo [5]. Doing this plane-sweep is highly suitable for parallel implementation on graphics processing units (GPUs) [37]. Note that planesweeps are discrete by nature, as is the nature of the coded patterns in many cases. In order to obtain convergence, and allow sub-pixel precision, we minimize a linearly-interpolated photoconsistency, along with the quadratic distance in the second term of Equation 9. The depth estimated at each pixel is set according to the minimum of the interpolated cost function.

Solving for α_j Given a patch estimate $P_j z$, an update of the patch resorts to a standard sparse representation problem. Specifically, if we take our sparse prior to be of an L_1 type, we can update α_j using iterative shrinkage [3],

$$\alpha_i^{n+1} = S_{\lambda t}\left(\alpha_j^n - 2tD^T\left(D\alpha_j^n - P_j z\right)\right), \qquad (10)$$

where t is a gradient descent step, chosen to be small enough, and $S_{\lambda t}(\cdot)$ denotes the soft shrinkage operator,

$$S_{\lambda t}(y) = \begin{cases} 0, & |y| \le \lambda t \\ y - \lambda t, & y > \lambda t \\ y + \lambda t, & y < -\lambda t \end{cases} \qquad (11)$$

While faster iterative methods exist for L_1 minimization (see [36] for a few examples), because of the alternating minimization nature of our scheme, more complex steps may not lead to faster convergence. We therefore chose to use the original iterative shrinkage scheme.

3.1. Learning a Sparse Depth Prior

In order to learn a surface model from range images, several properties of the data must be taken into account. Since reconstruction errors are of an outlier nature, algorithms such as KSVD [7] that assume an additive white Gaussian noise model require pre-processing and outlier removal. Furthermore, since many of the patches in range scans are of smooth surfaces, and since the KSVD algorithm is initialization-dependent, care must be taken to provide a diversified initial dictionary. We focus the algorithm on the scarcer edge patches by clustering the data first using the mean-shift algorithm [6]. The resulting dictionary obtained from a set of 50 range scans is shown in Figure 2. We note that the examples used for testing are not part of this dataset, avoiding overfitting for a specific subject. While the training data is from a specific class of human faces, the learned primitives are quite general, as can be seen in Figure 2. We leave the effect of different dictionary and training data choices for future research.
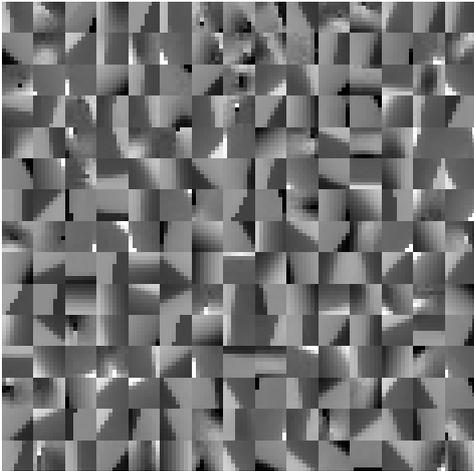
Figure 2. An example of the dictionary of 300 words obtained from a set of 50 range scans.

4. Results

In order to test the proposed scheme, we use a standard structured light setup similar to [30], with 10 striped patterns, along with an all-ones and all-zeros pattern. The camera images are sampled at 320×240, and projector patterns are shot using a 1024×768 DLP projector. In order to simulate low-SNR conditions, we have added Gaussian noise to the camera images before reconstruction. Results are shown in Figure 3 for the case of structured-light images with intensity Gaussian noise of standard deviations 5 and 10.

In order to quantitatively validate our method, we take as ground truth an almost-noiseless range image of the head statue, and measure range errors compare to it. We com-

pare both L_1 and robustified L_2, truncated at 10 millimeters, and compared to median post-processing, taken with the smallest filter size that removed range outliers from the face, in order to avoid oversmoothing. The results of this comparison are given in Table 1. For 320×240 images, the dictionary trained was of patch size 8×8.

We compare our results to several approaches. A common way of removing reconstruction artifacts is by median filtering, as was done in [27]. Yet another approach treats the problem as a denoising problem with a strong prior and impulse noise assumption. An example of this type of method would be to take the same depth prior we use, but solve a denoising problem with an L_1 fidelity term.

$$\operatorname*{argmin}_{z} \int_{\mathbf{x}} \|z - z_0\| d\mathbf{x} + \tilde{c}_1 \sum_i \tilde{\psi}\left(P_j z\right), \quad (12)$$

where z_0 is the reconstruction results without a prior. This approach would be similar, in a sense, to the depth image denoising suggested in [35]. This approach is marked in Table 1 under the *Sparse Denoise* column. In addition, it would be interesting to try a weaker prior for reconstruction such as TV regularization as suggested in Section 3. This approach is shown in the table as column *TV*. For all of the methods, parameters were chosen so as to obtain optimal robust L_2 results, while preventing remaining depth outliers. The table demonstrates the effectiveness of the proposed algorithm. While the computational cost of our algorithm is far higher with current Matlab code, the algorithm is highly parallelizable and one future line of work involves fast parallel implementation of this algorithm.

In Figure 4 we demonstrate the results of our algorithm on artifacts caused by head motion in the vertical direction. Even though the assumption of constant $a(\mathbf{x}), b(\mathbf{x})$ breaks, the algorithm overcomes many of the errors caused by reconstruction followed by outlier removal. The size of the median filter is chosen to be the smallest size that filters the motion artifacts over the eyes and mouth regions, a 7×7 filter in this case. We note that at this filter size, the mouth and nose areas merge, while artifacts remain on the eyelids.

5. Conclusions

In this paper we presented a novel model for regularized structured light reconstruction. Incorporating a sparse surface prior into a physically-motivated probabilistic outlook on structured light decoding, we demonstrate accurate results in scenarios where the usual approach for decoding structured light tends to fail.

The results obtained merit the coupling of a strong surface prior with a probabilistic model for structured light reconstruction, and motivate further exploration of the benefits of the proposed method as well as investigating the use of this approach for different types of depth scanners. An

Noise Level	Raw L_2 error	Median L_2 error	TV L_2 error	Sparse Denoising L_2 error	Sparse Reconst. L_2 error	Raw L_1 error	Median L_1 error	TV L_1 error	Sparse Denoising L_1 error	Sparse Reconst. L_1 error
2.5	1.4608	0.8411	0.8744	0.8680	0.8191	0.5996	0.4255	0.4240	0.4298	0.3379
5	2.6443	1.1033	1.1508	1.1768	0.9584	1.2013	0.5696	0.5689	0.6356	0.4135
7.5	3.9080	1.5315	1.715	1.8136	1.3489	2.1032	0.7384	0.7164	0.9489	0.5603
10	4.9841	1.9399	2.3866	2.758	1.7490	3.0949	0.9840	1.216	1.288	0.7571

Table 1. Error measurement at various noise levels, for structured light reconstruction, and noise reduction by median post-processing, reconstruction with TV prior, reconstruction followed by sparse denoising, and reconstruction using a sparse prior as shown in Algorithm 1. Errors are shown as robust L_2 (truncated at 10mm) and L_1 errors, in millimeters, over the region of the scanned object.

additional line of work involves implementing the current algorithm in an efficient manner, exploiting the high level of parallelism available in each phase.

Acknowledgment

References

[1] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *CVPR*, pages 60–65, 2005.

[2] Y. Cai, A. Nee, and H. Loh. Geometric feature detection for reverse engineering using range imaging. *J. of Vis. Comm. and Image Repres.*, 7(3):205–216, September 1996.

[3] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Process.*, 7(3):319–335, 1998.

[4] D. L. Chopp. Computing minimal surfaces via level set curvature flow. *J. Comput. Phys.*, 106(1):77–91, May 1993.

[5] R. T. Collins. A space-sweep approach to true multi-image matching. In *CVPR*, pages 358–363, 1996.

[6] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24:603–619, May 2002.

[7] M. Elad and M. Aharon. Image denoising via learned dictionaries and sparse representation. In *CVPR*, volume 1, pages 895–900, 2006.

[8] A. W. Fitzgibbon, D. W. Eggert, and R. B. Fisher. High-level model acquisition from range images. *Computer-Aided Design*, 29(4):321–330, 1997.

[9] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Trans. Image Process.*, 5(7):932–946, 1995.

[10] S. Gould, P. Baumstarck, M. Quigley, A. Y. Ng, and D. Koller. Integrating visual and range data for robotic object detection. In *ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications (M2SFA2)*, 2008.

[11] O. Hall-Holt and S. Rusinkiewicz. Stripe boundary codes for real-time structured-light range scanning of moving objects. In *ICCV*, volume 2, pages 359–366. IEEE, 2001.

[12] G. Hetzel, B. Leibe, P. Levi, and B. Schiele. 3D object recognition from range images using local feature histograms. In *CVPR*, pages 394 – 399, 2001.

[13] B. Huhle, T. Schairer, P. Jenke, and W. Straßer. Fusion of range and color images for denoising and resolution enhancement with a non-local filter. *Computer Vision and Image Understanding*, 114(12):1336–1345, 2010.

[14] K. Jia, X. Wang, and X. Tang. Optical flow estimation using learned sparse model. *ICCV*, 0:2391–2398, 2011.

[15] R. Keriven and O. Faugeras. Complete dense stereo-vision using level set methods. In *ECCV*, 1998.

[16] E. Kim and G. G. Medioni. 3D object recognition in range images using visibility context. In *IROS*, pages 3800–3807, 2011.

[17] T. Koninckx and L. Van Gool. Real-time range acquisition by adaptive structured light. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(3):432–445, 2006.

[18] D. Levin. The approximation power of moving least-squares. *Math. Comput.*, 67(224):1517–1531, 1998.

[19] T.-W. R. Lo and J. P. Siebert. Local feature extraction and matching on range images: 2.5D SIFT. *Computer Vision and Image Understanding*, 113:1235–1250, December 2009.

[20] M. Mahmoudi and G. Sapiro. Sparse representations for three-dimensional range data restoration. IMA Preprint 2280, University of Minnesota, 2009.

[21] L. Matthies, T. Balch, and B. Wilcox. Fast optical hazard detection for planetary rovers using multiple spot laser triangulation. In *ICRA*, pages 859–866. IEEE Press, 1997.

[22] R. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *ISMAR*, pages 127–136, 2011.

[23] R. A. Newcombe, S. Lovegrove, and A. J. Davison. DTAM: Dense tracking and mapping in real-time. In *ICCV*, pages 2320–2327, 2011.

[24] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(4):353–363, Apr. 1993.

[25] J. Posdamer and M. Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1):1 – 17, 1982.

[26] M. Quigley, S. Batra, S. Gould, E. Klingbeil, Q. Le, A. Wellman, and A. Y. Ng. High-accuracy 3D sensing for mobile manipulation: improving object detection and door opening. In *ICRA*, pages 3604–3610, Piscataway, NJ, USA, 2009. IEEE Press.

[27] O. Rubinstein, Y. Honen, A. Bronstein, M. Bronstein, and R. Kimmel. 3D-color video camera. In *3DIM*, pages 1505–1509, oct 2009.

[28] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D Letters*, 60:259–268, 1992.

[29] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8):2666 – 2680, 2010.

[30] K. Sato and S. Inokuchi. Three-dimensional surface measurement by space encoding range imaging. 2(1):27–39, 1985.

[31] O. Schall, A. Belyaev, and H.-P. Seidel. Adaptive feature-preserving non-local denoising of static and time-varying range data. *Computer-Aided Design*, 40(6):701–707, 2008.

[32] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-Time human pose recognition in parts from single depth images. June 2011.

[33] J. Stuehmer, S. Gumhold, and D. Cremers. Real-Time Dense Geometry from a Handheld Camera. In *Pattern Recognition (Proc. DAGM)*, pages 11–20, Darmstadt, Germany, Sept. 2010.

[34] J. P. Tardif and S. Roy. A MRF formulation for coded structured light. In *3DIM*, volume 0, pages 22–29, Washington, DC, USA, 2005. IEEE Computer Society.

[35] I. Tosic, B. A. Olshausen, and B. J. Culpepper. Learning sparse representations of depth. *CoRR*, abs/1011.6656, 2010.

[36] A. Y. Yang, A. Ganesh, Z. Zhou, S. Sastry, and Y. Ma. A review of fast l1-minimization algorithms for robust face recognition. *CoRR*, abs/1007.3753, 2010.

[37] R. Yang, G. Welch, and G. Bishop. Real-time consensus-based scene reconstruction using commodity graphics hardware. In *Pacific Conf. on Comp. Graphics and Applications*, PG '02, pages 225–, Washington, DC, USA, 2002. IEEE Computer Society.

[38] Z. Yang and D. Purves. A statistical explanation of visual space. *Nat Neurosci*, 6(6):632–40, June 2003.

[39] S. Yoshizawa, A. Belyaev, and H. P. Seidel. Smoothing by example: mesh denoising by averaging with similarity-based weights. *Proceedings of International Conference on Shape Modelling and Applications*, pages 38–44, 2006.

[40] L. Zhang, B. Curless, and S. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *3DPVT*, pages 24–36. IEEE, 2002.
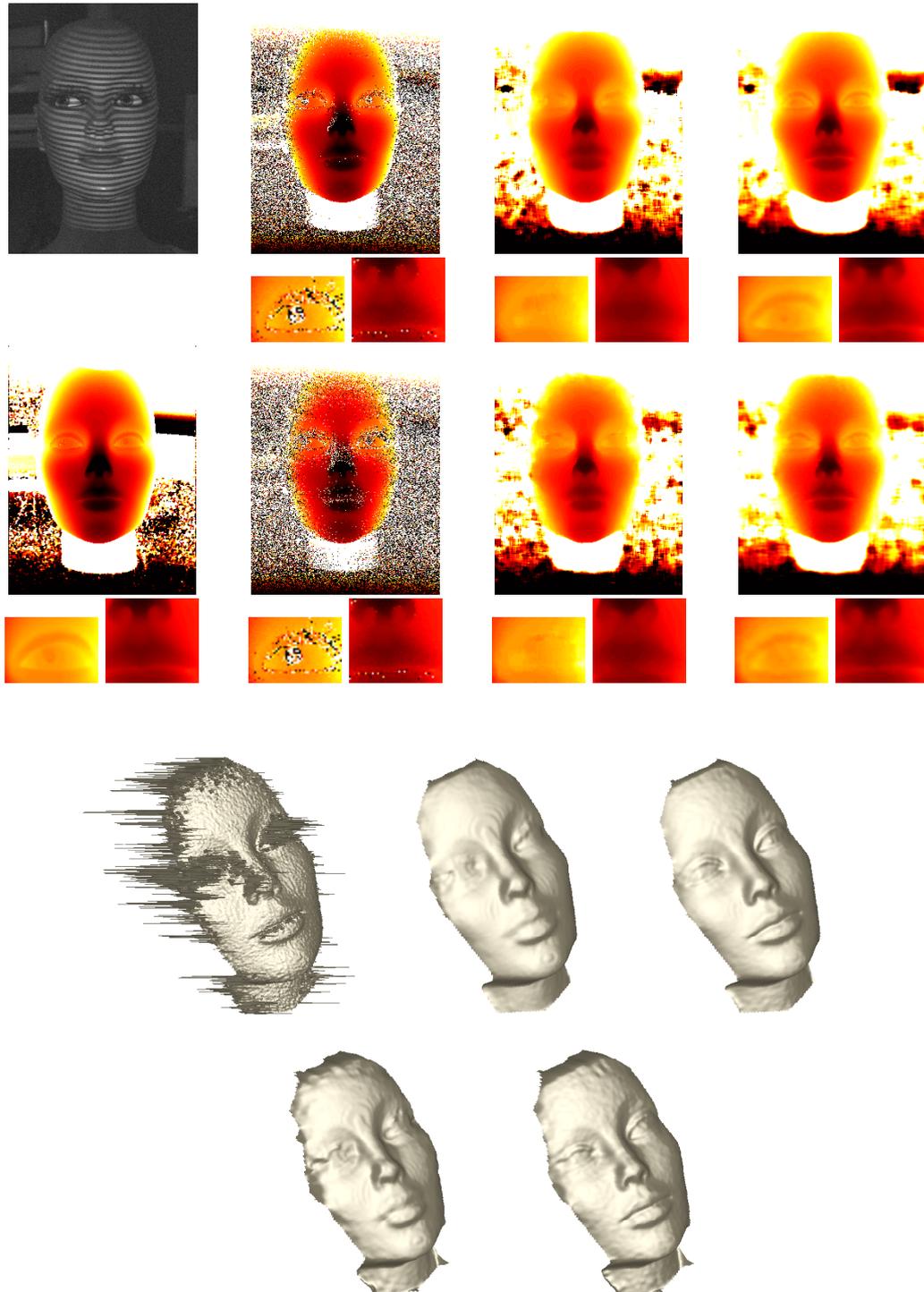
Figure 3. First row, left-to-right: An example textured pattern, reconstruction results, reconstruction with median filtering, reconstruction with sparse prior, where camera images were added Gaussian noise with standard deviation of 5, with close-up on the right eye region and the nose and mouth region. Second row, left-to-right: ground-truth reconstruction obtained from noiseless reconstruction, same sequence of results, where camera images were added Gaussian noise with standard deviation of 10. Third row, left-to-right: 3D raw reconstruction results, reconstruction with median post-processing and with a sparse prior for the case of $\sigma = 5$ noise. Fourth row, left-to-right: (3D raw reconstruction omitted since it was too noisy), reconstruction with median post-processing and with a sparse prior for the case of $\sigma = 10$ noise. In order to view the range images, color and/or online viewing is suggested.
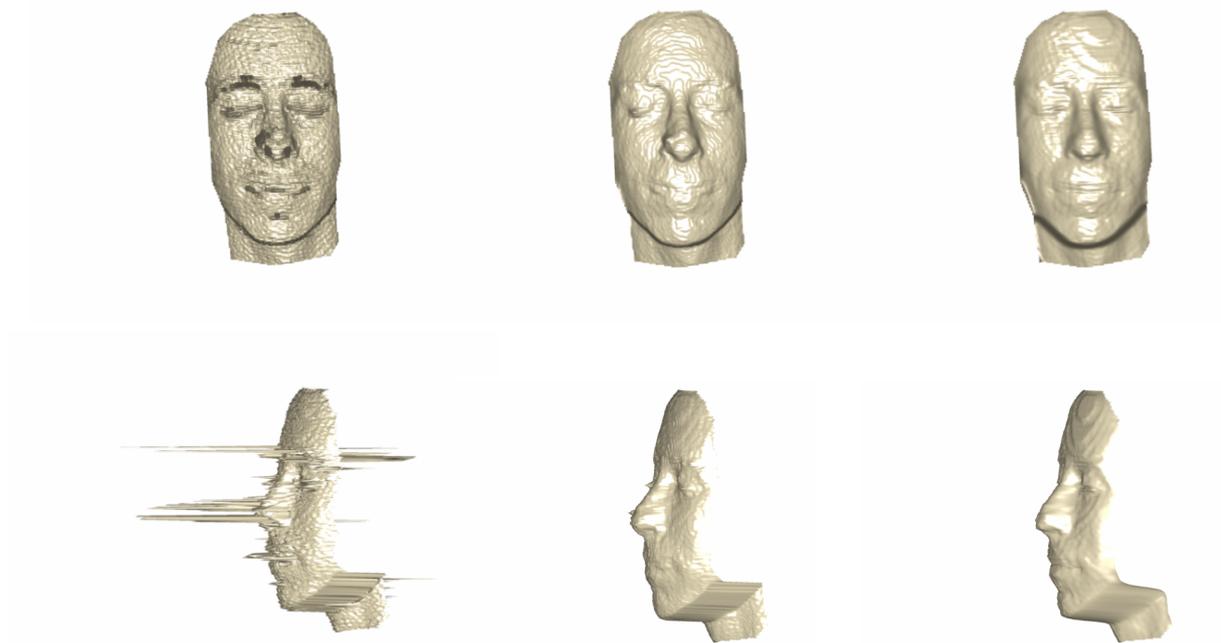
Figure 4. Left-to-right: An example with artifacts caused by vertical head motion, a median-filtered result, the result of the proposed method. Note the merging of the mouth and nose area in the median filter, and the remaining artifacts around the left eye and nose area.