This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2018.2830181, IEEE Transactions on Computational Imaging



# Information-Driven Adaptive Structured-Light Scanners

Guy Rosman, Daniela Rus and John W. Fisher III,

***Abstract*—Sensor planning and active sensing, long studied in robotics, adapt sensor parameters to maximize a utility function while constraining resource expenditures. Here we consider information gain as the utility function. While these concepts are often used to reason about 3D sensors, these are usually treated as a predefined, black-box, component. In this paper we show how the same principles can be used as part of the 3D sensor.**

**We describe the relevant generative model for structured-light 3D scanning and show how adaptive pattern selection can maximize information gain in an open-loop-feedback manner. We then demonstrate how different choices of relevant variable sets (corresponding to the subproblems of localization and mapping) lead to different criteria for pattern selection and can be computed in an online fashion. We show results for both subproblems with several pattern dictionary choices and demonstrate their usefulness for pose estimation and depth acquisition.**

***Index Terms*—Depth sensors, structured-light, information-gain, sensor planning, 3D scanners, uncertainty, simultaneous localization and mapping, generative models.**

## I. Introduction

Range sensors have revolutionized computer vision in recent years, with commodity RGB-D scanners allowing us to easily tackle challenging problems such as articulated pose estimation [1], Simultaneaous Localization and Mapping (SLAM) [2], [3], [4], and object recognition [5], [6]. Reasoning about 3D sensors often utilizes simplified, black-box, models of the acquisition process, that are only loosely coupled to the photometric principles behind the design of the scanner. 3D sensors abstract a significant complexity of the relations between the acquired images and the underlying scene behind them, giving us encapsulated representations of the environment that are simple and convenient to exploit in higher-level processes [7]. Given such intermediate representations, we can employ computer vision algorithms to understand the world and act based on the resulting understanding of the scene.

Significant efforts have been devoted to optimal planning and deployment of sensors under resource constraints, *e.g.*, on energy, time, or computation. Such *sensor planning* has been employed in many aspects of vision and robotics, including positioning of 3D sensors and cameras, as well as other *active sensing* problems, see for example [8], [9], [10], [11], [12], [13], [14]. The goal is to focus sensing on the aspects of the environment or scene most relevant to the specific inference task, taking all considerations into account.

G. Rosman, D. Rus and J.W. Fisher III are with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 02139 USA e-mail: (see http://www.michaelshell.org/contact.html).

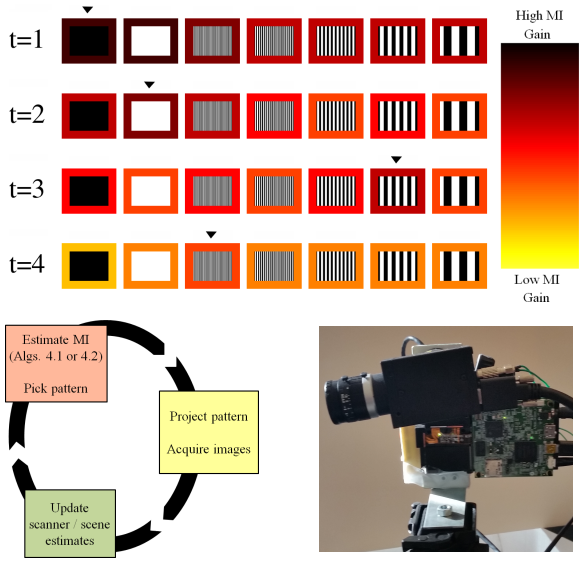


Fig. 1. Illustration of patterns selection. Each row illustrates another turn of pattern selection. For each pattern, the information gain is estimated, shown by different border color around each pattern, and the different stem heights in the plot on the left. Black arrowheads and red circles in the plot mark the selected pattern at each turn. Note the different patterns selected, and diminishing information gain over time. Bottom row: Left: the proposed open-loop w/ feedback 3D scanning with pattern selection flowchart. Right: the project/camera system used for 3D scanning.

However, the same optimality principles are generally not used to examine the operation mode of the 3D sensor itself, and the 3D representation hides properties of the acquisition process that could be leveraged to tailor the acquisition process for the specific task at hand. In reality, the true measurements in the system are acquisitions by a photosensitive sensor, and not the 3D representation. Parameters of the sensors, including any active illumination, should be considered as *action* parameters (in the decision-theoretic sense [15]) to be optimized and planned.

In this paper we focus on temporal structured light 3D scanners [16], [17] as an example for 3D sensing. In structured light 3D scanners, the trade-offs of temporal speed, robustness, and accuracy are well known – for example, Kinect's single shot patterns afford speed and lack of motion artifacts, but lack in spatial accuracy, whereas Gray-code temporal multiplexed patterns offer accuracy but lack speed and are sensitive to motion artifacts. We reformulate adaptive selection of patterns in structured-light scanners as the following resource constrained sensor-selection process, affording us an adaptive approach that can leverage context and reason about these tradeoffs.

We treat the choice of the projected pattern at each time as a planning choice, and the number of projected patterns as a resource. Our goal is to minimize the number of projected patterns while maximizing the task-specific *information gain*. We compute information gain from the (predicted) observation of the scene given previous observations and a new proposed projected pattern. This allows us to pick the next pattern, project it, and update our world state estimate in an online fashion, corresponding to the greedy selection regime in sensor selection, as is illustrated in Figure 1.

The contributions of this paper are: (i) We devise a probabilistic generative graphical model for the 3D scanning process, depicted in Figure 2. We estimate mutual information between the observed images and variables in the model in Algs. 1,3. (ii) For the task of range estimation, we demonstrate greedy open-loop pattern selection for the projector as an instance of focused active inference [18], in Subsec. IV-A. (iii) For the task of pose estimation, we show which parts of the scene are informative, for several cases of interest, in Subsec. IV-B.

We have presented the concept of focused active inference in 3D scanners in a recent conference submission [19]. We now extend upon this submission with addition discussion of the model and its components, as well as empirical exploration of the model's relation to more elaborate illumination models, describing the implication to the potential gain from active sensing. We moreover give a complete description of the algorithms involved and describe the use of active sensing as open-loop-with-feedback control.

We note that sensor planning is an instance of experimental design, studied in a variety of domains, including economics [20], medical decision making [21], robotics [22], [23], and sensor networks [24], [25], [26], [27], [28], [29]. While many optimality criteria have been proposed, one commonly used criterion is information gain. It is well-known that selection problems have intractable combinatorial complexity. However, it has been shown that tractable greedy selection heuristics, combined with open-loop feedback control [30] guarantee near-optimal performance [26], [31], due to the submodular property of conditional mutual information (MI). We therefore focus on this approach of greedy pattern selection in this paper. This assumes one can evaluate the information measure for the set of sensing choices (patterns in our current context). We derive a physics-based model for structured-light sensing that simultaneously lends itself to tractable information evaluation while producing superior empirical results in a real system. We also characterize the informational utility of a given pattern (or class of patterns) in the face of varying *relevant* versus *nuisance* parameter choices [18]. In the process, we demonstrate that the value of a given structured-light pattern changes depending on the specific inference task. We exploit commonly available graphics hardware to efficiently estimate the information gain of a selected pattern and reason about the effect of the dependency structure in the probabilistic model.

The choice of parameterization for the latent variables in the model is crucial for efficient information gain estimation. This can be seen in the common tasks of range sensing and pose estimation. We consider these two important applications and demonstrate how a careful choice of the scene and scanner representation lends itself to estimation of conditional mutual information.

Good inference and uncertainty estimation hinge on finding a scene parameterization that affords easy and efficient computation. Such a representation should model the sensing process faithfully, and ideally be suited to inference and uncertainty estimation in several tasks. Within a single model, this ability is often achieved by inferring only a subset of variables, or computing the *focused* mutual information [18] with respect to them. Choosing parameterizations of the scene which lend themselves to tractable uncertainty estimation, and further processing along a machine vision pipeline has been intensely studied in the robotics community, with several choices each having its own advantages and disadvantages for several applications. We explore in this paper some of these tradeoffs in the context of 3D reconstruction and computer vision from range scanners.

More concretely, our contribution is in defining a framework for inference and uncertainty estimation in active illumination 3D scanners. We first describe the model on which we base our framework, and then detail several techniques for estimating the mutual information between relevant variable sets, such as depth and pose, within this model. A significant difference of this problem compared to the classical geometry-based reasoning of information planning is the introduction of a much more complex set of nuisance variables. We discuss this important point as we develop our model, which captures both the geometry and photometry aspects of structured-light reconstruction in sufficient detail, allowing both inference and uncertainty-based sensory selection of scanner patterns.

In the field of structured-light reconstruction, several studies have suggested adaptive scanners (see for example [32], [33], [34], [12], [45]), and energy-efficient designs [35]. Similar to the design of fixed-sequence scanners, they trade off acquisition speed (and sensitivity to motion artifacts), robustness to various modes of corruption, and accuracy, but often fail to take into account the distribution of posterior uncertainty in illumination and geometry, or to generalize to multiple pattern libraries. In this paper we show how given a generative model for the sensing process we can obtain an adaptive scanner for various tasks, constraints, and pattern choices, forming a decision-theoretic *purposive* [36] 3D scanner that can be adapted to specific tasks and relevant sets for inference.

We formulate 3D acquisition as a probabilistic inference process within a detailed model for the scene and sensor in Section II. We discuss methods of representing uncertainty in a manner appropriate for a specific task. In Section III we show how MI estimation can be combined with standard approaches for reconstruction in several cases of interest, and demonstrate the integration of MI estimation into a structured-light scanner. Section IV demonstrates the proposed system in several experiments that exemplify the usefulness of the proposed approach. Section V concludes the paper and describes possible new directions.

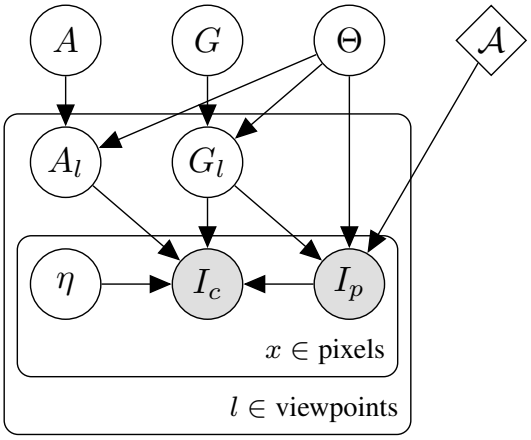


Fig. 2. Proposed model for classification with active illumination.

## II. Modelling Active 3D Computer Vision

We now describe the generative model used for pattern selection and inferring depth. We adopt a model that describes structured-light and time-of-flight imaging devices and standard cameras or camera-and-projector systems. Estimation of information gain is central to our method and thus impacts the choice of parameterization. We emphasize that approximations we use for estimating information gain and choosing patterns generally do not carry over when we compute the reconstruction. To our knowledge, this is the first analysis of active information-based planning in this setting. The model parameters are roughly partitioned into agent pose, geometry of the scene, and photometry of the scene. This is described in Figure 2, with the notation below:

- $A$ is the global scene appearance, assigning an intensity model to each geometry element in $G$. For the case of Lambertian illumination, this includes two random variables $a(x), b(x)$ that describe the photometry of the visible surface with respect to the perceived projector illumination. Their mean and standard deviation parameters, $a_0(x), b_0(x), \sigma_a(x), \sigma_b(x)$, are incorporated into $A$.
- $G$ is a global scene representation. For simplicity's sake, we chose a range image with range/height pixels distributed i.i.d with a Gaussian prior per pixel, $r(x) \sim \mathcal{N}\left(\mu_r, \sigma_r^2\right), \quad x \in \mathbb{R}^2$, and our scene representation is the projection with that range image from the camera's point of view at $\Theta = Id$. This representation is further described in Section III.
- $\Theta$ denotes the scanner pose – which we define as the transformation between the camera and the scene. We assume it to be distributed as a Gaussian in the tangent space around $(Id, \mathbf{0})$ in the Lie-group $SE(3)$. For the case of range estimation, $\Theta$ can be fixed at the identity, specifying a default gauge for the problem.
- $G_l$ is the *local* (viewpoint) scene representation. In our case, it is the rotation of elements of $G$ according to $\Theta$, using a deterministic function.
- $A_l(x) \in \mathbb{R}^2$ is the local appearance model. For each element in the local geometry $G_l$ we attach its perceived illumination model, $a, b$, sampled from the corresponding element in $A$. We note $A_l, G_l$ are not deterministic functions of $A, G, \Theta$ due to unmodeled aspects (*e.g.*occlusions). The geometry and pose determine camera and projector coordinates at each pixel.
- $\mathcal{A}$ represents the action space in a sensory-planning notation. In our case it is the choice of pattern.
- $I_p$ at image pixel $x$ is the projection of the projector patterns based on the reprojected point $\Pi^P_{r,\Theta}(x)$. While this can be a random variable, in practice the noise levels of the projector light are negligible compared to the camera noise $\eta$.
- $I_c(x) = a\left(\Pi^C_{r,\Theta}(x)\right) I_p\left(x\right) + b\left(\Pi^C_{r,\Theta}(x)\right) + \eta(x)$. Note that we assume the sampling process to be accurately taking us from $I_c(x)$ to a value of $I_p$ which we, in practice, sample bilinearly. Also, note that given past image measurements, $I_c$ is still Gaussian. This allows us to compute current estimates for $I_c$ without having to perform complex operations such as matrix inversion on parallel hardware.
- $\eta$ captures the pixel noise, as well as additional phenomenon unaccounted for (such as occlusions). In the examples we show we used additive Gaussian white noise with the same parameter for all pixels.

Computing the operators $\Pi^C_{r,\Theta}(x), \Pi^P_{r,\Theta}(x)$ involves first computing a world-coordinates representation $x^3$ of the geometry $G$ at each geometry element $x$. In order to compute $x^3$ in our representation, we back-project $x$ to a distance of $r(x)$ from the canonical viewpoint (chosen in our case to coincide with that of the camera at rest gauge).

$$\begin{aligned} \tilde{x} =& K^{-1}_{cam} x^3 \\ x^3 =& \frac{r(x) \begin{pmatrix} (\tilde{x})_1 \\ (\tilde{x})_2 \\ 1 \end{pmatrix}}{\sqrt{(\tilde{x})_1^2 + (\tilde{x})_2^2 + 1}}, \end{aligned} \tag{1}$$

where we abuse the notation of $x$ for its homogeneous coordinates representation. We then compute the rigid transformation $\bar{x}^3 = T_\Theta(x^3)$, followed by projection onto the camera and projector's viewpoints to form $\Pi^C_{r,\Theta}(x), \Pi^P_{r,\Theta}(x)$. Based on $\Pi^{(\cdot)}_{r,\Theta}(x)$, we sample $I_c$,$I_p$, and the appearance coefficients with bilinear interpolation.

The generative graphical model of Figure 2 depicts the relationships of the variables. Observations are denoted by shaded circles, latent variables by white circles, and parameters by diamonds. As shown in Figure 2, the model factorizes as

$$\begin{aligned} & p\left(A, G, \Theta, A_l, G_l, \eta, I_c, I_p; \mathcal{A}\right) \\ & = p\left(\Theta\right) p\left(A\right) p\left(G\right) \\ & \quad \prod_l p\left(A_l | A, \Theta\right) p\left(G_l | G, \Theta\right) \\ & \quad \prod_{l,x} p\left(I_c | A_l, G_l, I_p, \eta\right) p\left(I_p | G_l, \Theta; \mathcal{A}\right) p\left(\eta\right), \end{aligned} \tag{2}$$

where the first line includes prior terms for the scene. The second incorporates projection onto a specific viewpoint of the projector images and world model, and the last line involves sensor image rendering, and noise realization.

We note that depending on the inference task, various latent variables alternate their roles as either relevant or nuisance. We choose patterns in order to maximize *focused* information gains [18], i.e., information regarding the *relevant* set, rather than information of the non-relevant, or nuisance, variables. We follow the notation of [18] where $\mathcal{R} \subseteq \mathcal{U}$ denotes the relevant set and $\mathcal{U}$ denotes the set of all nodes. Nuisance parameters have certainly been considered in existing 3D reconstruction methods. Examples include the standard *binarize-decode-reconstruct* approach for time-multiplexed structured-light scanners or the choice of view-robust descriptors for 3D reconstruction from multiple views [37]. The utility of the generative model is that nuisances are dealt with in a mathematically-consistent fashion.

### A. Inference and Sensor Planning in 3D Vision

We consider several inference tasks of interest in 3D computer vision and the pattern selection issues which arise from them. For example, inference of $G_l$ given $I_c, I_p, \Theta$ amounts to 3D reconstruction, where $G_l$ is assumed to approximate $G$ and $A_l$ is treated as a nuisance. Previous methods adopt a probabilistic model for improving structured-light reconstruction [38], [39], but assume a predetermined set of patterns. Alternatively, SLAM methods incorporate inference steps for the geometry and pose parameters alternating between pose ($\Theta$) updates conditioned on the geometry ($G_l$) and vice-versa. Updates to the 3D map may be posed as inference of $G$ given $G_l, \Theta$. In all cases, limiting assumptions regarding occlusions, the relation of appearance parameters and 3D geometry, and the relation between different range scans of the same scene are typically invoked.

For structured-light acquisition, one can associate pixels in $I_c$ and $I_p$ given the range $r$ at each pixel $x$ (which is a choice for $G_l$) and the pose $\Theta$. The set of pixels in $I_p$ are obtained via $\Pi_{r,\Theta}(x) \in \mathbb{R}^2$ by back-projecting $x$ into the 3D world and projecting it into the projector image plane. The relation between the intensity values of these pixels can be given as

$$I_c(x) = a(x) I_p\left(\Pi_{r,\Theta}(x)\right) + b(x) + \eta(x), \tag{3}$$

where $a, b$ depend on the ambient light, normals, and albedo of the incident surface. For sufficiently large photon count, $\eta$ is assumed Gaussian accounting for sensor noise and unmodeled phenomena such as occlusions and non-Lambertian lighting components. Utilizing time-multiplexed structured-light, plane-sweeping [39] enables efficient inference of $G_l$ from $I_c, I_p$, and incorporation of priors on the scene structure $G$. For our purposes, one can assume a fixed pose, and limit the inference to estimation of $G_l$. Figure 3 provides an example of $I_c, I_p, a, b, r$ for a reconstructed scene with random smoothed patterns (as described in Subsection IV-A). The resulting 3D reconstruction is superior to the classic binarize-decode-triangulate pipeline with respect to robustness to artifacts such as specularities and low SNR conditions.

Our goal is to efficiently compute the relevant mutual information quantities $\mathcal{I}^{\mathcal{A}}(\mathbf{x}_{\mathcal{R}}; I_C)$ for different definitions of $\mathcal{R}$, and choices from the set $\mathcal{A}$, alternately considering $\Theta$, $G$, and $A$ as the relevant variable set $\mathbf{x}_{\mathcal{R}}$. Nonlinear correspondence operators (back-projection and projection) linking $I_c, I_p$ complicate dependency analysis within the model and preclude analytic forms. We exploit common graphics hardware for a straightforward and efficient sampling approach that follows the generative model.

### B. Photometric Entropy in Active Illumination 3D Scanning

When describing 3D scanner, the interplay of photometric models and the reconstruction can lead to improved results [40], [41] and warrants examination. In Equation 3, coefficients $a$ and $b$ capture illumination variability. A slightly more detailed description of the photometric model is given by a Lambertian model,

$$I_c = \rho \frac{1}{r_p(x)^2} \langle n(x), l \rangle \, I_p(\pi_r(x)) + \rho I_{amb} + \eta(x). \tag{4}$$

Here, $\rho = \rho(\pi_r(x))$ is the albedo coefficient, $n(x)$ is the surface normal at a given image location $x$, $l$ is the projector direction, and $I_{amb}$ is the ambient lighting. $r_p$ is the distance from the projector, and $I_p(\pi_r(x))$ is the projector intensity. Comparing this model with Equation 3, we can approximate

$$a(x) \approx \rho \frac{1}{r_p(x)^2} \langle n(x), l \rangle ; \qquad b(x) \approx \rho I_{amb}. \tag{5}$$

While we can use the model of Equation 4 for inference, additional variables make inference less stable, and often require more observations for the same certainty in the variables of interest. We therefore use the simplified model from Equation 3, which have proven itself in structured-light reconstruction (see, for example, [39]).

However, we can still use the model of Equation 4 to explore the contributions of the different factors, similar to an ablation study [42]. Observing the pixel-wise intensity entropy associated with different simplifications of this model provides us with intuition on the relative importance of various factors and gives us some bounds on how much information can be gained from modification of the patterns. Specifically, when looking at different patterns, we can contrast a pattern that changes in an i.i.d manner in the image plane to a fixed pattern that is deformed according the the surface geometry and projection operator. The difference in image entropy between the two hints at the possible information gain achievable by better selection of patterns, and bounds the maximum information gain per pixel. we perform a simulation of the perceived illumination and measure the per-pixel entropy. We use a Lambertian illumination model as in Equation 4. The range of the surface affects the illumination through several terms in the equation: notably, there is (i) intensity attenuation of the projector due to distance, (ii) effect on the normal, and the (iii) texture change due to the triangulation. In addition, there is (iv) the projector intensity in a specific projector ray Removing each of the factors from the model in Equation 4, we get an estimate of the contribution of each factor. We assume a Markov Random Field (MRF) model for the surface depth with uniform weights, and a constant surface albedo. The estimated entropy values are shown in Figure 4, for different variability levels of the range image, and at different base ranges.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2018.2830181, IEEE Transactions on Computational Imaging



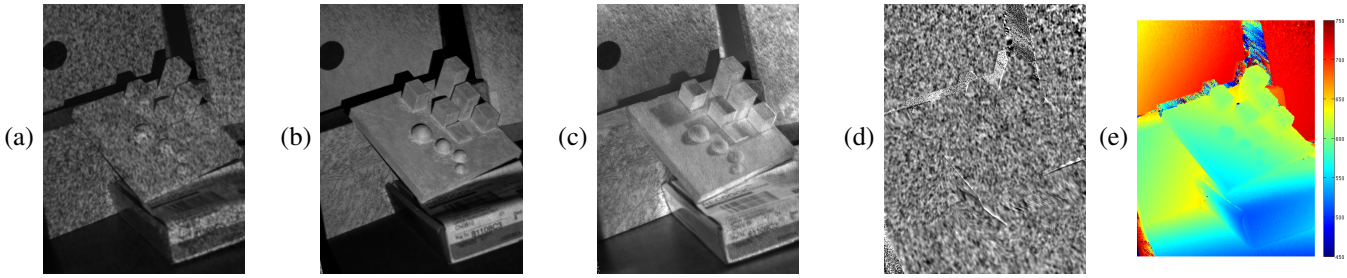


Fig. 3. Left-to-right: a) $I_c$, b) coefficients $a$ and c) $b$ from Equation 3 for the MAP-estimated range d) $I_p$ in the camera image plane, e) the range image in millimeters. Note how parameter $b$ captures scene illumination, whereas parameter $a$ captures the reflectance coefficient of the surface with respect to the projector.

Ideally, a pattern should utilize at each region the full intensity range to probe the full range of geometric uncertainty. The difference between variation sources (iii, black curves) and (iv, blue curves) hints at the maximal achieved variation in the image intensity for patterns designed to investigate our specific range uncertainty at each pixel. It demonstrate how by varying the illumination according to the current uncertainty, we can produce patterns that explore each area more efficiently. In our examples, the average range of the surface is set to be either 20cm, 1 meter, or 10 meters, with range uncertainty of up to 10cm. The ratio of ambient light intensity to maximum projected pattern variation is 1. We note that at in our tests, at reasonable ranges the normals effect on the intensity dominates that of the distance attenuation. It is, however, a standard assumption to avoid the effects relating to surface normals in structured-light and time-of-flight scanners.

We further note that at larger distances, the change due to geometry's effect on the pattern diminishes, requiring a larger disparity between the projector and camera, or higher resolutions of the projector, patterns, and cameras. This is a known limitation in the design of structured-light scanners, and can be seen in the transition between the $20cm$ and $10m$ plots. However, the effect of the change in the projected pattern dominates other effects, much more than the change in projector ray-surface location due to the range. This suggests that by adapting the patterns according to uncertainty in range, there is a lot to be gained in these cases. Integrating photometric normal-based reconstruction and code could be done within the model we propose, but it beyond the scope of this work. Directly modifying the pattern leads to much higher entropy compared to displacing a fixed pattern due to changes in the range image. It hints at the amount of modification for $P(I_c|r), P(I_c)$ possible via different choices of patterns, assuming full knowledge of the scene, and justifies adapting the patterns based on known world model and uncertainty. The viewpoint we take here treats pattern selection as an experimental design problem, as we will demonstrate. We note that theoretically, an infinitely dense pattern would yield the most variation. Lacking an estimate of the uncertainty in depth, structured-light scanner pattern are often designed to accomodate multi-scale search over the range, or other techniques to guarantee the range is uniquely identified, see for example standard grey-code striped light. Our simulation highlights the role of choosing the patterns according to the scene model and uncertainty — in most cases of structured-light scanning, we already have an estimate of the scene parameters from previous time frames, and this prior knowledge should be used.

## III. Estimating Uncertainty in 3D Scanners

We present two important cases of estimating mutual information gain for pattern selection in structured-light scanners. In each, we consider inference over different subsets of variables, and the mutual information between them and the observed images. Differing assumptions on the fixed/inferred variables and dependency structure in the image formation model lead to different algorithms for MI estimation given as Algorithms 1 and 3.

An important observation is that given the pose, range measurements and camera image pixel values can be approximated as an independent estimation problem per-pixel (here we model the effect of surface self-occlusions as noise). This provides an efficient and parallelizable estimation procedure for the case of range estimation. This assumption has been exploited in plane-sweeping stereo, and inference of structured light with priors [39]. We now utilize it for MI estimation. We note that even where the inter-pixel dependency is not negligible, we can compute an upper bound for the information gain. For example, for the case of pose and range estimation we obtain

$$\begin{aligned}\mathcal{I}(I_c;\Theta,r) =& H(I_c) - H(I_c|\Theta,r) \leq \\ & \sum_x H(I_c(x)) - \sum_x H(I_c(x)|\Theta,r(x)) \triangleq \hat{\mathcal{I}}(I_c;\Theta),\end{aligned} \tag{6}$$

where $\hat{\mathcal{I}}$ is the *pixel-wise mutual information* between the sensor and the inferred parameter.

### A. Range Image MI Estimation

We start with the simple, yet instructive, case of estimating mutual information between the scene geometry and the observed images given a known pose, for predetermined set of illumination patterns. Here, inference is over $G_l$ as represented by the range at each camera pixel $r \equiv r(x)$. We assume a Gaussian prior for $a$ and $b$, with means 1,0, and standard deviation large enough to be uninformative.

We compute the pixel-wise mutual information individually and sum the results. In this subsection, we assume a deterministic choice of pose; the patterns are deterministic throughout

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2018.2830181, IEEE Transactions on Computational Imaging



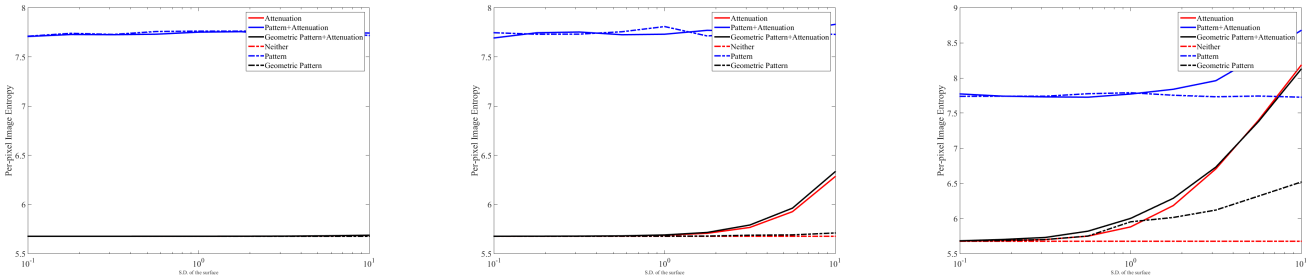


Fig. 4. Left-to-right: The per-pixel image entropy as at a range of $1000cm$, $100cm$ and $20cm$, for different levels of surface range variation. Different curves in each plot represent the predicted entropy with and without range attenuatation, and with either no pattern, projected pattern intensity change based on the scene geometry only, or a pattern drawn directly in the image plane, representing a maximally varying pattern. While attenuation can have an effect, it is small compared to the pattern's disparity contribution. The maximal entropy due to changing the patterns is higher than the mere change in pattern due to different intersection with the surface, suggesting a stronger possible signal for infering the range for patterns designed with the specific range uncertainty in mind. As can be seen in the leftmost plot, the small disparity of the projector and long range make the geometry-based variation in the pattern ineffective, which demonstrates the limitations of structured light scanners at a long range with limited projector and camera resolutions and small projector-camera baseline.

the paper, and hence omitted from the notation for $\mathcal{I}$. The mutual information between $I_c$ and $G_l$ given $\theta$,$I_p$ is given by

$$\hat{\mathcal{I}}\left(I_c; G_l|\theta\right) = \sum_x \mathcal{I}\left(I_c(x); r(x)|\theta\right) \tag{7}$$
$$= \sum_x E_{I_c(x),r(x)|\theta}\left[\log \frac{p(I_c(x)|r(x),\theta)}{p\left(I_c(x)|\theta\right)}\right].$$

While computing $p(I_c(x)|r(x),\theta)$ is straightforward, we are still forced to estimate $p(I_c(x)|\theta)$, which can be done by marginalizing over $r$ according to our posterior estimates,

$$p(I_c(x)|\theta) = E_r[p\left(I_c(x)|r(x),\theta\right)]. \tag{8}$$

For each sample of $\theta, r$, we can then compute the log of the likelihoods ratio, and integrate it. We note the existence of alternatives such as using GMMs or Laplace approximations, for efficient implementation.

We perform one sampling loop in order to estimate $p(I_c|\theta)$. We then use another set of samples in order to estimate $\hat{\mathcal{I}}\left(I_c; G_l|\theta\right)$. Algorithm 1 describes computation of the MI gain for frame $T$.

Since $a, b, \eta^{(0..T)}$ are all are assumed to be Gaussian conditioned on $r$, $p\left(a, b, I_c^{(t)}|I_p^{(0...t)}, I_c^{(0...t-1)}\right)$ is Gaussian. We can compute the pdf of $a, b$ and $I_c^{(T)}$ given $I_p^{(0...T)}$ and $I_c^{(0...T-1)}$, by conditioning on each image $t$ at a time, computing $p\left(a, b, I_c^t|I_c^{0..t-1}\right)$ for each $t = 0..T$ iteratively. This allows fast computation on parallel hardware such as graphics processing units (GPUs), without explicit matrix inversion or other costly operations at each kernel.

**Online Range Estimation:** This observation of structure also lends itself to approximate online updates of the range estimation after each pattern selection and projection, forming the update step in Figure 1, by updating a posterior distribution of each pixel's range value. This could be done either explicitly for a set of range values, or by estimating posterior likelihoods after each image and then refitting a Gaussian to these estimates, as described in Algorithm 2. In this algorithm, the update/correction step, once a new pattern was projected, is done by updating per $r$ value the distribution of $a, b$, assumed to be Gaussian. This essentially does a plane-sweeping update per new image received, yet tracks the effect of previously seen images on the estimated photometric coefficients' distribution. A Gaussian assumption of $a, b$, and the Gaussian assumption on $\eta^T$ allows a simple analytic update by looking at $a, b, I_c$ and using the expression for conditional distributions in a multivariate Gaussian vector. Full treatment of efficient iterative approaches for time-multiplexed structured light range estimation are beyond the scope of this paper.

**Algorithm 1** MI estimation / pattern selection for range image

**Input**: a set of candidate patterns $\hat{I}_p$ to be projected as $\hat{I}_p^t$, past projected patterns $I_p^{1...(t-1)}$ and observations $I_c^{1...(t-1)}$

**Output** pattern $p_{max}$ that will maximize average information gain on range image.

**for** pattern $p$, in each pixel $x$ **do**
  **for** samples $i = 1, 2, \ldots, N_{hist}$ **do**
    Sample a range value for $x$ according to $p(r)$.
    Raytrace $I_p^{1..(t)}$, sample $I_c^{1..(t-1)}$. Compute the statistics of $a, b, I_c^{(t)}$ conditioned on previous image measurements $I_c^{1..(t-1)}$.
    Compute probability $p(I_c^{(t)}|r)$.
    Update the estimated per-pixel histogram, $p(I_c^{(t)})$.
  **end for**
  **for** samples $i = 1, 2, \ldots, N_{MI}$ **do**
    Draw a new range value for $x$ according to a proposal distribution $p(r)$.
    Raytrace $I_p^{1..(t)}$, sample $I_c^{1..(t-1)}$. Compute the statistics of $a, b, I_c^{(t)}$ conditioned on previous image measurements $I_c^{1..(t-1)}$.
    Compute probability $p(I_c^{(t)}|r)$, estimate $\log\left(\frac{p(I_c^{(t)}|r)}{p(I_c)}\right)$.
    Update the estimated mutual information.
  **end for**
**end for**
Pick pattern $p_{max}$ with maximum average MI over the image. Set $\hat{I}_p^t$ to be pattern $p_{max}$.

**Algorithm 2** Open-loop-with-feedback adaptive range estimation

**Input**: a set of candidate patterns, initial range and intensity estimates

**Output** update range estimates at each time $t$.

**for** $t = 1, 2, \ldots, \ldots$ **do**
  Compute MI values using Algorithm 1, selecting optimal pattern $I_p^t$
  Project selected pattern $I_p^t$, acquire image $I_c^t$
  Use plane sweeping over $r$ to update posterior probability of $r, a, b$:
  **for** $r = 1, 2, \ldots, \ldots$ **do**
    Update the posterior distribution of $a, b$, assuming Gaussian distribution with previously estimated values as prior.
  **end for**
**end for**

### *B. Pose MI Estimation with Structured-Light*

A second important case we explore is typical of pose estimation problems, where we try to infer a low-dimensionality latent variable set with global influence, in addition to range uncertainty. In 3D pose estimation, we usually estimate $\Theta$ given a model of the world $G$.

In visual SLAM, $G, A, A_l$ are commonly used to infer $\Theta, G_l$, either as online inference [3], or in batch-mode [43], where usually a specific function of the input (feature locations from different frames, or correspondence estimates) is taken. In depth-sensor based SLAM, the range sensors obtain a measurement $G_l$ under some active illumination. $\Theta$ is then approximated from $G, G_l$. Unlike the usual usage of 3D sensors for generating a range map and then using it within a 3D SLAM algorithm, this allows direct posterior update of the pose parameters given the projector patterns and acquired camera image, before a complete range image can be formed.

We now describe computation of the MI between the pose and the images. As before, we parameterize $G_l$ by $r(x)$, and given $(\Theta, r)$ we re-establish a correspondence between $I_p$ and $I_c$. This is done by computing a back-projected point $x_j^3$ (denoting it is a 3D point), transforming it according to $\Theta$ to get $\bar{x}_j^3$, and projecting $\bar{x}_j^3$ onto the camera and projector image. A similar situation would arise where inferring a class variable, where instead of merely inferring $\Theta$ we also infer a categorical variable $C$ that determines the class of the observed object. Here too, we can still use the following observations: (i) given the pose parameters, the problem can still be approximated as a per-pixel process – this assumption underlies most visual servoing approaches. (ii) the pose parameter space is low-dimensional and can be sampled from, as is often done in particle filters for pose estimation. We can therefore write

$$\mathcal{I}\left(I_c(x); \Theta | G_l\right) = E_{I_c(x),\Theta,r}\left(\log \frac{P(I_c(x)|\Theta)}{P\left(I_c(x)\right)}\right), \quad (9)$$

where as before, $P\left(I_c(x)|\theta\right)$ is computed by marginalization over $r(x)$. This procedure is detailed as Algorithm 3. When computing $p(I_c(x)|\Theta)$, $p(\Theta)$ can be conditioned on previous observations, and sampled from the current uncertainty estimate for the pose and range.

**Algorithm 3** MI estimation / pattern selection for pose estimation

**Input**: a set of candidate patterns $\hat{I}_p$ to be projected as $\hat{I}_p^t$, past projected patterns $I_p^{1\ldots(t-1)}$ and observations $I_c^{1\ldots(t-1)}$

**Output** pattern $p_{max}$ that will maximize information gain on pose $\theta$.

**for** pattern $p$, in each pixel $x$ **do**
  Set $I_p^t$ to be pattern $p$
  **for** samples $i = 1, 2, \ldots, N_{hist}$ **do**
    Draw pose sample $\theta_i$, compute $T_{\theta_i}$
    **for** range sample $r(x)$, from $N_r$ samples **do**
      Back-project $x^3$, compute $\bar{x}^3 = T_{\theta_i, r}\left(x\right)$.
      Project $\bar{x}^3$ and sample $I_p^{1\ldots t}$, sample $I_c^{1\ldots(t-1)}$.
      Compute the statistics of $a, b, I_c^t$ conditioned on previous image measurements $I_c^{1..(t-1)}$ and $r(x)$.
      Update the estimated per-pixel histogram, $P(I_c^t)$
    **end for**
  **end for**
  **for** samples $i = 1, 2, \ldots, N_{MI}$ **do**
    Draw pose sample $\theta_i$ and associated transformation $T_{\theta_i}$
    **for** range sample $r(x)$, from $N_r$ samples **do**
      Back-project $x^3$, compute $\bar{x}^3 = T_{\theta_i, r}\left(x\right)$.
      Project $\bar{x}^3$ and sample $I_p^{1..(t)}$, sample $I_c^{1..(t-1)}$.
      Compute $a, b, I_c^t$ estimates conditioned on previous image measurements $I_c^{1..(t-1)}$, and $r(x)$.
      Estimate $log\left(\frac{P(I_c^t|a,b,I_p,T_{\theta_i})}{P(I_c^t)}\right)$.
      Update the mutual information gain estimate.
    **end for**
  **end for**
**end for**
Pick pattern $p_{max}$ with maximum MI sum over the image. Set $\hat{I}_p^t$ to be pattern $p_{max}$

We note that when sampling the pose, different variants of the range images can be used, allowing us to marginalize w.r.t. range uncertainty as well.

When sampling a conditioned image model per pixel, collisions in the projected pixels can occur. While these can be arbitrated using atomic operations on the GPU, the semantics of write hazards on GPUs are such that invalid pixel states can be avoided, and in practice there was no need to use atomic operations. Furthermore, to allow efficient computation on the GPU, we must consider memory access patterns. In our implementation we compute proposal image statistics given $\theta$, and then aggregate the contribution into the accumulators for the mutual information per pixel.

**Online Pose Estimation:** In the case of range estimation, one can use the observed images directly to update a pose estimate, without reconstructing a range image, similar the approach shown for range estimation, in Algorithm 2. For the case of a known map, this simplifies the update step in Figure 1 to updates of the pose based on the recently acquired

observations, based on the likelihood of the images, as described Subsection III-A, by conditioning on new observations $I_c^{(t)}$ as they arrive. For the case where both the map and the pose are unknown, this can be done, for example, by a Rao-Blackwellized particle filter [44]. However such estimation is beyond the scope of this paper.

**Extension to classification** we could incorporate categorical variables, including object classes as part of $\Theta$. This requires merely changing lines 4,14, in Algorithm 3 to sample a distribution over $\bar{x}_j^3(\theta, C, r)$ instead of $\bar{x}_j^3(\theta, r)$. This allows us to choose patterns for object classification tasks, which is beyond the scope for this paper.

While sampling the full space of appearance and range per-pixel is computationally expensive, running the algorithm without any optimizations on a GPU takes approximately one second on an Nvidia Quadro K2000.

## IV. Numerical Results

We conducted several experiments aimed at giving an intuition for the approach proposed in this paper, and demonstrating its utility, with several choices of projector patterns and scenes, including striped light and smoothed random patterns with both Gaussian foviated, and striped-masked modulation. In terms of the relevant sets of variables, we have focused on range sensing and pose estimation. Our priors for photometric and range variables are set to be uninformative unless otherwise stated.

### A. Pattern Choice for Range Sensing

We first demonstrate the setup used. For pattern libraries we used a set of random patterns generated by smoothing i.i.d. Gaussian noise with Gaussian filters of various scales, and striped patterns of the sort used for gray-code structured-light. While these are chosen to represent both well structured and weakly structured pattern libraries, other choices of patterns [45] can be used. They are shown in Figures 6 and 10, respectively. We used as test objects both fabricated models with various scales of features, see Figure 6, and coated/raw wooden art models. The PointGrey Grasshopper II camera and TI LightCrafter projector used are shown in Figure 1. Pixel noise standard deviation was about $2.5/255$ for most experiments. We validate the use of the smoothed Gaussian patterns for reconstruction in Figure 5, demonstrating the decrease in the average range L2 error measured as we use more patterns for reconstruction. We use the reconstruction from a set of 120 patterns as a ground-truth estimate, making the assumption that the reconstruction is an unbiased estimator, so that reconstruction using all patterns is considered a ground-truth.

In Figure 6 we show the MI gain collected over the scene, averaged over 50 random pattern sequences. The amount of information gained from the patterns decreases as we add more patterns, as expected with MI, and surfaces that are well-illuminated and frontal-facing having faster uncertainty reduction. We look at the average MI gain per pattern over various random sequences of patterns, in Figure 7. We highlight several interesting cases. The first case (which often occurs in practice) assumes high uncertainty of the range or the appearance coefficients. The second and third cases involve less and more certainty in the appearance coefficients respectively. The fourth case involves having a good initial guess (std. of 7mm) for the range. As expected, the certainty of the appearance coefficients increases the MI between the images and the range. Having a good range prior decreases the amount of information gained per frame and the overall MI.

We then proceed to perform selection according to MI gain based on the proposed model. Although we perform greedy (pattern at a time) selection, there are bounds guaranteeing the performance of a greedy vs. optimal selection of the whole pattern sequence – see [31] for such bounds and the relevant terminology. In our test we initialize each attempt from a pair of randomly chosen patterns. At each turn we try ten randomly chosen patterns and compute their image-range MI. We pick the the most informative pattern, and contrast this with a random pattern selection. The MI gains for two scenes are measured in Table I, collected over ten instantiations of the selection process.

In one scenario, we modulate the patterns by spatial bands in the projector's image plane: 14 bands in the $x$ and in the $y$ directions with 15 random textures instantiations for each band, see example in Figure 8(a). We note that modulating patterns is physically feasible in a real system, and is warranted when overcoming projector intensity / background illumination limitations, as discuss, for example, in [46]. From these we greedily select patterns in ten sequences, and unify them into 69 unique patterns. The patterns are mostly those that illuminate the region of interest, as expected by their high MI gain. The region of interest is defined as the silhoutte of an object (the hand) in the image, and serves as a relevant set to focus the sensing on, with an intuitive definition. A similar test was done with patterns modulated by an exponentially, radially decreasing envelope, illuminating local regions of the projector field of view at each time (see Figure 8(d)). 20 random patterns are taken, modulated by 15 random locations. Of these, 65 are selected after removing repetitions. Here the region of interest was the mannequin. We use these pattern sets to reconstruct the range image, and compare to randomly choosing the same number of patterns. Qualitatively, the selected patterns often illuminated parts of the objects which were poorly reconstructed, as expected. As we show in Figure 8, we get significantly more accurate reconstruction compared to random selection—18.9mm RMS, compared to 24.1mm RMS for the hand example, and 51.3mm compared to 59.1mm in the mannequin example. This demonstrates the usefulness of our selection criteria when judged by reconstruction accuracy.

Finally, in order to demonstrate that greedy selection improves reconstruction, on average, per pattern selection, we perform ten greedy selection steps, selecting a single pattern out of ten randomly drawn ones, and demonstrate the resulting reconstruction. We take striped gray-code patterns modulated by radially-decreasing piece-wise smooth masks, centered at various locations, for a total of 240 patterns. The results of adding patterns at random vs. greedy selection show that even when we do not yet have reasonable reconstruction, greedy

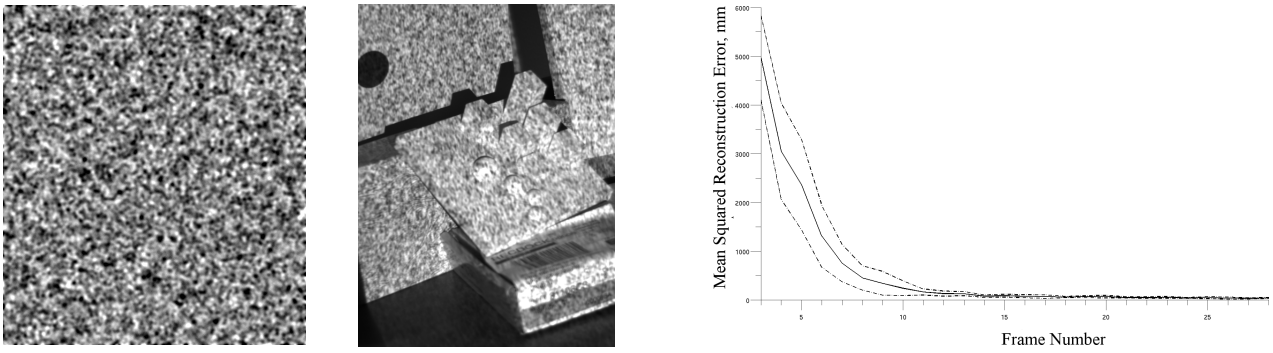


Fig. 5. Left-to-right: a projected Gaussian-smoothed pattern, a captured image, average reconstruction error as a function of the number of patterns used. Dashed lines mark the standard deviation over pattern sequences.

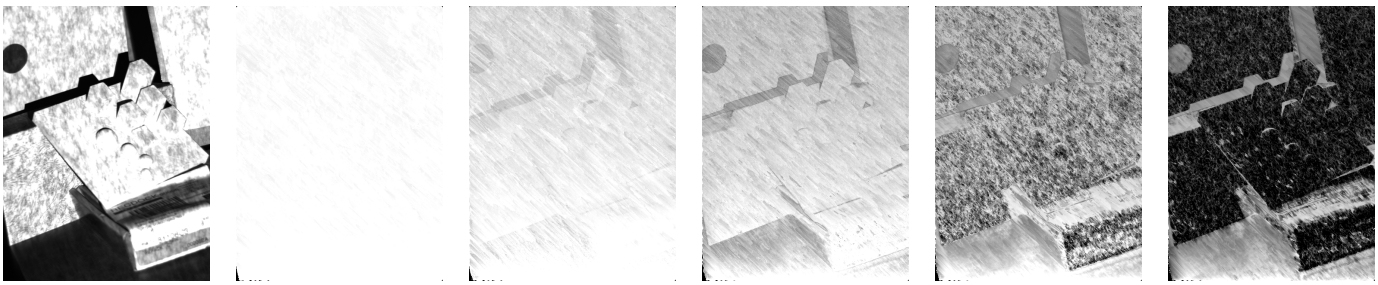

Fig. 6. Left-to-right: An indicator image of reflected patterns amplitudes, followed by the mutual information between the image and the range, for random Gaussian-smoothed patterns. The initial patterns are dominated by well-illuminated areas, followed by poorly-illuminated areas (a secondary trend relates to the surface illumination angle).

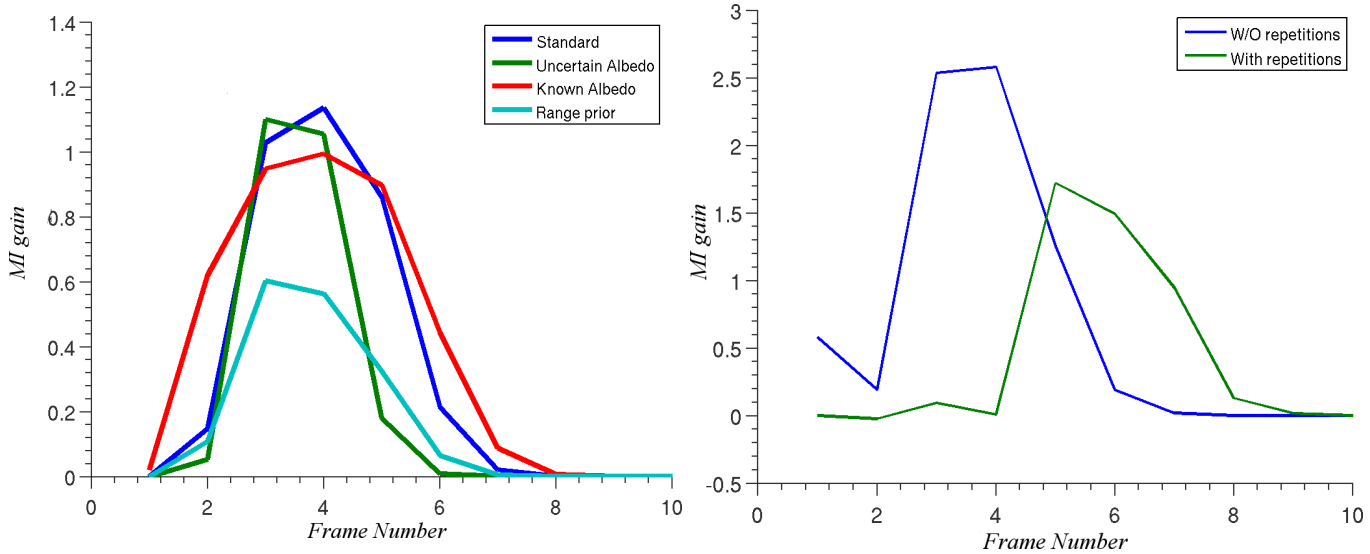


Fig. 7. Left: Mutual information gain under different assumptions on the scene: Blue line - the standard case of large range and albedo uncertainty of $\sigma_r = 300mm$, $\sigma_a = 3, \sigma_b = 300$. Red line - $\sigma_a = 30, \sigma_b = 3000$ (high uncertainty of the appearance). Green line - $\sigma_a = 0.3, \sigma_b = 20$ (strong prior on the appearance). Cyan line - $\sigma_r = 7mm$ (low initial uncertainty of the range). Given a good prior on the nuisance parameters of the albedo, range is estimated more quickly in terms of frames. Given a strong range prior, the region does not require as many patterns for estimation, and overall MI gain is smaller. Right: Blue - information gain for a set of different patterns. Green - where only half of the patterns are shown, but they are repeated twice. The information gain is much lower in the second case.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2018.2830181, IEEE Transactions on Computational Imaging



| | Hand Mean MI, Greedy | STDev, Greedy | Mean MI, Random | STDev Random | Mannequin Mean MI, Greedy | STDev, Greedy | Mean MI, Random | STDev Random |
|---|---|---|---|---|---|---|---|---|
| Step 1 | 0.4168 | 0.2820 | 0.1267 | 0.0957 | 0.1688 | 0.0561 | 0.0756 | 0.0504 |
| Step 2 | 0.7904 | 0.2803 | 0.3263 | 0.2457 | 0.2404 | 0.0694 | 0.0653 | 0.0484 |
| Step 3 | 0.8129 | 0.1820 | 0.2686 | 0.1694 | 0.3030 | 0.0916 | 0.1199 | 0.0695 |
| Step 4 | 0.6232 | 0.1125 | 0.2125 | 0.1591 | 0.2911 | 0.0806 | 0.0997 | 0.0939 |
| Step 5 | 0.1562 | 0.0995 | 0.0903 | 0.1317 | 0.1334 | 0.0450 | 0.0744 | 0.0656 |
| Step 6 | 0.0229 | 0.0264 | 0.0376 | 0.0433 | 0.0400 | 0.0232 | 0.0482 | 0.0486 |

TABLE I

MI GAIN STARTING FROM TWO RANDOM PATTERNS, WHEN USING GREEDY SELECTION, COMPARED TO RANDOM PATTERN SELECTION. RESULTING MI GAINS ARE SHOWN FOR THE HAND AND MANNEQUIN EXAMPLES FROM FIGURE 8. OUR MI-GREEDY APPROACH OBTAINS A LARGER INFORMATION GAIN, AND DOES SO FASTER (IN FRAME COUNTS) THAN A RANDOM ORDERING OF FRAMES.

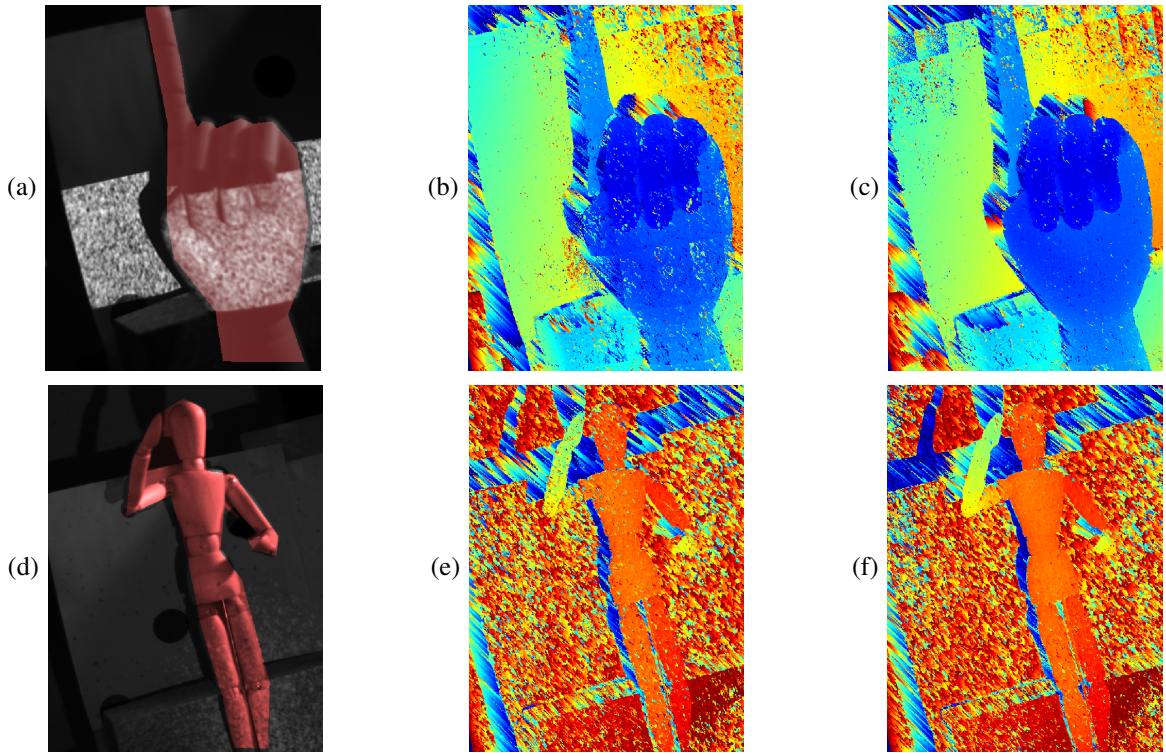


Fig. 8. Left-to-right: camera image with a projected pattern on the marked object (red overlay marks the mask used for MI integration). The area covered by the mask received significantly more pattern coverage and the reconstruction with these bands is considerably better than random selection. Top: reconstruction with a random set of 69 bands (range RMS=24.1mm) vs. reconstruction with the set of 69 bands selected by a greedy selection (range RMS=18.9mm). Bottom: reconstruction with a random set of 65 blobs (range RMS=59.1mm) - random vs. greedy.

selection according to MI improved L2 reconstruction error. Despite the fact the L2 reconstruction error does not directly coincide with MI, we show that computing MI gain according to our model results early on in the reconstruction sequence in improved reconstruction results, as shown in Figure 9. For example, the depth reconstruction error obtained by 10 random patterns is obtained with less than six patterns in the greedy case, representing a 40% speedup.

### *B. Pattern Choice for Pose Estimation*

In Figures 10–13 we show computed per-pixel MI between a new camera image and the pose, assuming a highly certain range image, as estimated by Algorithm 3. We start in Figure 10 with a synthetic case where the results are easy to interpret, with a scene made of a single large corner. The pattern set for this experiement is the standard gray-code striped patterns, shown in the first row. We assume only translational uncertainty; we leave reasoning about the full SE(3) pose space to future work as it is less instructive. We use stripes going from coarse to fine, stopping at a pattern of four pixels stripe width in the projector image plane. At this phase, the appearance coefficients $A, G$ are well estimated. In this example the camera and the projector are facing the $z$ direction, and in front of them there is a large smoothed corner. We compare a case of uncertainty in the $xy$ plane, to that of uncertainty in the $z$ plane in terms of the pixel-wise MI gain. The large sloped corner and the edges are the main source of uncertainty reduction in $xy$ since the rest of the scene is planar. In the $z$ uncertain case, the full image is informative to the same extent. The intermediate case is a mix between the two, as expected.

For pattern selection, in Figure 11 we demonstrate pattern choice according to the proposed criteria for choosing patterns in a structured-light scanner. This shows that for an unknown

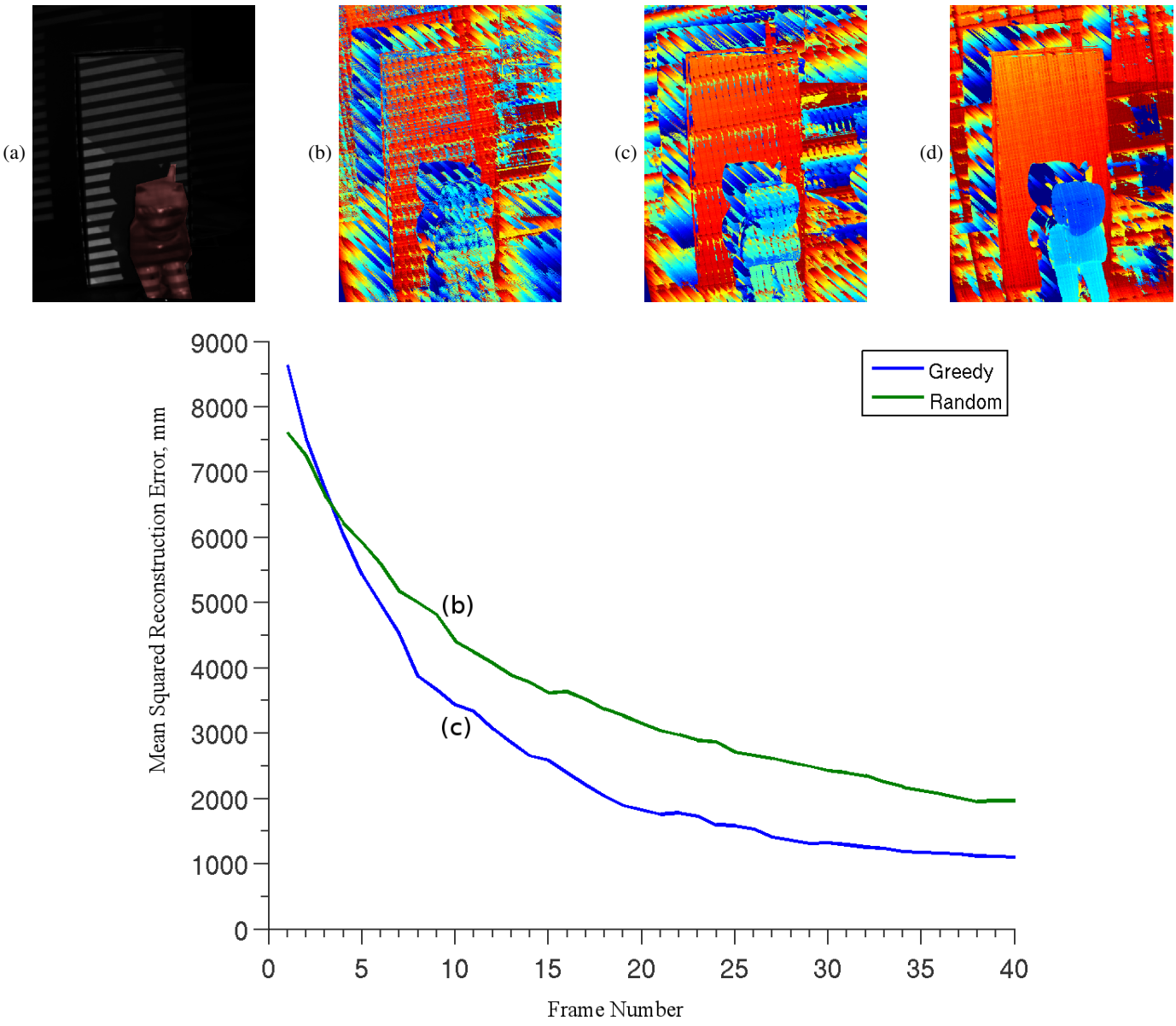


Fig. 9. Top, Left-to-right: camera image with a projected pattern on the marked object (MI integration mask shown in red), range image of the scene as reconstructed by random selection of 10 patterns, greedy selection of 10 patterns and the full set of 240 patterns, reconstruction squared error as a function of the number of patterns addded, averaged over 20 trials. Bottom: error between partial frame sets reconstruction and the full 240 frames reconstruction, where frames are added at random (green) or using our approach (blue). Greedy selection based on our model improves reconstruction results with significantly fewer frames (50%), as demonstrated by Subfigures (b) and (c).

pose information can be obtained from edges and corners; given a reasonable model of the scene, we can use mutual information to suggest which pattern to use to project only informative parts of the scene. The patterns chosen consist of a striped pattern projected only along a partial band of the projector screen. Figure 12 demostrates a different set of patterns, of stripes modulated by a Gaussian mask, allowing to focus a pattern in a small region, which is important in practical applications. As can be seen, the top-ranking patterns are those that illuminate edges in the scene, which should give us high uncertainty reduction. MI for pose estimation can also be seen with real scenes. In Figure 13 we show pixelwise pose estimate for Gaussian smoothed patterns. The most informative pixels are edges and sloped areas, where the perceived projector intensity changes rapidly as a function of the pose.

## V. Conclusions

In this paper we present a novel information-driven approach to planning into 3D sensors at the sensor level. We demonstrate how different uncertainty estimates and sensor models lead to different criteria for pattern selection. Future work includes the completion of a prototype scanner based on the proposed approaches. This decision-theoretic approach where action choice is identified with pattern selection in structured-light easily extends to other reconstruction techniques such as depth-from-focus (see for example [47]) and time-of-flight [48], [49]. We intend to explore these in future

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2018.2830181, IEEE Transactions on Computational Imaging



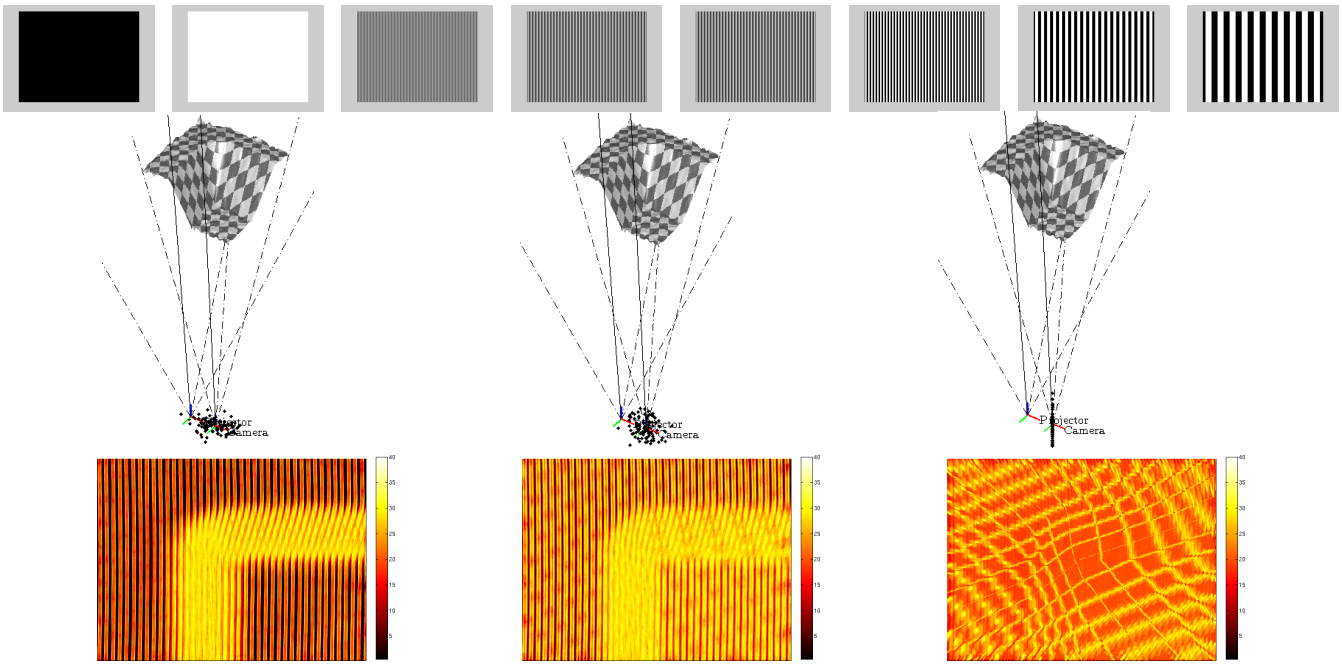


Fig. 10. Per-pixel information gain for the case of initial uncertain scanner position. Left-to-right, top row: a set of patterns used for 3D sensing for pose estimation. Middle row: a rendering of the scene with sensor pose samples (black dots) in 3 scenarios, and the fields of view of the projector and camera. Bottom row: pixelwise mutual information estimates: with high uncertainty in the x-y plane of the scanner, uncertainty in x-y-z, and z-only uncertainty in scanner position. Yellow and red marking high and low information gain, respectively. Surfaces at sharp angles to the projector and camera provide greater uncertainty reduction in the x-y directions, whereas for uncertainty in the z axis, all surfaces are informative.

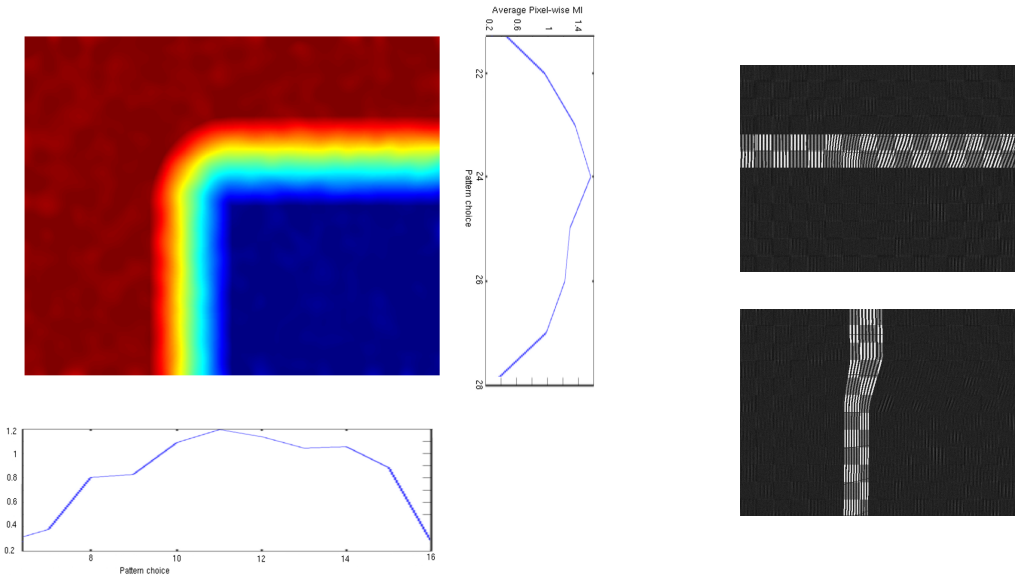


Fig. 11. Left: the depth image and the MI scores of vertical and horizonal stripe masks of the patterns with respect to pose estimation in the $xy$ plane. Right: the top-scoring horizontal and vertical patterns, as seen when projected onto the scene. As can be seen, the patterns that were selected are the ones illuminating the edges and corner.

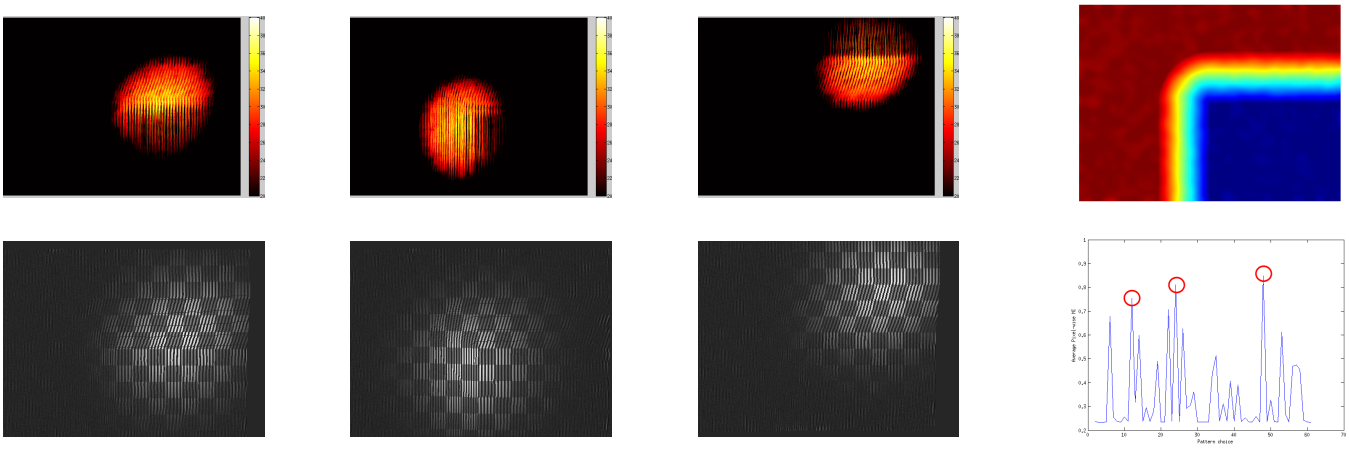

Fig. 12. Left-to-right, top row: the top 3 selected masks from a set of 60 masks, and the range image. Bottom row: a MAP estimated images for the 3 masks, used when estimating the MI for each pattern, followed by the average MI scores for the patterns. Red circles mark the patterns shown.

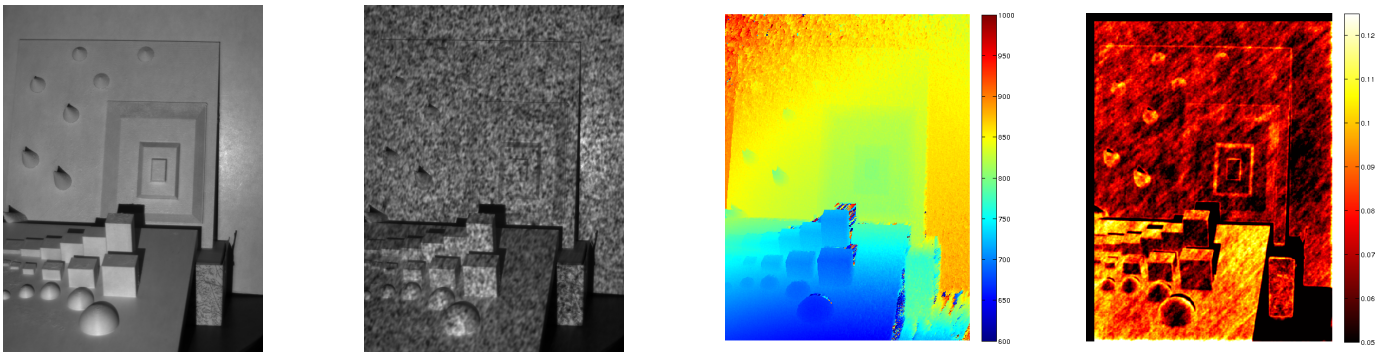

Fig. 13. Left-to-right, top row: an image of the scene, one of the projected patterns as capture, the range image, the pixelwise mutual information with respect to the pose, which initial uncertainty in the camera's $xy$ plane. The main informative areas are the cones, and regions that face the $x, y$ directions.

work. While we show several standard pattern libraries in the results section, selecting the optimal library is not the focus of this paper, even though it heavily interacts with the problem we address. Indeed, many of the libraries are designed so that a complete scanning cycle is efficient and effective. However, once we look at the online pattern selection problem given a changing scene, other patterns may be relevant. Exploration of the interaction between different nuisance factors in the online case, and the efficiency gap between a fixed-order scanning plans and adaptive plans, or the complementarity of different pattern sets in dynamic scenes is left for future work.

**Acknowledgements**

The authors thank Christopher Dean for general and helpful discussions. Support for this research has been provided by ONR MURI N00014-09-1-1051, N00014- 11-1-0688, and ARO MURI W911NF-11-1-0391. We are grateful for this support.

## REFERENCES

[1] 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*, Jun. 2011.

[2] J. J. Leonard and H. F. Durrant-Whyte, "Simultaneous map building and localization for an autonomous mobile robot," in *IROS*, 1991, pp. 1442–1447.

[3] S. Thrun, "Robotic mapping: A survey," in *Exploring Artificial Intelligence in the New Millenium*, G. Lakemeyer and B. Nebel, Eds. Morgan Kaufmann, 2002.

[4] M. F. Fallon, H. Johannsson, and J. J. Leonard, "Efficient scene simulation for robust Monte Carlo localization using an RGB-D camera," in *ICRA*, May 2012.

[5] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view RGB-D object dataset," in *ICRA*, 2011.

[6] P. K. Nathan Silberman, Derek Hoiem and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *ECCV*, 2012.

[7] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press, 2005.

[8] K. Ikeuchi and J.-C. Robert, "Modeling sensor detectability with VANTAGE geometric sensor modeler," Carnegie-Mellon University.Computer science. Pittsburgh (PA US), Tech. Rep. CMU-CS-89-120, 1989.

[9] L. Paletta, M. Prantl, and A. Pinz, "Learning temporal context in active object recognition using bayesian analysis," in *ICPR*, vol. 1, 2000, pp. 695–699.

[10] J. Denzler and C. Brown, "Information theoretic sensor data selection for active object recognition and state estimation," *IEEE-TPAMI*, vol. 24, no. 2, pp. 145–157, Feb 2002.

[11] F. S. Cohen and D. Cooper, "A decision theoretic approach for 3-d vision," in *CVPR*, Jun 1988, pp. 964–972.

[12] Y. Zhang, Z. Xiong, P. Cong, and F. Wu, "Robust depth sensing with adaptive structured light illumination," *J. Vis. Comm. and Image Representation*, vol. 25, no. 4, pp. 649–658, 2014.

[13] L. Valente, Y.-H. R. Tsai, and S. Soatto, "Information-seeking control under visibility-based uncertainty," *Journal of Mathematical Imaging and Vision*, vol. 48, no. 2, pp. 339–358, 2014.

[14] M. Sheinin and Y. Y. Schechner, "The next best underwater view," in *CVPR*, 2016, accepted.

[15] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An*

*Introduction*. Cambridge, MA: MIT Press, 1998. [Online]. Available: citeseer.ist.psu.edu/sutton98reinforcement.html

[16] F. M. Wahl, "A coded light approach for depth map acquisition," in *Mustererkennung 1986*. Springer, 1986, pp. 12–17.

[17] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," *Pattern Recognition*, vol. 43, no. 8, pp. 2666 – 2680, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S003132031000124X

[18] D. S. Levine and J. P. How, "Sensor selection in high-dimensional Gaussian trees with nuisances," in *NIPS*, 2013, pp. 2211–2219.

[19] G. Rosman, D. Rus, and J. W. Fisher III, "Information-driven adaptive structured-light scanners," in *CVPR*, 2016.

[20] R. A. Howard, "Information value theory," *IEEE Trans. Systems Science and Cybernetics*, vol. 2, no. 1, pp. 22–26, Aug 1966.

[21] Y. Fumie, "Value of information analysis in environmental health risk management decisions: Past, present, and future," *Risk analysis : an international journal.*, 2004,.

[22] D. S. Levine, "Information-rich path planning under general constraints using rapidly-exploring random trees," Master's thesis, MIT, Dept. of Aero.-Astro., June 2010. [Online]. Available: http://acl.mit.edu/papers/LevineSM.pdf

[23] B. J. Julian, M. Angermann, M. Schwager, and D. Rus, "Distributed robotic sensor networks: An information-theoretic approach," *I. J. Robotic Res.*, vol. 31, no. 10, pp. 1134–1154, 2012.

[24] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong, "Model-driven data acquisition in sensor networks," in *VLDB*. VLDB Endowment, 2004, pp. 588–599.

[25] J. L. Williams, J. W. Fisher III, and A. S. Willsky, "Approximate dynamic programming for communication-constrained sensor network management," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 3995–4003, August 2007. [Online]. Available: publications/papers/williams07a.pdf

[26] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *JMLR*, vol. 9, pp. 235–284, Jun. 2008.

[27] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *Signal Processing Magazine, IEEE*, vol. 19, no. 2, pp. 61 –72, mar 2002.

[28] E. Ertin, J. W. Fisher III, and L. C. Potter, "Maximum mutual information principle for dynamic sensor query problems," in *IPSN*. Springer, Feb 2003, pp. 558–561.

[29] C. Kreucher, K. Kastella, and A. Hero, "Sensor management using an active sensing approach," *Signal Processing*, vol. 85, no. 3, pp. 607–624, 2005.

[30] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Athena Scientific, 2000.

[31] J. L. Williams, J. W. Fisher III, and A. S. Willsky, "Performance guarantees for information theoretic active inference." *JMLR*, vol. 2, pp. 620–627, 2007. [Online]. Available: http://dblp.uni-trier.de/db/journals/jmlr/jmlrp2.html#WilliamsFW07

[32] L. V. Gool and T. P. Koninckx, "Real-time range acquisition by adaptive structured light," *IEEE-TPAMI*, vol. 28, no. 3, pp. 432–445, March 2006.

[33] Q. Li, M. Biswas, M. Pickering, and M. Frater, "Dense depth estimation using adaptive structured light and cooperative algorithm," in *CVPR Workshops*, June 2011, pp. 21–28.

[34] X. Maurice, P. Graebling, and C. Doignon, "Real-time structured light coding for adaptive patterns," *Journal of Real-Time Image Processing*, vol. 8, no. 2, pp. 169–178, 2013.

[35] M. O'Toole, S. Achar, S. G. Narasimhan, and K. N. Kutulakos, "Homogeneous codes for energy-efficient illumination and imaging," *ACM Trans. on Graphics*, vol. 34, no. 4, pp. 35:1–35:13, 2015.

[36] S. K. Nayar, V. Branzoi, and T. E. Boult, "Programmable Imaging: Towards a Flexible Camera," *Int. J. of Computer Vision*, Oct 2006.

[37] S. Soatto, "Steps towards a theory of visual information: Active perception, signal-to-symbol conversion and the interplay between sensing and control," *CoRR*, vol. abs/1110.2053, 2011.

[38] J. P. Tardif and S. Roy, "A MRF formulation for coded structured light," in *3DIM*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 22–29.

[39] G. Rosman, A. Dubrovina, and R. Kimmel, "Sparse modeling of shape from structured light," in *3DIMPVT*. Washington, DC, USA: IEEE Computer Society, 2012, pp. 456–463. [Online]. Available: http://dx.doi.org/10.1109/3DIMPVT.2012.20

[40] L. Yu, S. K. Yeung, Y. Tai, and S. Lin, "Shading-based shape refinement of RGB-D images," in *CVPR*, 2013, pp. 1415–1422.

[41] R. Or-El, G. Rosman, A. Wetzler, R. Kimmel, and A. M. Bruckstein, "RGBD-fusion: Real-time high precision depth recovery," in *CVPR*, 2015, pp. 5407–5416.

[42] P. R. Cohen and A. E. Howe, "Toward ai research methodology: three case studies in evaluation," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 3, pp. 634–646, May 1989.

[43] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *ISMAR*, Nara, Japan, Nov. 2007.

[44] G. Grisetti, C. Stachniss, and W. Burgard, "Improved techniques for grid mapping with rao-blackwellized particle filters," *IEEE transactions on Robotics*, vol. 23, no. 1, pp. 34–46, 2007.

[45] M. Gupta and S. K. Nayar, "Micro phase shifting," in *CVPR*. IEEE, 2012, pp. 813–820.

[46] M. Gupta, Q. Yin, and S. K. Nayar, "Structured light in sunlight," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 545–552.

[47] X. Yuan, P. Llull, X. Liao, J. Yang, G. Sapiro, D. J. Brady, and L. Carin, "Low-cost compressive sensing for color video and depth," in *CVPR*, 2014, accepted.

[48] J. C. Howell, G. A. Howland, A. Kirmani, A. Colaco, and V. K. Goyal, "Compressive depth map acquisition using a single photon-counting detector: Parametric signal processing meets sparsity," in *CVPR*, 2012, pp. 96–102.

[49] A. Adam, C. Dann, O. Yair, S. Mazor, and S. Nowozin, "Bayesian time-of-flight for realtime shape, illumination and albedo," *CoRR*, vol. abs/1507.06173, 2015.

**Guy Rosman** Guy Rosman is a post-doctoral fellow at MIT / CSAIL, where he received the Technion-MIT post-doctoral Fellowship and is working with the Distributed Robotics Lab and the Sensing, Learning and Inference group. He obtained in 2004 his BSc Summa Cum Laude, in 2008 MSc Cum Laude, and in 2013 PhD at the Technion (with the Jacobs-Qualcomm fellowship), in the Computer Science Department. He has worked at several companies/labs, including IBM/HRL, RAFAEL, Medicvision, and Invision Biometrics. His research interests include computer vision and 3D sensing, as well as inference and machine learning techniques.

**Daniela Rus** Daniela Rus is the Andrew (1956) and Erna Viterbi Professor of Electrical Engineering and Computer Science and Director of the Computer Science and Artificial Intelligence Laboratory (CSAIL) at MIT. Rus's research interests are in robotics, mobile computing, and data science. Rus is a Class of 2002 MacArthur Fellow, a fellow of ACM, AAAI and IEEE, and a member of the National Academy of Engineering. She earned her PhD in Computer Science from Cornell University. Prior to joining MIT, Rus was a professor in the Computer Science Department at Dartmouth College.

**John W. Fisher III** (M"98) received the Ph.D. degree in Electrical and Computer Engineering from the University of Florida (UF), Gainesville, in 1997. He is currently a Principal Research Scientist in the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology (MIT), Cambridge, and affiliated with the Laboratory for Information and Decision Systems, MIT. Prior to joining MIT, he has been affiliated with UF as both a faculty member and graduate student since 1987, during which time he conducted research in the areas of ultrawideband radar for ground penetration and foliage penetration applications, radar signal processing, and automatic target recognition algorithms. His current area of research focus includes information theoretic approaches to signal processing, multimodal data fusion, machine learning, and computer vision.