Sparse Modeling of Shape from Structured Light

Guy Rosman, *Student Member, IEEE*, Anastasia Dubrovina, *Student Member, IEEE*, and Ron Kimmel, *Fellow, IEEE*

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 and illumination model is crude, and is executed separately from the decoding phase which often assumes a generic interference model. In this paper we bridge this gap by viewing the reconstruction process via a probabilistic model combining illumination and shape. Specifically, we present an alternating minimization algorithm for structured light reconstruction, incorporating a sparsity-based prior for the local surface model. Integrating this 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.

Index Terms—3D/stereo Scene Analysis, 3D Reconstruction, Range data, Structured light, Sparse Priors, Nonlinear Optimization, Probabilistic Models.

1 Introduction

With ever more prevalent sources for 3D data, 3D acquisition and processing is an increasingly important part of scene analysis. Active illumination range scanners are used for scene understanding [14], [23], [18], robotics [25], [12], [30], object modeling [10], [3], indoor scene mapping [26], and human computer interaction [37], among other tasks.

Structured light systems usually consist of a calibrated camera-projector pair, where coded light pattern sequences emitted by the projector are acquired by the camera, allowing robust triangulation and depth reconstruction. Time-multiplexed structured light systems trade-off spatial for temporal resolution. They allow us to obtain dense and accurate reconstruction at low cost, with relatively simple hardware and without too many limiting assumptions on the scene. Other alternatives for structured light attempt to trade-off resolution for coding robustness by incorporating decoding schemes for larger neighborhoods which add a certain assumption of regularity. For a review of existing structured light techniques see, for example, [34].

In order to improve reconstruction robustness, 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 [48], [20], or temporal objects behavior [13], [48], [20]. This regularity, however, is usually based on channel decoding error approaches, and does not relate to the geometry of the scene or the image formation model. As such, its optimality is often limited due to the inaccurate reconstruction error model.

Yet, modeling these assumptions in a more complete way is crucial when the captured illumination patterns are of low SNR, for example due to long scanning range and short camera exposure times. In the case of dynamic scenes, 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 cause more reconstruction artifacts.

The probabilistic model we present here relates the time-multiplexed structured light to methods for spatio-temporal stereo reconstruction [8], [39]. In our case, however, we are estimating the expected camera luminous intensity, rather than assuming brightness constancy.

Here, we obtain improved reconstruction results from structured light scanners [29], [34], in face of challenging illumination conditions and motion artifacts, by providing strong priors for the imaging model and surface shape. Instead of using strong shape priors for range image correction, the approach we suggest incorporates shape and illumination priors into the reconstruction itself, giving us a principled approach of combining powerful surface priors and probabilistic understanding of the acquisition process. We use patch-based range image priors, similar to those successfully utilized for images, depth images,

• *G. Rosman, A. Dubrovina and R. Kimmel are with the Department of Computer Science, Technion, Haifa, Israel, 32000. E-mail: {rosman,nastyad,ron}@cs.technion.ac.il*

This research was supported by European Community's FP7- ERC program, grant agreement no. 267414.

and surface processing [2], [9], [41], [36], [46], [47], [21]. We demonstrate the priors obtained from range images to be quite intuitive and meaningful.

This paper builds upon a previous conference paper [31], discussing more completely the reconstruction model and demonstrating additional priors. Furthermore, we add more examples in order to test the behaviour of the algorithm in real-life low SNR conditions, and add an additional structured-light patterns scheme.

Specifically, in Section 2 we develop our reconstruction model. In Section 3 we describe the resulting reconstruction algorithm. We demonstrate our results and several aspects of the model's behavior on real images in Section 4. Section 5 concludes the paper and discusses future venues of research.

2 Regularized Structured Light Model

In shape from structured light, we reconstruct the geometric structure of the scene based on active illumination. We illuminate the scene with projected patterns $I_P = \{I_P^{(i)}\}_{i=1}^N$, where N is the number of patterns, and capture a sequence of images $I_C = \{I_C^{(i)}\}_{i=1}^N$ with a camera. Let us 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 $z(\mathbf{x})$. $\mathbf{x} \in \mathbb{R}^2$ is the (two-dimensional) camera image coordinates vector.

In this work we assume a Lambertian surface model for objects, and a projector emitting directional light in a temporal sequence of patterns. The main source of image noise is assumed to be the sensor/imaging process. Although other sources of deviations from the model exist (for example object motion), in many cases they can be overcome as we will show. Since structured-light systems decode a set of patterns and need all of the patterns to be decoded correctly, we can assume relatively low noise levels – 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 dynamic scenes, and thus requiring short exposure intervals.

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

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

$a(\mathbf{x})$ 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. $\Pi_z(\mathbf{x})$ denotes the depth-dependent intensity 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. $n^{(i)}(\mathbf{x})$ is the pixel noise, assumed to be additive white Gaussian noise, independent and identically distributed (i.i.d.) in space and time. The above model assumes a linear camera gain model. In practice, this approximate model works well enough so as to obtain good reconstruction results. Incorporating camera gain nonlinearity is deferred as future work since it is not necessary for this model.

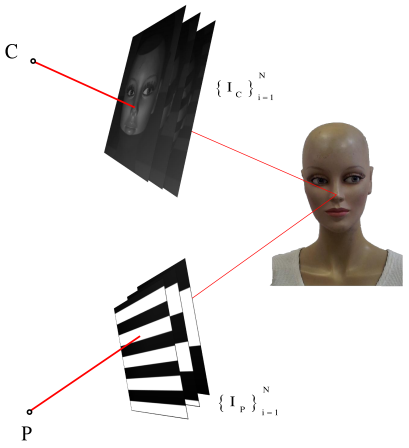

Fig. 1. An example of a structured light system setup.

We wish to formulate and maximize a probability function of the depth given the known camera images and projected textures. In reconstruction we are looking for the depth value $z(\mathbf{x})$ that maximizes the probability

$$\begin{aligned} z &= \underset{z}{\operatorname{argmax}} \min_{a,b} P(z, a, b | I_P, I_C) \\ &= \underset{z}{\operatorname{argmax}} \min_{a,b} \frac{P(z, a, b, I_P, I_C)}{P(I_P, I_C)} \\ &= \underset{z}{\operatorname{argmax}} \min_{a,b} \frac{P(I_P, I_C, a, b | z) P(z)}{P(I_P, I_C)} \\ &= \underset{z}{\operatorname{argmax}} \min_{a,b} P(I_P, I_C, a, b | z) P(z) \\ &= \underset{z}{\operatorname{argmin}} \min_{a,b} \left(-logP(I_P, I_C, a, b | z) - logP(z)\right), \end{aligned} \quad (2)$$

where we have applied Bayes' rule, and switched to log-probability domain. 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 [6] when seeking the optimum value of z. In the framework of probabilistic inference, this is known as max-sum elimination. Minimizing the

negative log-probability over a and b, we have

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

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 $\mathbf{x}, \Pi_z(\mathbf{x})$, respectively,

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

Inserting the optimal a, b as a function of z and noting the conditional independence (given z) of neighboring pixel values $I_C(\mathbf{x}), I_P\left(\Pi_z(\mathbf{x})\right)$ provides us with a functional to minimize with respect to $z(\mathbf{x})$, similar to [40],

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

The expression $\rho_{SL}\left(z; I_C, I_P, \mathbf{x}\right)$ denotes a penalty for the photoconsistency assumption. In standard structured-light techniques, this term is often optimized per pixel in 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. In order to achieve robustness to noise and computational efficiency, these approaches treat binarization or code-word identification errors as general channel decoding errors, using robust codes which have a firm theoretical background, but which do not really model the channel characteristics for this specific problem. These characteristics should result from the imaging model and geometric relations, and should not be ignored.

The term $\psi(z)$ 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 [45]. In Section 2.1 we describe several possible regularization priors for depth images.

Incorporating Illumination and Reflectance Characteristics An additional improvement to the model can be made if we take into account the regularity of scene characteristics such as object albedo, illumination and normals. These assumptions have been utilized in the context of shape-from-single-image reconstruction [1], but in our case a simpler model suffices. We augment the photoconsistency assumption by adding a prior for the choice of a, b,

$$\min_z \left(\min_{a,b} \left[\sum_i \frac{\left(a(\mathbf{x}) I_P^{(i)}\left(\Pi_z(\mathbf{x})\right) + b(\mathbf{x}) - I_C^{(i)}(\mathbf{x})\right)^2}{\sigma_I^2} + \frac{\left(a(\mathbf{x}) - \mu_a\right)^2}{\sigma_a^2} + \frac{\left(b(\mathbf{x}) - \mu_b\right)^2}{\sigma_b^2} \right] \right), \quad (6)$$

where μ_a, μ_b are taken from a locally computed average. σ_a, σ_b are constants set manually, since estimating second-order moments from a small neighborhood of a noisy signal can be quite sensitive. This allows us to compute a, b even if only a few noisy frames are available, as is often the case with multiple color structured light systems, as shown for example in Figure 4.

2.1 Regularization Terms for Depth Images

We now describe a few possible regularization terms for the depth image, representing various tradeoffs between model robustness and computational efficiency.

Total-Variation Regularization The minimum area [5] and *total-variation* [33] (TV) priors, and related smoothness measures have been suggested in several forms for regularization of range images [27] and surface reconstruction [17], [38], [19]. TV regularization for structured light can be expressed as

$$\operatorname*{argmin}_z \int_{\mathbf{x}} \rho_{SL}\left(z; I_C, I_P, \mathbf{x}\right) + \tilde{c}\|\nabla z\| dx, \quad (7)$$

where $\|\nabla z\|$ is the total variation of the range image, for some coefficient $\tilde{c}$. This form of regularization is strongly related to MRF-based structured light [40]. A related prior is the second order total variation,

$$\operatorname*{argmin}_z \int_{\mathbf{x}} \rho_{SL}\left(z; I_C, I_P, \mathbf{x}\right) + \tilde{c}\|Hz\| dx, \quad (8)$$

where H denotes the matrix of second order derivatives of z, $H(z) = (z_{xx}, \sqrt{2} z_{xy}, z_{yy})$. This prior can be computed quite efficiently and lends itself to parallel computation [42]. Furthermore, it is well suited to the

often-made approximation of the scene as a piecewise-linear surface.

Patch-based L_1 Prior 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 [22] 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 [36], [15], [24], [41].

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

$$\underset{z}{\operatorname{argmin}} \int_{\mathbf{x}} \rho_{SL}(z; I_C, I_P, \mathbf{x})\, d\mathbf{x} + \tilde{c}_1 \sum_j \tilde{\psi}(P_j z), \quad (9)$$

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

$$\underset{z,\alpha_j}{\operatorname{argmin}} \int_{\mathbf{x}} \rho_{SL}(z; I_C, I_P, \mathbf{x}) + \tilde{c}_1 \left(\sum_j \|P_j z - D\alpha_j\|^2 + \lambda \|\alpha_j\|_1 \right), \quad (10)$$

where D denotes a dictionary for depth image patches, P_j denotes a matrix extracting block j from the image in column-stacked notation, and α_j denotes the representation coefficients of patch $P_j z$ in that dictionary.

Gaussian Mixtures Prior for Structured Light Since depth images are expected locally to be very sparse, another approach of modeling them is by a Gaussian-mixture patch model, which in a sense first selects the support set of atoms from a structured-sparsity dictionary where each Gaussian component defines a support set, and then estimates its coefficients. In this approach, data patches are assumed to be generated from a sparse Gaussian-Mixture model in patch-space, similar to the approach suggested by Yu et al. [47], and Zoran and Weiss [49]. The relation of this image model to sparsity has been thoroughly discussed in [47]. In our case, Gaussians are pre-learned from a set of depth images, although an adaptive approach, learning the component distributions from the processed image itself is also possible. Unlike the case of natural images, such a learning process would have to account for the bias of the depth image patches, and the nature of the noise in the initial reconstruction results, as described in Subsection 3.1. As the reconstruction errors are far from the standard additive noise model, learning under such an outliers noise is not trivial and is left for future research. The components of the Gaussian components form natural features of the range images, adapted to edges and corners. This is not surprising, and has often been demonstrated in sparsity-related literature. The components of the patch distribution obtained by a Gaussian mixture model (GMM) are given in Figure 3. The optimization problem can be written as

$$\underset{z,\alpha_j,k_j}{\operatorname{argmin}} \int_{\mathbf{x}} \rho_{SL}(z; I_C, I_P, \mathbf{x}) + \sum_j \tilde{c}_1 \|P_j z - U_j f_j\|^2 + \lambda \left(f_j^k\right)^T \Sigma_k^{-1} \left(f_j^k\right) + \frac{\lambda}{2} log\left(\pi^N \left|\Sigma_k^{-1}\right|\right), \quad (11)$$

where f_j^k denotes the coefficients used to represent patch j in terms of Gaussian component k_j. Σ_k denotes the covariance matrix of component k, N is the number of pixel in each patch, and U_j define the principal directions of the Gaussian component used for patch j. As is often the case in patch-based priors, the patches' mean is subtracted before coefficients estimation, and added before the synthesis of the new patch.

3 Alternating Minimization Algorithm for Regularized Structured Light

We now describe the specific algorithm used to solved structured-light reconstruction with a sparsity-based prior, as shown in Equation 10. A complete algorithmic description is given as Algorithm 1.

We assume the coded light pattern can be initially 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 photo-consistency term between the structured light patterns and the resulting camera image intensities [28], when estimating the illumination conditions. Note that this function depends only on the depth value and camera intensities per pixel. In order to obtain the regularized solution we suggest to use an alternating minimization. By adding a set of auxiliary variables, we decouple the problems of regularization and structured light decoding. This is done by minimizing the functional in Equation 5, which is of a half-quadratic form [11]. Minimization with respect to the regularization term given z results in a denoising problem. For patch-based priors, the resulting approach is similar to the one shown in [16]. We now detail each of the minimization steps.

Solving for z The update of z depends on the structured light patterns, and may not even be continuous. We note that for all of the regularization terms presented in subsection 2.1, the term coupling the regularization to the photoconsistnecy term is quadratic in $z_i(\mathbf{x})$, the patch-dependent representation of $z(\mathbf{x})$ in patch i. Therefore 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 $\tilde{z}(\mathbf{x})$, an averaged version of $z(\mathbf{x})$ in all of the patches containing this pixel, with an aggregate weight $w(\mathbf{x})$,

for every $\mathbf{x}$

$$z^{n+1} = \underset{z}{\operatorname{argmin}}\, \rho_{SL}(z) + \tilde{c}_1 w \|z - \tilde{z}\|^2. \quad (12)$$

A solution can be obtained by sweeping the set of possible z values, similar to stereo algorithms [6]. Doing this plane-sweep is highly suitable for parallel implementation on graphics processing units (GPUs) [44]. Note that plane-sweeps are discrete by nature, as are the coded patterns in many cases. This does not constitute a further disadvantage as they are of approximately the same resolution. In order to obtain convergence, however, and in order to allow sub-pixel precision, we minimize a linearly-interpolated photoconsistency, along with the quadratic distance in the second term of Equation 12. The depth estimated at each pixel is set according to the minimum of the interpolated cost function, allowing us to incoporate sub-pixel precision into the plane-sweeping operator, as can be seen in the results section. Achieving sub-pixel resolution is important both in terms of accuracy and in terms of the visual artifacts that accompany discrete-pixel reconstruction, as seen for example in Figure 5, where the staircasing effect in the noise of the median-filtered reconstruction is typical of discrete-patterns structured-light systems.

Solving an L_1 regularization Given a patch estimate $P_j z$, an update of its representation becomes a standard sparse approximation problem. Specifically, if we take our sparse prior to be of an L_1 regularity type, we can update α_j using iterative shrinkage [4],

$$\alpha_j^{n+1} = S_{\lambda t}\left(\alpha_j^n - 2tD^T\left(D\alpha_j^n - P_j z\right)\right), \quad (13)$$

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

$$S_{\lambda t}(y) = \begin{cases} 0, & |y| \le \lambda t \\ y - \lambda t, & y > \lambda t \\ y + \lambda t, & y < -\lambda t \end{cases} \quad (14)$$

While faster iterative methods exist for L_1 minimization (see [43] 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. We note that the dictionary in our case is pretrained from a set of depth images. The exact training procedure is defined in Subsection 3.1.

Solving a GMM regularization prior In this case, the choice of Gaussian component and its coefficients are given by going over the Gaussian components, computing the corresponding linear estimator, and the resulting log-probability term. Given that patch $P_j z$ belongs to Gaussian mixture component k with basis U_k and covariance matrix Σ_k, the linear estimator for the coefficients f_j is given by

$$f_j^k = \left(\tilde{c}_1 U_k^T U_k + \lambda \Sigma_k^{-1}\right)^{-1}\left(\tilde{c}_1 U_k^T\right) P_j z. \quad (15)$$

Algorithm 1 Alternating Minimization Sparse Structured Light

1: Compute initial reconstruction z by plane-sweeping, according to Equation 6.
2: **for** $k = 1, 2, \ldots,$ until convergence **do**
3: Update auxiliary variable for regularization:
- Update $\tilde{z}^k$ by TV denoising, or second order TV denoising, according to [42], or
- Update $\alpha_j^k(\mathbf{x})$ for all j, according to Equation (13), for L_1 regularization, or
- Update $\tilde{f}_j^k$ by GMM component selection and linear estimation according to Equation 15,16, for GMM regularization.

5: Update $z^k(\mathbf{x})$, according to Equation (12).
6: **end for**

The component k for each patch is chosen so that the minimum regularized error is achieved over all components,

$$\underset{k}{\operatorname{argmin}} \left(\begin{array}{l} \tilde{c}_1 \|P_j z - U_j f_j^k\|^2 + \\ \lambda \left(f_j^k\right)^T \Sigma_k^{-1} \left(f_j^k\right) + \frac{\lambda}{2} log\left(\pi^N \left|\Sigma_k^{-1}\right|\right) \end{array} \right), \quad (16)$$

where the GMM coefficients f_j^k are computed according to Equation 15. We refer the reader to Yu et al. [47] for further elaboration on the method, and comment that the Gaussian mixture component are pretrained on a dataset of images, as defined in Subsection 3.2.

3.1 Learning a Depth Dictionary

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 [9] that assume an additive white Gaussian noise model. Such algorithms require some form of pre-processing and outlier removal in order to train on data with outliers. 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 less-frequent edge patches by clustering the data first using the mean-shift algorithm [7]. 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. Thus we avoid 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.

3.2 Learning a Gaussian-Mixture Model for Depth Images

For the GMM prior we have used 200 Gaussian components, learned from the same dataset as the sparse

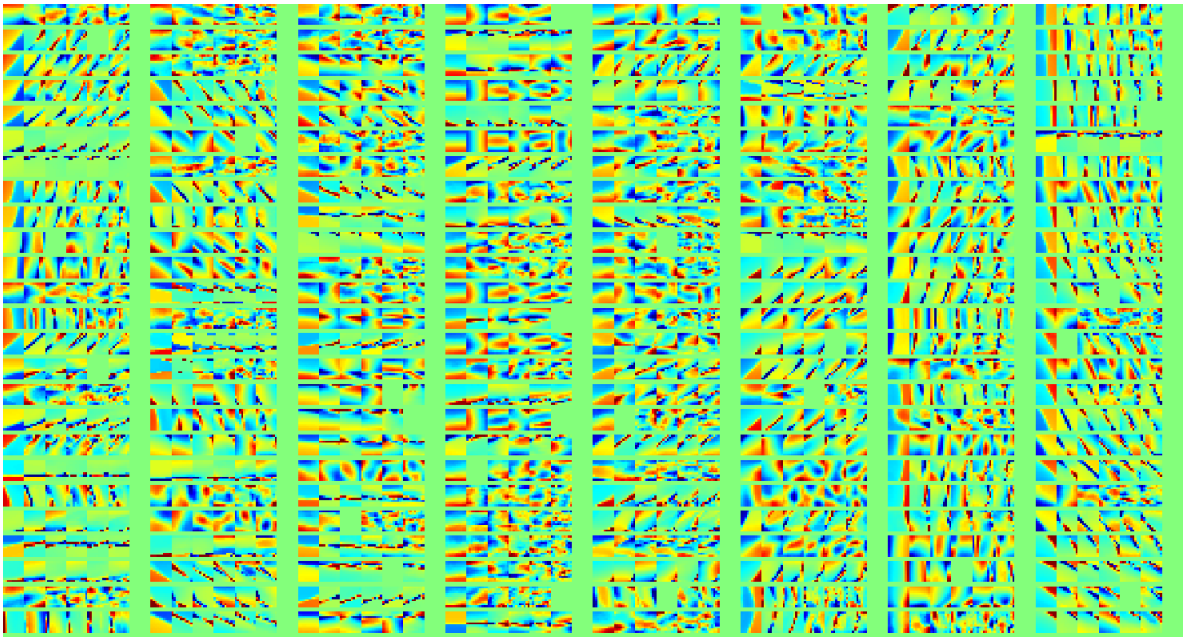

Fig. 3. An example of 200 Gaussian mixture components obtained from a set of 50 range scans. Each 6-columns group represents the principle directions of Gaussian components. Each row represents a Gaussian component, with the leftmost columns representing the more variable directions in the mixture.

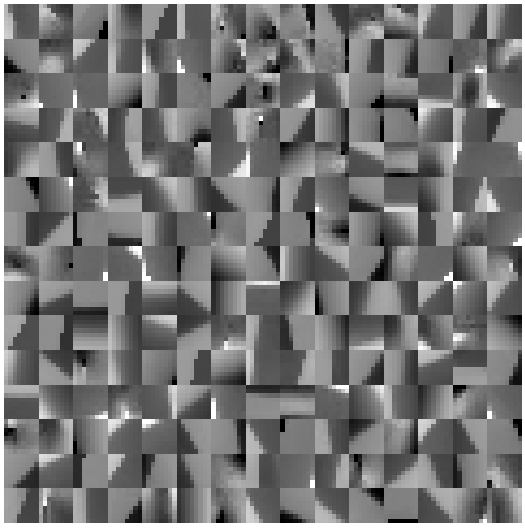

Fig. 2. Example atoms from a dictionary of 300 words obtained from a set of 50 range scans.

dictionary prior, and using the same type of pruning for flat and outlier-containing patches. Learning the GMM component was done in a standard way (see [47] for more details). The components of the patch distribution obtained by a Gaussian mixture model are given in Figure 3.

4 Results

We now proceed to demonstrate the results of the proposed scheme. We first note that merely by using a sweeping approach instead of the usual decoding approach, we can improve the reconstruction. This is not surprising since the channel noise model used in the standard gray-code reconstruction was inaccurate to begin with. This improvement is obtained even without an additional regularization term, as shown in Figure 4. In this figure, in order to measure the amount of reconstruction outliers detected, we measure the deviation of the current depth beyond the $(0.4, 0.6)$ quantiles of the local neighborhood depth for a small (9×9) region.

The importance of using a better per-pixel model can be clearly observed around the eyes of a reconstructed face, which is often a problematic area in 3D reconstruction due to the low reflection coefficient of the pupil. This is demonstrated in Figure 5, where reconstructing and then post-processing the depth image does not provide reasonable reconstruction of the eyes region. Similarly, the sides of the face which are poorly illuminated by the projector suffer from reconstruction artifacts as well.

The main structured-light patterns scheme we experimented with is a standard structured light setup similar to [35], with 10 striped black and white patterns, along with an all-ones and all-zeros pattern. The camera images are sampled at a resolution of 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 6,7 for the case of structured-light images with intensity Gaussian noise of standard deviations 5 and 10.

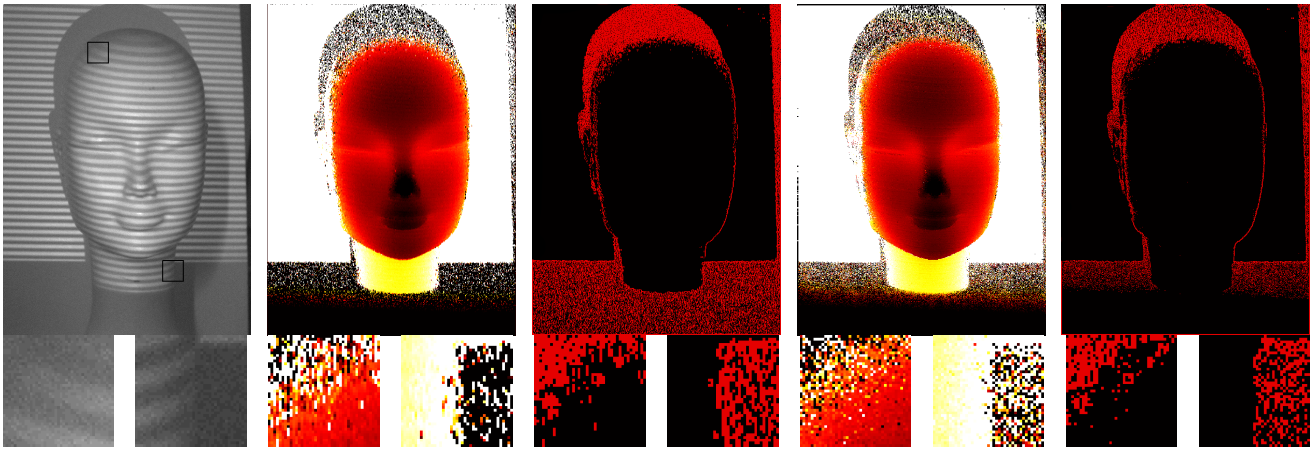

Fig. 4. Top row, left-to-right: One of the texture images, the result obtained by the method of [32], and an outlier map (red pixels signify gross errors in the reconstruction), the result obtained by plane-sweeping according to Equation 6, and an outlier map. Bottom row: two zoomed-in areas of low SNR, marked as boxes in the texture intensity image. In these areas of weak illumination, plane-sweeping results in fewer outliers compared to a standard decoding approach for structured-light.

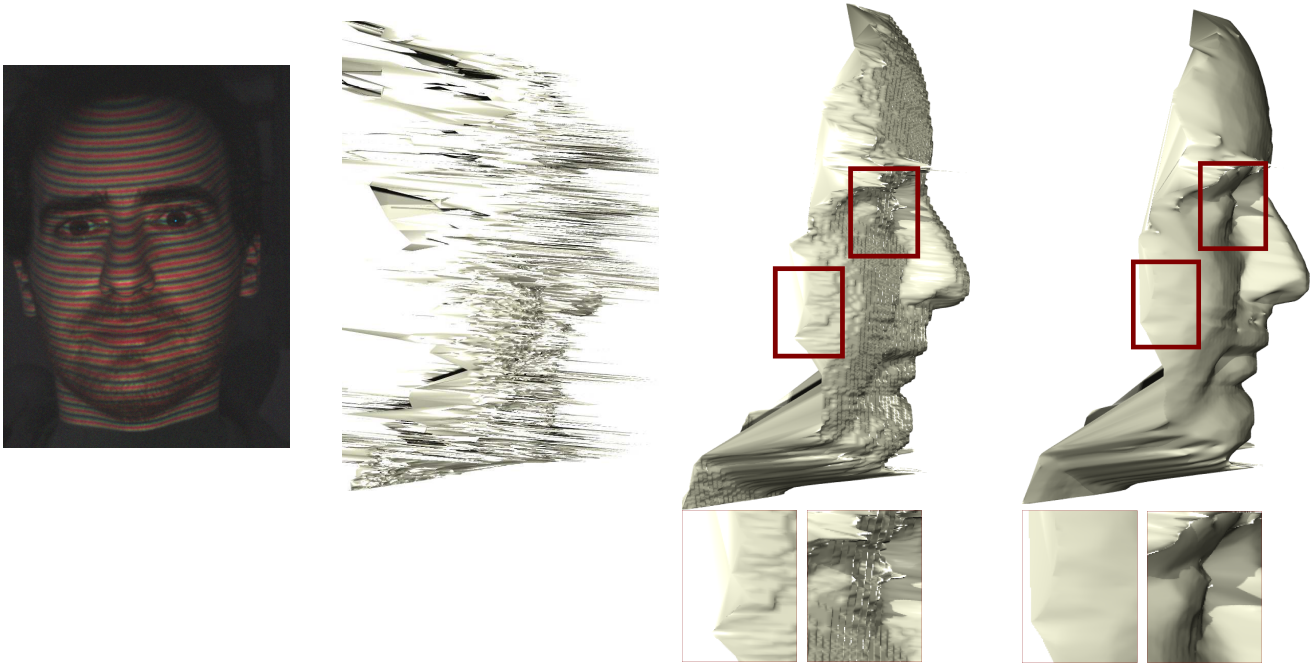

Fig. 5. An example reconstruction of the eye region of a person. Left-to-right: the intensity image based on the structured-light setup of [32], the result obtained by plane-sweeping according to Equation 6 with no post-processing, the result after median filtering, and the result of regularized reconstruction using Equation 8.

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 compare both L_1 and robustified L_2, truncated at 10 millimeters. The error measurements are performed over a manually segmented mask of the 3D object in the image domain. The results of this comparison are given in Table 1. For all of the images, the dictionary trained for patched-based priors was of patch size 8×8. As can be seen, the error of the median filtered result is smaller than those of sparse denoising with robust fitting term, or that of TV regularized reconstruction. This is due to the fact that TV regularization is too weak to overcome errors in the data term, and denoising with an L_1 term is still somewhat sensitive to the strong outliers found in structured-light reconstructed depth images.

We compare our results to several approaches. A common way of removing reconstruction artifacts is by median filtering, as was done in [32]. We compared to median post-processing, taken with the smallest filter size that removed range outliers from the face, in

order to avoid oversmoothing. 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

$$\underset{z}{\operatorname{argmin}} \int_{\mathbf{x}} \|z - z_0\| d\mathbf{x} + \tilde{c}_1 \sum_j \tilde{\psi}\left(P_j z\right), \qquad (17)$$

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 [41]. 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 quite high with current Matlab code, the algorithm is highly parallelizable and one future line of work involves fast parallel implementation of this algorithm.

In Figure 8 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 down, the algorithm overcomes many of the errors caused by a decoding-based 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.

4.1 Color Structured Light Example

Another example patterns scheme we used involves a color pattern projector, similar to [32]. In this setup, a single grayscale camera is used, operating at a resolution of 480×360, at 180 frames-per-second. The exposure time is $5.56ms$, due to synchronization between the projector and camera (see [32] for more details). A DLP projector emits color patterns sequentially in each cycle, and 12 patterns are used, 4 at each channel. In such a patterns set, since only 4 patterns are available per color channel, estimation of a, b is sensitive to image noise.

It is quite important in this setup to have a prior for a, b as part of the model. Incorporating such a prior as shown in Equation 6 contributed greatly to the reconstruction performance. The result of the reconstruction is shown in Figures 5,9. The noise levels in these examples are not very high, but these examples are important because they demonstrate a real structured-light scenario, with real sensor short exposure artifacts. The frame-rate of the camera, about 15Hz, is still relatively low. It is therefore important to stress that in faster scanners short exposure time (and the resulting artifacts) is likely to play an even more significant role.

In Figure 9 we demonstrate the results using the Gaussian mixture model prior shown in Figure 3. This result demonstrates the generality of the proposed framework using a different regularization term. The Gaussian mixture components we used are shown in Figure 3.

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. Even in the case of no regularization, incorporating a realistic illumination model into the reconstruction cost function leads to a more robust reconstruction of each range pixel. An additional line of work involves implementing the current algorithm in an efficient manner, exploiting the high level of parallelism available in each phase. Other relevant venues of research include online learning of the surface model, and the incoproration of a more complete illumination model into the reconstruction.

Acknowledgments

This research was supported by the European Community's FP7-ERC program, grant agreement no. 267414. The authors would like to thank Mr. Alon Zvirin for work with the structured light scanner.

References

[1] J. T. Barron and J. Malik. Shape, albedo, and illumination from a single image of an unknown object. In *CVPR*, pages 334–341, 2012.

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

[3] 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.

[4] 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.

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

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

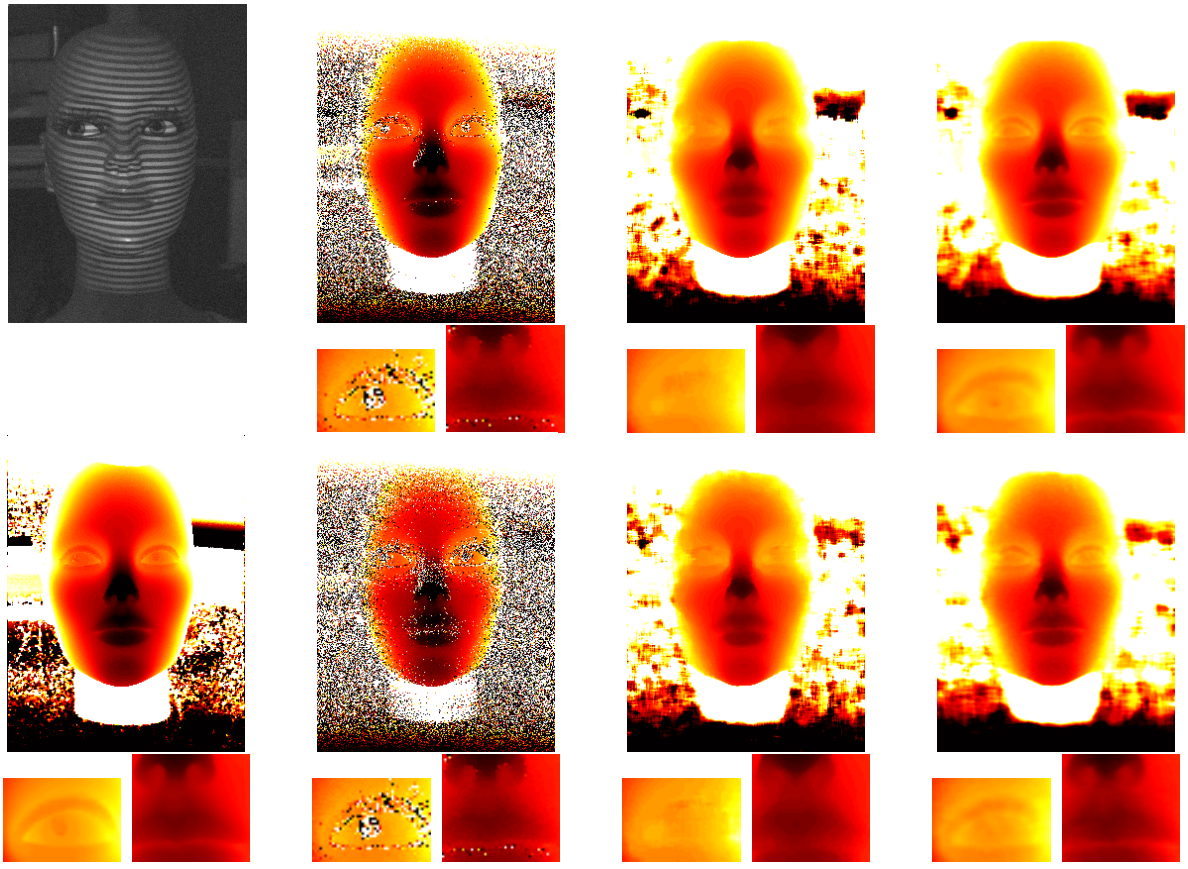

Fig. 6. First row, left-to-right: An example textured pattern, reconstruction results, reconstruction with median filtering, reconstruction with sparse prior, where camera images were added Gaussian noise with standard deviation of 5, with close-up on the right eye region and the nose and mouth region. 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. In order to view the range images, color and/or online viewing is suggested.

[7] 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.

[8] J. Davis, D. Nehab, R. Ramamoorthi, and S. Rusinkiewicz. Spacetime stereo: A unifying framework for depth from triangulation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(2):296–302, Feb. 2005.

[9] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15(12):3736–3745, 2006.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] K. Jia, X. Wang, and X. Tang. Optical flow estimation using learned sparse model. In *ICCV*, pages 2391–2398. IEEE Computer Society, 2011.

[17] R. Keriven and O. Faugeras. Complete dense stereovision using level set methods. In *ECCV*, 1998.

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

[19] R. Kimmel. 3D shape reconstruction from autostereograms and stereo. *Journal of Visual Communication and Image Representation*, 13(1-2):324–333, 2002.

[20] 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.

[21] F. Lenzen, K. I. Kim, R. Nair, S. Meister, H. Schafer, F. Becker, C. Garbe, and C. Theobalt. Denoising strategies for time-of-flight data. In *Time-of-Flight Imaging: Algorithms, Sensors and Applications*, 2012.

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

[23] T.-W. R. Lo and J. P. Siebert. Local feature extraction and matching on range images: 2.5D SIFT. *Computer Vision and*

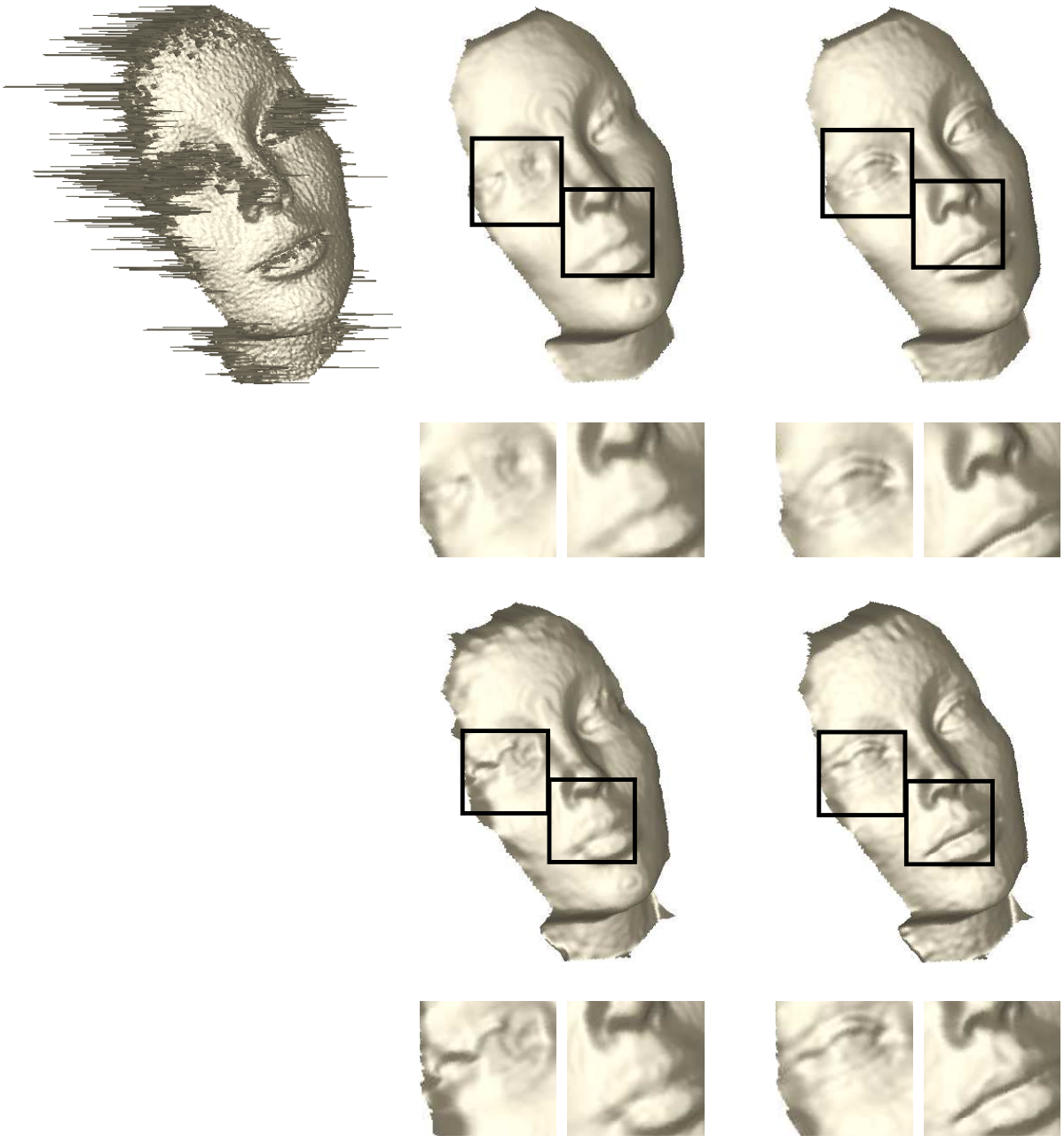

Fig. 7. First row, left-to-right: 3D raw reconstruction results, reconstruction with median post-processing and with a sparse prior for the case of $\sigma = 5$ noise. Second row, left-to-right: (3D raw reconstruction omitted since it was too noisy), reconstruction with median post-processing and with a sparse prior for the case of $\sigma = 10$ noise. In order to view the range images, color and/or online viewing is suggested.

Image Understanding, 113:1235–1250, December 2009.

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

[25] 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.

[26] 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.

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

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

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

[30] 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.

[31] G. Rosman, A. Dubrovina, and R. Kimmel. Sparse modeling of shape from structured light. In *Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing,*

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.

Visualization & Transmission, 3DIMPVT '12, pages 456–463, Washington, DC, USA, 2012. IEEE Computer Society.

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

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

[34] 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.

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

[36] 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.

[37] 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. In *CVPR*, June 2011.

[38] 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.

[39] Y. Swirski, Y. Y. Schechner, and T. Nir. Variational stereo in dynamic illumination. In *ICCV*, pages 1124–1131, Washington, DC, USA, 2011. IEEE Computer Society.

[40] 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.

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

[42] C. Wu and X.-C. Tai. Augmented lagrangian method, dual methods, and split bregman iteration for ROF, vectorial TV, and high order models. *SIAM J. Img. Sci.*, 3:300–339, July 2010.

[43] 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.

[44] 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.

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

[46] 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.

[47] G. Yu, G. Sapiro, and S. Mallat. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *CoRR*, abs/1006.3056, 2010.

[48] 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.

[49] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In *ICCV*, pages 479–486, 2011.

G uy Rosman is currently pursuing his PhD in the Computer Science Department, at the Technion, Israel. He graduated in 2004 his BSc Summa Cum Laude at the Technion, and in 2008 his MSc, Cum Laude, at the Technion, in the Computer Science Department. During his studies, he worked at IBM Haifa Research Labs, and as an algorithm developer in Rafael Advanced Defense Systems, Medicvision Imaging Solutions and Invision Biometrics LTD. He is the recipient of the Jacobs-Qualcomm award, Intel PhD award, and CS faculty excellence prizes. His research interests include fast variational methods and PDE based image processing, as well as algorithms for motion and structure estimation, surface processing and 3D reconstruction.

A nastasia Dubrovina is currently pursuing her PhD in the Computer Science Department, at the Technion. She graduated in 2008 her BSc Summa Cum Laude at the Technion, and in 2010 her MSc at the Technion, in the Electrical Engineering Department. During her studies, Anastasia worked at at Rafael Advanced Defense Systems, IBM Haifa Research Labs and HP Labs Israel. Her research interests include non-rigid and deformable surface matching, spectral surface processing, 3D reconstruction, and image segmentation. Anastasia's PhD research was awarded the Jacobs-Quallcom prize.

P rofessor Ron Kimmel is a researcher in the areas of computer vision, image processing, and computer graphics. He is a tenured Professor at the Computer Science Dep., Technion where he holds Montreal Chair in Sciences. He is a Technion graduate (PhD/DSc 1995), spent his postdoctoral years (1995-1998) at Berkeley, and was a visiting professor at Stanford (2003-2004).

Prof. Kimmel has published over a hundred and fifty articles and papers in scientific journals and conferences. He is on the editorial board of International Journal of Computer Vision (IJCV), IEEE Transactions on Image Processing (TIP), and Journal of Mathematical Imaging and Vision (JMIV), and co-chaired conferences like Scale-Space Theories in Computer Vision and SIAM conf. on imaging, First Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (NORDIA)-2008 and Tenth Asian Conference on Computer Vision (ACCV)-2010.

Ron Kimmel is the author of Numerical Geometry of Images, published by Springer in 2003 and coauthor of Numerical Geometry of Non-Rigid Shapes, published by Springer in October 2008. Ron Kimmel is IEEE Fellow for contributions to image processing and non-rigid shape analysis since 2009. He was awarded "Test of Time" award for ICCV1995, Geodesic Active Contours paper in 2011.

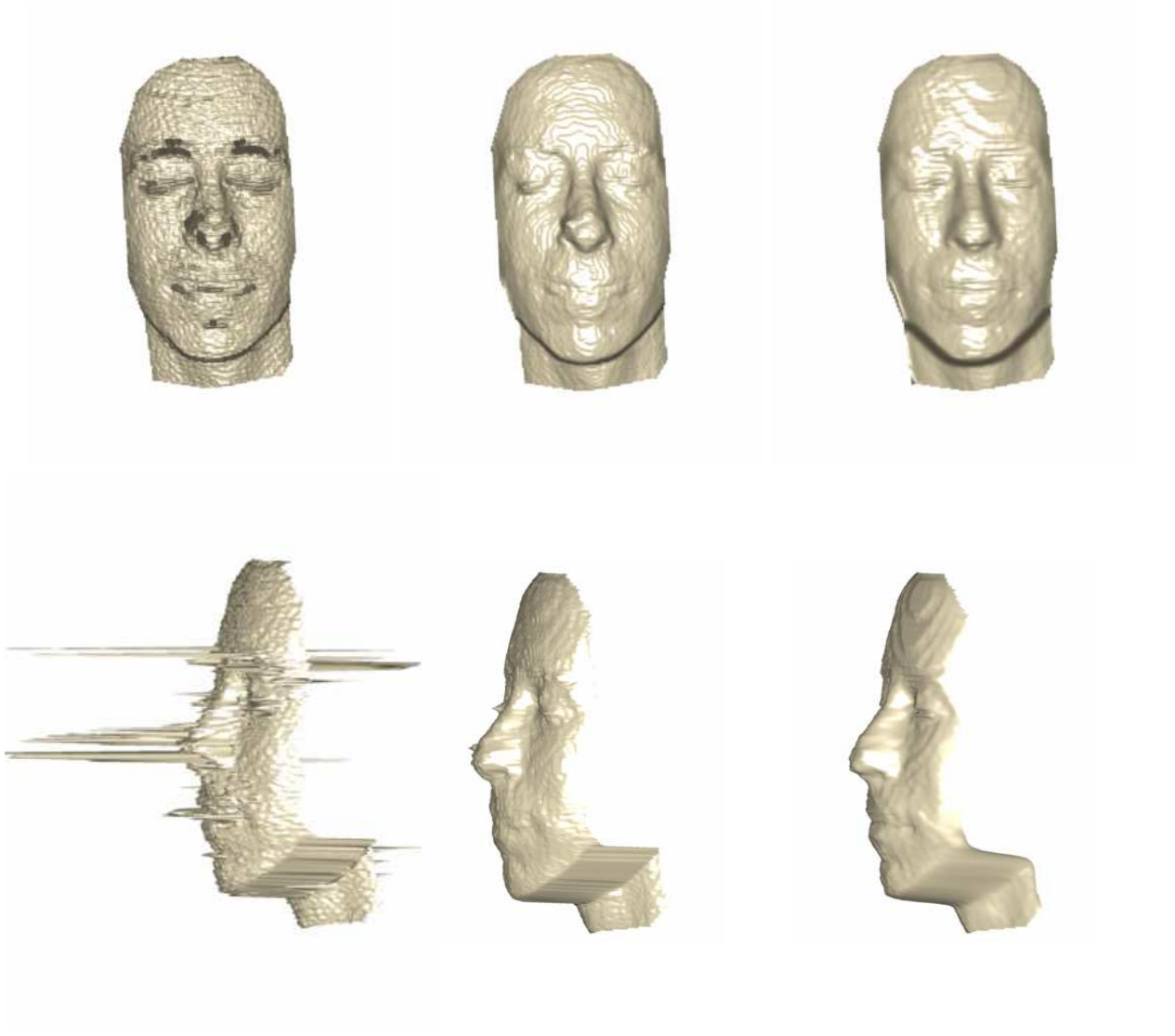

Fig. 8. Left-to-right: An example with artifacts caused by vertical head motion, a median-filtered result, the result of the proposed method. Note the merging of the mouth and nose area in the median filter, and the remaining artifacts around the left eye and nose area.

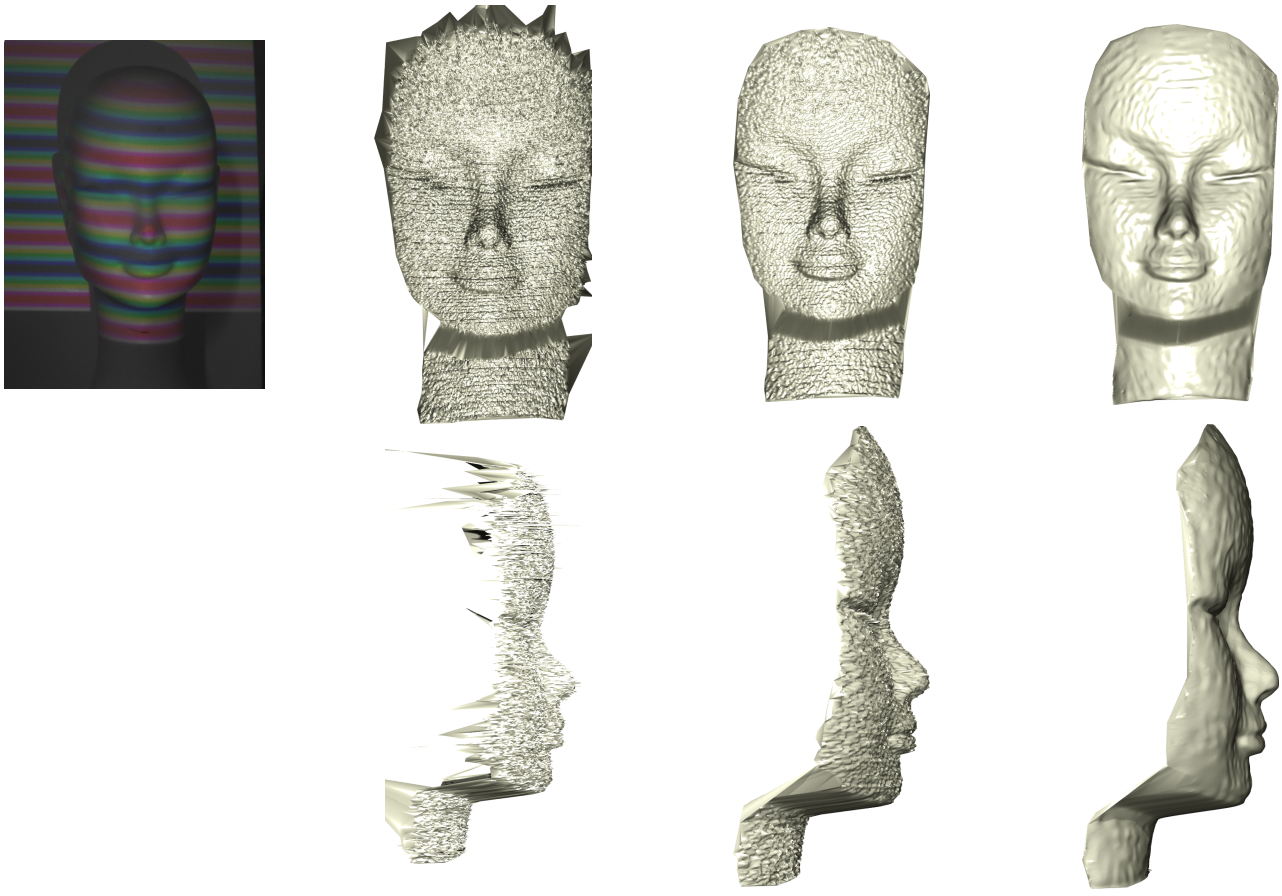

Fig. 9. Reconstructed surface based on the structured-light setup of [32], with GMM prior. Top, Left-to-right: one of the camera images in all 3 channels, raw reconstruction front view, median-filtered initial solution, regularized reconstruction with GMM prior. Bottom: reconstruction, side view.