

SCIENTIFIC REPORTS



OPEN

Tracking objects outside the line of sight using 2D intensity images

Jonathan Klein^{1,2}, Christoph Peters¹, Jaime Martín¹, Martin Laurenzis² & Matthias B. Hullin¹

Received: 06 July 2016

Accepted: 09 August 2016

Published: 31 August 2016

The observation of objects located in inaccessible regions is a recurring challenge in a wide variety of important applications. Recent work has shown that using rare and expensive optical setups, indirect diffuse light reflections can be used to reconstruct objects and two-dimensional (2D) patterns around a corner. Here we show that occluded objects can be tracked in real time using much simpler means, namely a standard 2D camera and a laser pointer. Our method fundamentally differs from previous solutions by approaching the problem in an analysis-by-synthesis sense. By repeatedly simulating light transport through the scene, we determine the set of object parameters that most closely fits the measured intensity distribution. We experimentally demonstrate that this approach is capable of following the translation of unknown objects, and translation and orientation of a known object, in real time.

The widespread availability of digital image sensors, along with advanced computational methods, has spawned new imaging techniques that enable seemingly impossible tasks. A particularly fascinating result is the use of ultrafast time-of-flight measurements^{1,2} to image objects outside the direct line of sight^{3–6}. Being able to use arbitrary walls as though they were mirrors can provide a critical advantage in many sensing scenarios with limited visibility, like endoscopic imaging, automotive safety, industrial inspection and search-and-rescue operations.

Out of the proposed techniques for imaging occluded objects, some require the object to be directly visible to a structured⁷ or narrow-band^{8–10} light source. Others resort to alternative regions in the electromagnetic spectrum where the occluder is transparent^{11–13}. We adopt the significantly more challenging assumption that the object is in the direct line of sight of neither light source nor camera (Fig. 1), and that it can only be illuminated or observed indirectly via a diffuse wall^{3–6,14}. All the observed light has undergone at least three diffuse reflections (wall, object, wall), and reconstructing the unknown object is an ill-posed inverse problem. Most solution approaches reported so far use a backprojection scheme as in computed tomography¹⁵, where each intensity measurement taken by the imager votes for a manifold of possible scattering locations. This *explicit* reconstruction scheme is computationally efficient, in principle real-time capable⁶, and can be extended with problem-specific filters^{3,16}. However, it assumes the availability of ultrafast time-resolved optical impulse responses, whose capture still constitutes a significant technical challenge. Techniques proposed in literature include direct temporal sampling based on holography^{1,17,18}, streak imagers², gated image intensifiers⁵, serial time-encoded amplified microscopy¹⁹, single-photon avalanche diodes²⁰, and indirect computational approaches using multi-frequency lock-in measurements^{21–23}. In contrast, *implicit* methods state the reconstruction task in terms of a problem-specific cost function that measures the agreement of a scene hypothesis with the observed data and additional model priors. The solution to the problem is defined as the function argument that minimizes the cost. In the only such method reported so far⁴, the authors regularize a least-squares data term with a computationally expensive sparsity prior, which enables the reconstruction of unknown objects around a corner without the need for ultrafast light sources and detectors.

Here we introduce an implicit technique for detecting and tracking objects outside the line of sight in real time. Imaged using routinely available hardware (2D camera, laser pointer), the distribution of indirect light falling back onto the wall serves as our main source of information. This light has undergone multiple reflections; therefore, the observed intensity distribution is low in spatial detail. Our method combines a simulator for three-bounce indirect light transport with a reduced formulation of the reconstruction task^{6,16}. Rather than aiming to reconstruct the geometry of an unknown object, we assume that the target object is rigid, and that its shape and material are either known and/or irrelevant. Translation and rotation, the only remaining degrees of

¹University of Bonn, Institute of Computer Science II, Bonn, Germany. ²French-German Research Institute Saint-Louis, France. Correspondence and requests for materials should be addressed to M.B.H. (email: hullin@cs.uni-bonn.de)

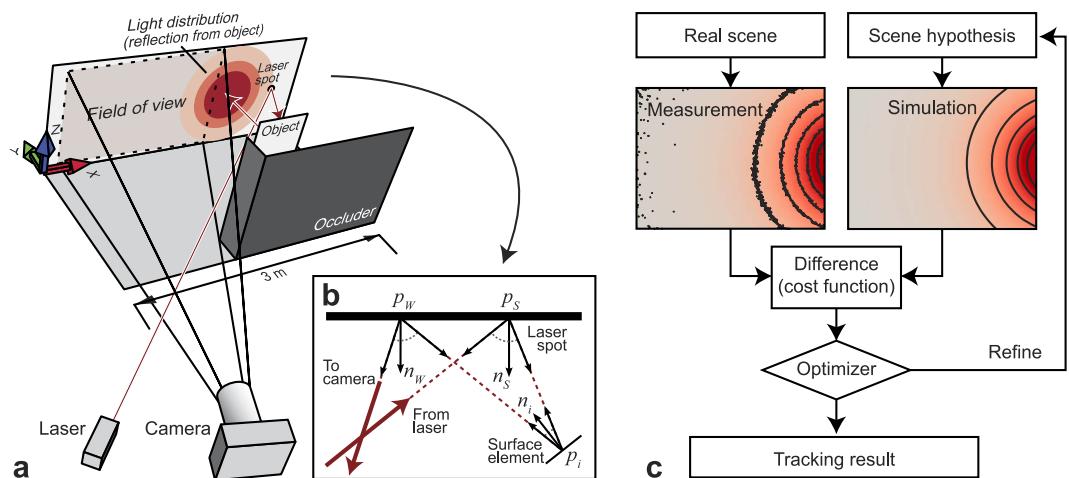


Figure 1. Tracking objects around a corner. (a) Our experimental setup follows the most common arrangement reported in prior work, except that it does not use time-of-flight technology. A camera observes a portion of a white wall. To the right of the camera's field of view, a collimated laser illuminates a spot that reflects light toward the unknown object. The light distribution observed by the camera is the result of three diffuse light bounces (wall–object–wall) plus ambient contributions. (b) Geometry of three-bounce reflection for a single surface element. (c) Flow diagram of our tracking algorithm. Given shape, position and orientation of an object (the “scene hypothesis”), we simulate light transport to predict the distribution that this object would produce on the wall. By comparing this distribution to the one actually observed by the camera, and refining the parameters to minimize the difference, the object's motion is estimated.

freedom, can now be found by minimizing a least-squares energy functional, forcing the scene hypothesis into agreement with the captured intensity image.

Our main contributions are threefold. We propose to use light transport simulation to tackle an indirect vision task in an analysis-by-synthesis sense. Using synthetic measurements, we quantify the effect of object movement on the observed intensity distribution, and predict under which conditions the effect is significant enough to be detected. Finally, we demonstrate and evaluate a hardware implementation of a tracking system. Our insights are not limited to intensity-only imaging, and we believe that they will bring non-line-of-sight sensing closer to practical applications.

Results

Light transport simulation (synthesis). At the center of this work is an efficient renderer for three-bounce light transport. Being able to simulate indirect illumination at an extremely fast rate is crucial to the overall system performance, since each object tracking step requires multiple simulation runs. Like all prior work, we assume that the wall is planar and known, and so is the position of the laser spot. The object is represented as a collection of Lambertian surface elements (*surfels*), each characterized by its position, normal direction and area. As the object is moved or rotated, all its surfels undergo the same rigid transformation. We represent this transformation by the *scene parameter* p , which is a three-dimensional vector for pure translation, or a six-dimensional vector for translation and rotation. The irradiance received by a given camera pixel is computed by summing the light that reflects off the surfels. The individual contributions, in turn, are obtained independently of each other as detailed in the Methods section, by calculating the radiative transfer from the laser spot via a surfel to the location on the wall observed by a pixel. Note that by following this procedure, like all prior work, we neglect self-occlusion, occlusion of ambient light, and interreflections. To efficiently obtain a full-frame image, represented by the vector of pixel values $S(p)$, we parallelized the simulation to compute each pixel in a separate thread on the graphics card. The rendering time is approximately linear in the number of pixels and the number of surfels. On an NVIDIA GeForce GTX 780 graphics card, the response from a moderately complex object (500 surfels) at a resolution of 160×128 pixels is rendered in 3.57 milliseconds.

To estimate the magnitude of changes in the intensity distribution that are caused by motion or a change in shape, we performed a numerical experiment using this simulation. In this experiment, we used a fronto-parallel view on a $2\text{ m} \times 2\text{ m}$ wall, with a small planar object (a $10\text{ cm} \times 10\text{ cm}$ white square) located at 50 cm from the wall. Object and laser spot were centered on the wall, but not rendered into the image. Figure 2 shows the simulated response thus obtained. By varying position and location of the object, we obtained difference images that can be interpreted as partial derivatives with respect to the components of the scene parameter p . Since the overall light throughput drops with the fourth power of the object-wall distance, translation in Y direction caused the strongest change. Translation in all directions and rotation about the X and Z axes affected the signal more strongly than the other variations. With differences amounting to several percent of the overall intensity, these changes were significant enough to be detected using a standard digital camera with 8- to 12-bit A/D converter.

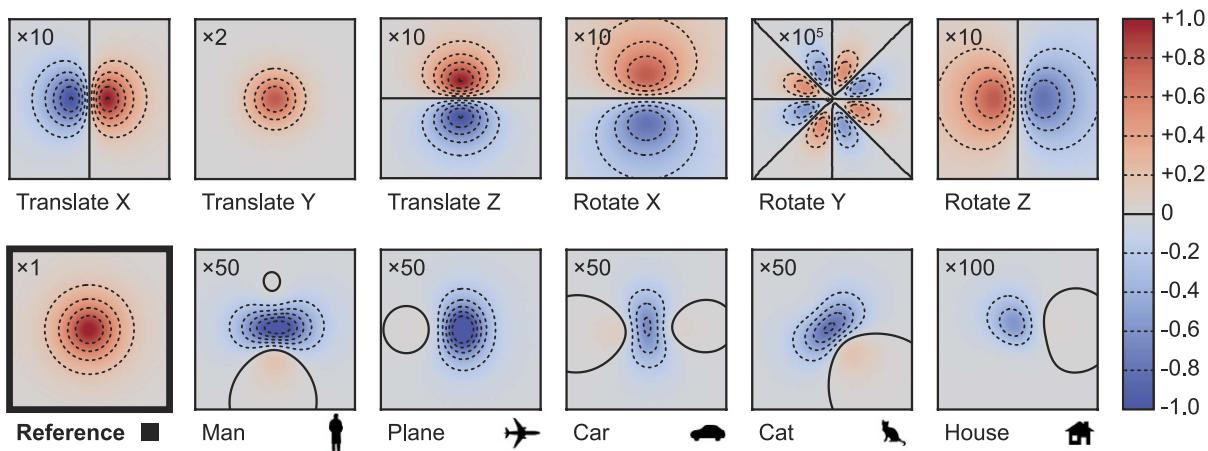


Figure 2. Intensity difference images. To investigate the effect of changes in object position and orientation on the intensity distribution observed on the wall, we performed a simplified synthetic experiment with an orthographic view of a $2\text{ m} \times 2\text{ m}$ wall, and laser spot and object centered with respect to the wall. The reference distribution (bottom left) was produced by a $10\text{ cm} \times 10\text{ cm}$ square-shaped object, located at 50 cm from the wall. Six difference images (top row), obtained by translating ($\pm 2.5\text{ cm}$) and rotating ($\pm 7.5^\circ$) the object about the X, Y and Z axes, illustrate the distribution and magnitude of the respective change in the signal. The images shown in the bottom row visualize the difference caused by a change in shape. For display, each difference image has been amplified by the indicated factor (2 to $100,000$) that also reflects the relative significance of the effect: Translations and rotations (except around the Y axis) caused the signal to change by roughly 1% per centimeter or per angular degree. A change in the object shape led to a peak difference around 1–2%, and rotation around the Y axis had a much smaller effect.

Experimental setup. Our experiment draws inspiration from prior work^{3,4,6,14,16}; the setup is sketched in Fig. 1(a). Here, due to practical constraints, some of the idealizing assumptions made during the synthetic experiment had to be relaxed. In particular, only an off-peak portion of the intensity pattern could be observed. To shield the camera from the laser spot and to avoid saturation and lens flare, we had to position the laser spot outside the field of view. The actual reflectance distribution of the wall and object surfaces was not perfectly Lambertian, and additional light emitters and reflectors, not accounted for by the simulation, were present in the scene. To obtain a measured image \mathbf{M} containing only light from the laser, we took the difference of images captured with and without laser illumination. Additionally, we subtracted a calibration measurement $\hat{\mathbf{B}}$ containing light reflected by the background. A specification of the devices used, and a more detailed introduction of the data pre-processing steps, can be found in the Methods section.

Tracking algorithm (analysis). With the light transport simulation at hand, and given a measurement of light scattered from the object to the wall, we formulate the tracking task as a non-linear minimization problem. Suppose \mathbf{M} and $\mathbf{S}(p)$ are vectors encoding the pixel values of the *measured* object term and the one predicted by the *simulation* under the transformation parameter or scene hypothesis p , respectively. We search for the parameter p that brings \mathbf{M} and $\mathbf{S}(p)$ into the best possible agreement by minimizing the cost function

$$f(p) = \|\mathbf{M} - \gamma(\mathbf{M}, \mathbf{S}(p)) \cdot \mathbf{S}(p)\|_2^2, \quad \text{where} \quad \gamma(a, b) = \frac{a^T \cdot b}{\|b\|_2^2}. \quad (1)$$

The factor $\gamma(a, b)$ projects b to a , minimizing the distance $\|a - \gamma(a, b) \cdot b\|_2^2$. By including this factor into our objective, we decouple the recovery of the scene parameter p from any unknown global scaling between measurement and simulation, caused by parameters such as surface albedos, camera sensitivity and laser power. To solve this non-linear, non-convex, heavily over-determined problem, we use the Levenberg-Marquardt algorithm²⁴ as implemented in the Ceres library²⁵. Derivatives are computed by numerical differentiation. When tracking six degrees of freedom (translation and rotation), evaluating the value and gradient of f requires a total of seven simulation runs, or on the order of 25 milliseconds of compute time on our system.

Tracking result. To evaluate the method, we performed a series of experiments. The physical object used in all experiments was a car silhouette cut from plywood and coated with white wall paint, shown in Fig. 3(a). While our setup is able to handle arbitrary three-dimensional objects (as long as the convexity assumption is reasonable), this shape was two-dimensional for manufacturing and handling reasons.

For a given input image \mathbf{M} and object shape, the cost function $f(p)$ in Eq. (1) depends on three to six degrees of freedom that are being tracked. Figure 3(b) shows a slice of the function for translation in the XY-plane, with all other parameters fixed. Although the global minimum is located in an elongated, curved trough, only four to five iterations of the Levenberg-Marquardt algorithm are required for convergence from a random location in the tracking volume. In real-time applications, since position and rotation can be expected to change slowly over

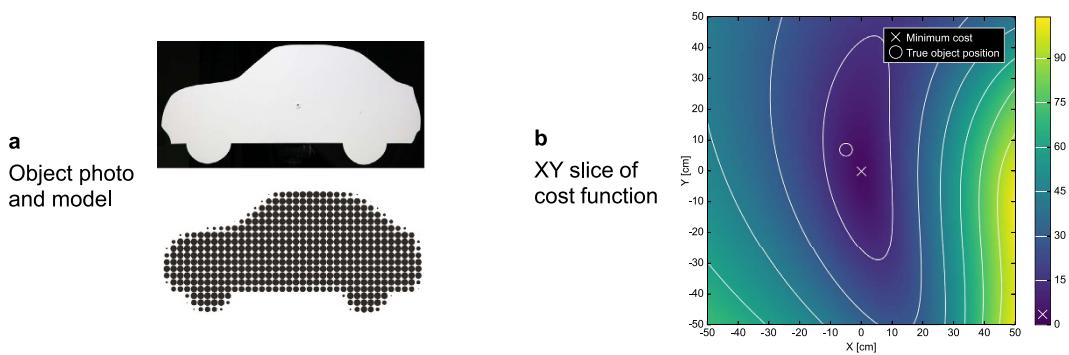


Figure 3. Object model and cost function used for tracking. (a) photo of an object cut from white plywood, and its representation as surface elements (surfels). Note that although we use a flat object for demonstration, our method is also capable of handling three-dimensional objects. (b) XY slice of the cost function for positional tracking, centered around the global minimum. With a perfect image formation model and in the absence of noise, the minimum (marked by cross) and the measured position of the object (marked by circle) should coincide at a function value of exactly $f(p) = 0$. Under real conditions, the reconstructed position deviated from the true one by a few centimeters, and the minimum was a small positive value.

time, the optimization effort can be reduced to two to three iterations per frame by using the latest tracking result to initialize the solution for the next frame.

In **Experiment 1**, we kept the object's orientation constant. We manually placed the object at various known locations in an $60\text{ cm} \times 50\text{ cm} \times 60\text{ cm}$ working volume, and recorded 100 camera frames at each location. These frames differ in the amount of ambient light (mains flicker) and in the photon noise. For each frame, we initialized the estimated position to a random starting point in a cube of dimensions $(30\text{ cm})^3$ centered in the tracking volume, and refined the position estimate by minimizing the cost function, Eq. (1). The results are shown in Fig. 4(a). From this experiment, we found positional tracking to be repeatable and robust to noise, with a sub-cm standard deviation for each position estimate. The root-mean-square distance to ground truth was measured at 4.8 cm, 2.9 cm and 2.4 cm for movement along the X, Y and Z axis, respectively. This small systematic bias was likely caused by a known shortcoming of the image formation model, which does not account for occlusion of ambient light by the object.

In **Experiment 2**, we kept the object at a (roughly) fixed location and rotated it by a range of $\pm 30^\circ$ around the three coordinate axes using a pan-tilt-roll tripod with goniometers on all joints. Again, per setting we recorded 100 frames that mainly differ in the noise pattern. We followed the same procedure as in the first experiment, except that this time we jointly optimized for all six degrees of freedom (position and orientation). The results are shown in Fig. 4(b). As expected, the rotation angles were tracked with higher uncertainty than the translational parameters, although the average reconstructions for each angle remain stable. We identify two main sources for the added uncertainty: the increased number of degrees of freedom and the pairwise ambiguity between X translation and Z rotation, and between Z translation and X rotation (Fig. 2). We recall that in the synthetic experiment, the effect of Y rotation was vanishingly small; here, the system tracked rotation around the Y axis about as robustly as the other axes. This unexpectedly positive result was probably owed to the strongly asymmetric shape of the car object.

So far, we assumed that the object's shape was known. Since this requirement cannot always be met, we dropped it in **Experiment 3**. Using the data already captured using the car object for the first experiment, we performed the light transport simulation using a single oriented surface element instead of the detailed object model. Except for this simplification, we followed the exact same procedure as in Experiment 1 to track the now unknown object's position. The results are shown in Fig. 5(a). Despite a systematic shift introduced by the use of the simplified object model, the position recovery remained robust to noise and relative movement was still detected reliably.

The need for a measured background term $\hat{\mathbf{B}}$ can hinder the practical applicability of our approach as pursued so far. In **Experiment 4**, we lifted this requirement. When omitting the term without any compensation, the tracking performance degraded significantly (Fig. 5(b)). However, we observed that the background image, caused by distant scattering, was typically smooth and well approximated by a linear function $g(u, v) = au + bv + c$ in the image coordinates u and v (Fig. 6). We extended the tracking algorithm to fit such linear models to both input images \mathbf{M} and $\mathbf{S}(p)$, and subtract the linear portions prior to evaluating the cost function (Eq. 2). This simple pre-processing step greatly reduced the bias in the tracking outcome and enabled robust tracking of object motion (Fig. 5(c)) even in unknown rooms.

The supplementary video to this paper shows two real-time tracking sessions (Session 1: translation only; Session 2: translation and rotation) using the described setup. A live view of the hidden scene is shown next to the output from the tracking software. The average reconstruction rate during these tracking sessions was 10.2 frames per second (limited by the maximum capture rate of our camera-laser setup) for Session 1, and 3.7 frames per second (limited by computation) for Session 2. The 2-dimensional car model was represented by 502 surfels; the total compute time required for a single tracking step was 72.9 ms for translation only, and 226.1 ms for translation and rotation.

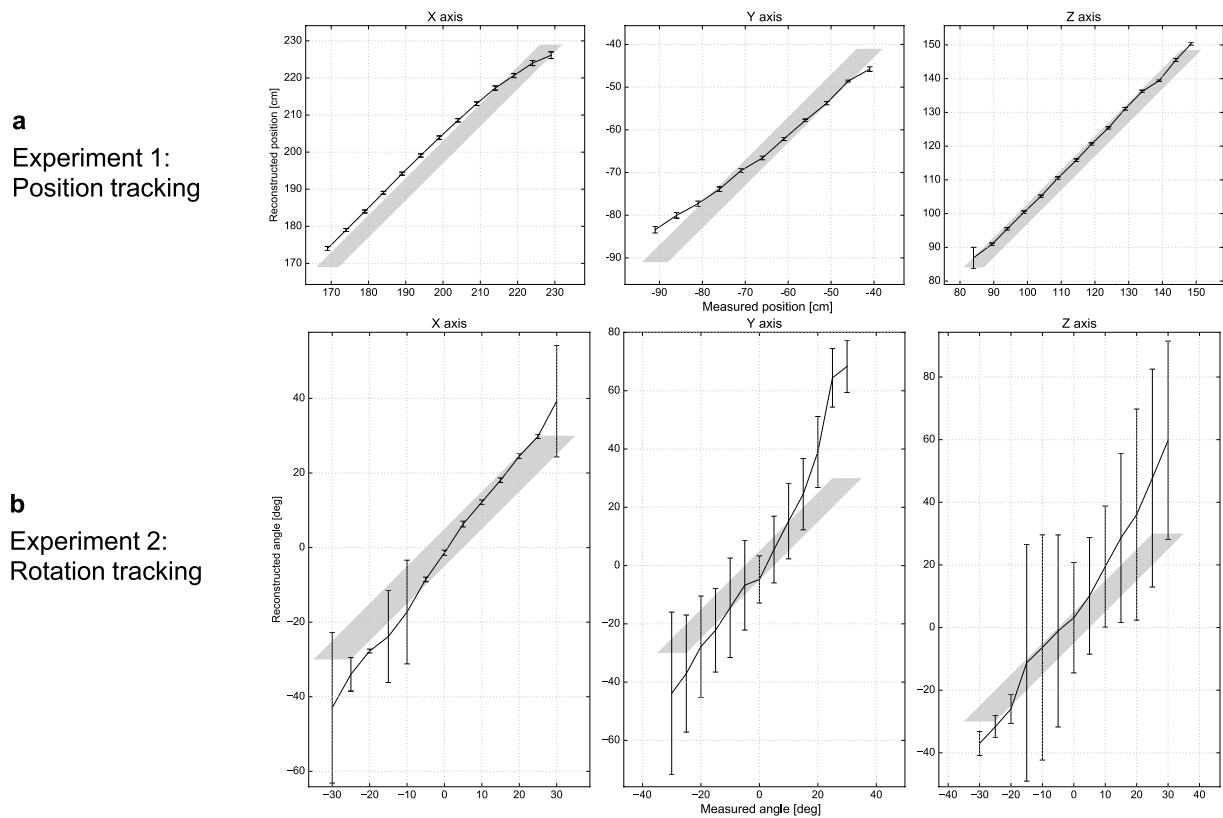


Figure 4. Tracking a known object. (a) Result of three tracking sessions where the object was translated along the X, Y and Z axes (Experiment 1). We recorded 100 input images at each position and reconstructed the object position for each input image independently. Plots and error bars visualize the mean and standard deviation of the recovered positions. The area shaded in gray is the confidence range for the true position which was determined using a tape measure. (b) Result of three tracking sessions where the object was rotated around the X, Y and Z axes (Experiment 2). From 100 input images, we jointly reconstructed translation and rotation. Shown are mean and standard deviation of the recovered rotation angle. The higher uncertainty reflects the fact that rotation in general has a smaller effect on the signal, and the ambiguity between translational and rotational motion (also see Fig. 2).

Discussion

The central finding of this work is that the popular challenge of tracking an object around a corner can be tackled without the use of time-of-flight technology. By formulating an optimization problem based on a simplistic image formation model, we demonstrated parametric object tracking only using 2D images with a laser pointer as the light source. In a room-sized scene, our technique achieves sub-cm repeatability, which puts it on par with the latest time-of-flight-based techniques^{6,16}. However, as our technique does not rely on temporally resolved measurements of any kind, it has the unique property of being scalable to very small scenes (down to the diffraction limit) as well as large scenes (sufficient laser power provided). We note that the analysis-by-synthesis approach *per se* is not limited to pure intensity imaging, but may form a valuable complement to other sensing modalities as well. For instance, a simple extension to the light transport model would enable it to accommodate time-of-flight or phase imaging.

A key feature of the analysis-by-synthesis paradigm is its transparency. Putting a virtual experiment (simulation) alongside the real experiment enables a rigorous quantitative analysis of the sensing problem. Using difference images, for instance, we investigated the influence that parameter changes have on the signal, and predicted the detectability of centimetre-scale motion. The same mechanism could also be used to obtain robustness estimates regarding additional unknowns in the scene model, such as non-diffuse object reflectance or the presence of additional objects. With these options, our approach offers a significant advantage over existing non-line-of-sight sensing techniques.

The real-time performance of our technique is determined by four main factors: the capture rate of the camera (constrained by exposure time and read-out bandwidth), performance of the compute system, the discretization of the model into surfels and the number of translational and rotational degrees of freedom afforded to the model. Other factors, in particular the question whether object and room are known, are irrelevant with this regard.

We identify four main limiting factors to the resolution and repeatability of our technique. Firstly, shortcomings in the models for scene and light transport can introduce a systematic bias. We exemplarily demonstrated how additional heuristic pre-processing steps can mitigate this bias. In usage scenarios where systematic errors preclude quantitative tracking, simpler sensing tasks, like the detection of object motion, will still remain possible.

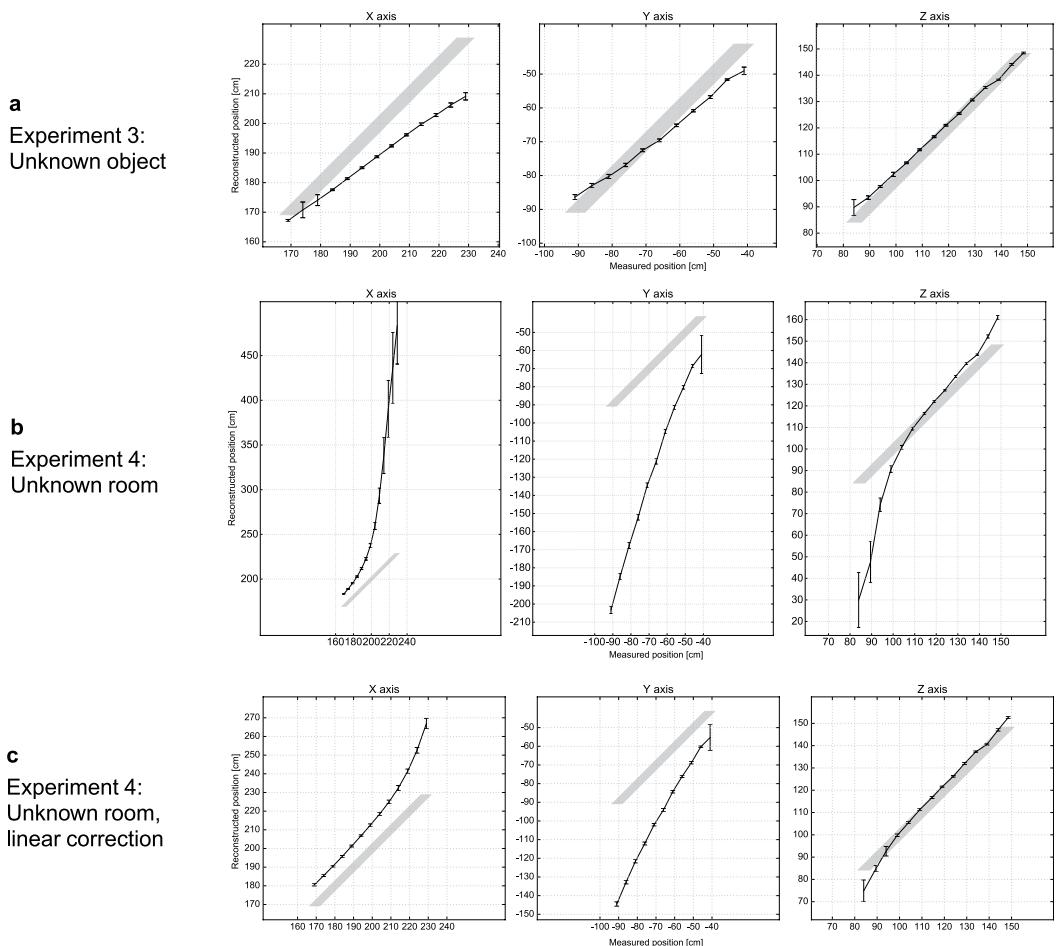


Figure 5. Tracking of an unknown object, or in an unknown room. (a) Result of Experiment 3: Positional tracking as in Experiment 1, but with no knowledge about the object shape. We used a single oriented surface element for the light transport simulation. (b) Result of Experiment 4: Positional tracking as in Experiment 1, but without subtracting the pre-calibrated room response. The estimated absolute position greatly deviated from the ground-truth position (shaded areas). (c) Subtraction of a linear fit significantly reduced the tracking error and made the tracking task feasible even in the absence of a background measurement. In all cases, the standard deviation (error bars) remained small, indicating that changes in position could still be robustly detected.

Secondly, the tracking of additional parameters like rotation, non-rigid objects or multiple object positions, is sensitive to image noise. The adoption of advanced filtering techniques or multi-frame averaging will further improve the tracking quality. Furthermore, certain applications will require a careful selection of the degrees of freedom afforded to the model. Thirdly, like in all prior work, we assumed knowledge about the geometry and angular reflectance distribution of a wall that receives light scattered by the unknown object. Thanks to recent progress in mobile mapping²⁶, highly detailed geometry and albedo texture data is already widely available for many application scenarios; if not, it can be recovered using existing line-of-sight sensing methods. Lastly, our tracking result is the outcome of a local parameter search (Levenberg–Marquardt) and hence not guaranteed to be the global optimum of the cost function, Eq. (1). Although we never experienced convergence problems in practice, some situations may necessitate a combination of global and local optimization strategies.

The prospective of being able to sense beyond the direct line of sight can benefit many application fields. So far, the deployment of existing approaches has been hindered by practical limitations such as long capture times and device costs. As we were able to show here, these limitations can in principle be overcome if the problem can be reduced to a small number of degrees of freedom. One of the first applications of such reduced models could be in urban traffic safety, where the motion of vehicles and pedestrians is constrained to the ground plane. Extrapolating from our results, we believe that more detailed forward models and efficient simulation techniques can become a source of profound insight about non-line-of-sight sensing problems—and, eventually, enable the first truly practical solutions for looking around corners.

Methods

Light transport simulation. Accurate simulation of indirect illumination is computationally expensive and can take hours to complete. By assuming that all light has undergone exactly three reflections, we achieved a

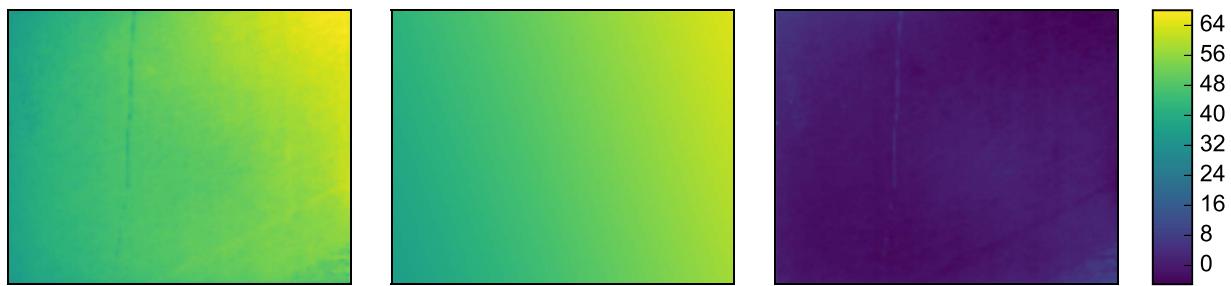


Figure 6. Approximating the background term by a linear model. From left to right, in arbitrary units: background term $\hat{\mathbf{B}}$ obtained through calibration, linear approximation of $\hat{\mathbf{B}}$, residual background term after subtraction of linear component.

reduced overall computational complexity that is linear in the number of pixels and the number of surfels n . The geometry of this simulation is provided in Fig. 1b. Each camera pixel observes a radiance value, L , leaving from a point on the wall, p_W , that, in turn, receives light reflected by the object's surfels. The portion contributed by the surfel of index $i \in \{1 \dots n\}$ is the product of three reflectance terms, one per reflection event; and the geometric view factors known from radiative transfer^{27,28}:

$$\begin{aligned}
 L_i: &= \rho_0 \cdot f_s(p_L - p_S, p_i - p_S) && \text{(laser spot)} \\
 &\cdot \frac{(n_S \circ (p_i - p_S)) \cdot (n_i \circ (p_S - p_i))}{\|p_S - p_i\|_2^2} \cdot f_i(p_S - p_i, p_W - p_i) \cdot A_i && \text{(\textit{i}th surfel)} \\
 &\cdot \frac{(n_i \circ (p_W - p_i)) \cdot (n_W \circ (p_i - p_W))}{\|p_i - p_W\|_2^2} \cdot f_W(p_i - p_W, p_C - p_W) && \text{(wall),}
 \end{aligned} \tag{2}$$

where the operator

$$v \circ w := \begin{cases} \frac{v^T \cdot w}{\|v\|_2 \cdot \|w\|_2} & \text{if } v^T \cdot w > 0 \\ 0 & \text{otherwise} \end{cases}$$

denotes a normalized and clamped dot product as used in Lambert's cosine law. Each line in Eq. (2) models one of the three surface interactions. n_S , n_i and N_W are the normal vectors of laser spot, surfel and observed point on the wall, and $f_{\{S,i,W\}}(\omega_{\text{in}}, \omega_{\text{out}})$ are the values of the corresponding bidirectional reflectance distribution functions (BRDF). The incident and outgoing direction vectors ω_{in} and ω_{out} that form the arguments to the BRDF are given by the scene geometry. In particular, the vectors p_L, p_S, p_i, p_W and p_C represent the positions of, in this order: the laser source, the laser spot on the wall, the i^{th} surfel, the observed point on the wall, and the camera (center of projection). A_i is the area of the i^{th} surfel, and ρ_0 a constant factor that subsumes laser power and the light efficiencies of lens and sensor. This factor is cancelled out by the projection performed in the cost function Eq. (1), so we set it to $\rho_0 = 1$ in simulation. The total pixel value is simply computed by summing Eq. (2) over all surfels:

$$L_{\text{total}} := \sum_{i=1}^n L_i \tag{3}$$

This summation neglects mutual shadowing or inter-reflection between surfels, an approximation that is justifiable for flat or mostly convex objects. For lack of measured material BRDFs, we further assume all surfaces to be of diffuse (Lambertian) reflectance such that $f_{\{S,i,W\}} := \text{const} = 1$, again making use of the fact that the cost function Eq. (1) is invariant under such global scaling factors. If available, more accurate BRDF models as well as object and wall textures can be included at a negligible computational cost.

Capture devices. Our image source was a Xenics Xeva-1.7-320 camera, sensitive in the near-infrared range (900 nm–1,700 nm), with a resolution of 320×256 pixels at 14 bits per pixel. We used an exposure time of 20 ms. The laser source (1 W at 1.550 nm) was a fiber-coupled laser diode of type SemiNex 4PN-108 driven by an Analog Technologies ATLS4A201D laser diode driver and equipped with a USB interface trigger input. On the output side of the fiber, we fed the collimated beam through a narrow tube with absorbing walls to reduce stray light.

A desktop PC with an NVIDIA GeForce GTX 780 GPU, 32 GB of RAM and an Intel Core i7-4930 K CPU controlled the devices and performed the reconstruction.

Measurement routine and image pre-processing. After calibrating the camera's gain factors and fixed pattern noise using vendor tools, we assumed that all pixels had the same linear response. All images were down-sampled to half the resolution (160×128 pixels) prior to further processing. Due to the diffuse reflections, apart from noise, the measurements do not contain any information of high spatial frequency. Thus, moderate down-sampling is a safe way to improve the performance of the later reconstruction.

The images measured by the camera are composed of several contributions, each represented by a vector of pixel-wise contributions: *ambient* light not originating from the laser, **A**; laser light scattered by static *background* objects present in the scene, **B**; and laser light scattered by the dynamic *object*, **O**. All measured images are further affected by noise, the main sources being photon counting noise and signal-independent read noise. We assume the scene to remain stationary at least during short time intervals between successive captures. Further assuming the spatial extent of the object to be small, shadowing of **A** and **B** by the object, as well as ambient light reflected by the object, can be neglected. By turning the laser on and off, and inserting and removing the object, the described kind of setup can capture the following combinations of these light contributions:

$$\begin{aligned} \text{Laser off (0), object absent (0): } \mathbf{I}_{00} &= \mathbf{A} + \text{noise} \\ \text{Laser on (1), object absent (0): } \mathbf{I}_{10} &= \mathbf{A} + \mathbf{B} + \text{noise} \\ \text{Laser off (0), object present (1): } \mathbf{I}_{01} &= \mathbf{A} + \text{noise} \\ \text{Laser on (1), object present (1): } \mathbf{I}_{11} &= \mathbf{A} + \mathbf{B} + \mathbf{O} + \text{noise} \end{aligned}$$

The input image to the reconstruction algorithm, **M**, was obtained as the difference of images captured in quick succession with and without laser illumination. Additionally, we subtracted a calibration measurement containing light reflected by the background:

$$\mathbf{M} := \mathbf{I}_{11} - \mathbf{I}_{01} - \hat{\mathbf{B}} \approx \mathbf{O} + \text{noise}, \quad (4)$$

The addition or subtraction of two input images increases the noise magnitude by a factor of about $\sqrt{2}$. The background estimate $\hat{\mathbf{B}}$ was captured with the object removed by recording difference images with and without laser illumination. We averaged $n = 300$ such difference images to reduce noise in the background estimate:

$$\hat{\mathbf{B}} := \frac{1}{n} \sum_{i=0}^n (\mathbf{I}_{10}^{(i)} - \mathbf{I}_{00}^{(i)}) \stackrel{n \gg 1}{\approx} \mathbf{B} \quad (5)$$

References

1. Abramson, N. Light-in-flight recording by holography. *Optics Letters* **3**, 121–123 (1978).
2. Velten, A., Raskar, R. & Bawendi, M. Picosecond camera for time-of-flight imaging. In *Imaging and Applied Optics IMB4* (Optical Society of America, 2011).
3. Velten, A. *et al.* Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications* **3**, 745 (2012).
4. Heide, F., Xiao, L., Heidrich, W. & Hullin, M. B. Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (2014).
5. Laurenzis, M. & Velten, A. Nonline-of-sight laser gated viewing of scattered photons. *Optical Engineering* **53**, 023102–023102 (2014).
6. Gariepy, G., Tonolini, F., Henderson, R., Leach, J. & Faccio, D. Detection and tracking of moving objects hidden from view. *Nature Photonics* **10** (2016).
7. Sen, P. *et al.* Dual Photography. *ACM Transactions on Graphics (Proc. SIGGRAPH)* **24**, 745–755 (2005).
8. Steinvall, O., Elmquist, M. & Larsson, H. See around the corner using active imaging. *Proc. SPIE* **8186**, 818605–818605–17 (2011).
9. Katz, O., Small, E. & Silberberg, Y. Looking around corners and through thin turbid layers in real time with scattered incoherent light. *Nature Photonics* **6**, 549–553 (2012).
10. Katz, O., Heidmann, P., Fink, M. & Gigan, S. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nature Photonics* **8**, 784–790 (2014).
11. Sume, A. *et al.* Radar detection of moving targets behind corners. *Geoscience and Remote Sensing, IEEE Transactions on* **49**, 2259–2267 (2011).
12. Adib, F. & Katabi, D. See through walls with wifi! *SIGCOMM Comput. Commun. Rev.* **43**, 75–86 (2013).
13. Adib, F., Hsu, C.-Y., Mao, H., Katabi, D. & Durand, F. Capturing the human figure through a wall. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* **34**, 219 (2015).
14. Buttafava, M., Zeman, J., Tosi, A., Eliceiri, K. & Velten, A. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics Express* **23**, 20997–21011 (2015).
15. Pan, X., Sidky, E. Y. & Vannier, M. Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction? *Inverse Problems* **25**, 123009 (2009).
16. Kadambi, A., Zhao, H., Shi, B. & Raskar, R. Occluded imaging with time-of-flight sensors. *ACM Transactions on Graphics* **35**, 15 (2016).
17. Abramson, N. Light-in-flight recording: high-speed holographic motion pictures of ultrafast phenomena. *Applied optics* **22**, 215–232 (1983).
18. Quercioli, F. & Molesini, G. White light-in-flight holography. *Applied Optics* **24**, 3406–3415 (1985).
19. Goda, K., Tsia, K. K. & Jalali, B. Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena. *Nature* **458**, 1145–1149 (2009).
20. Gariepy, G. *et al.* Single-photon sensitive light-in-flight imaging. *Nature Communications* **6** (2015).
21. Heide, F., Hullin, M. B., Gregson, J. & Heidrich, W. Low-budget transient imaging using photonic mixer devices. *ACM Transactions on Graphics (Proc. SIGGRAPH)* **32**, 45 (2013).
22. Kadambi, A. *et al.* Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* **32**, 167 (2013).
23. Peters, C., Klein, J., Hullin, M. B. & Klein, R. Solving trigonometric moment problems for fast transient imaging. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* **34**, 220 (2015).
24. Marquardt, D. W. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* **11**, 431–441 (1963).
25. Agarwal, S., Mierle, K. & Others [sic]. Ceres solver. <http://ceres-solver.org> (2015).
26. Puente, I., González-Jorge, H., Martínez-Sánchez, J. & Árias, P. Review of mobile mapping and surveying technologies. *Measurement* **46**, 2127–2145 (2013).
27. Çengel, Y. & Ghajar, A. *Heat and Mass Transfer: Fundamentals and Applications* (McGraw-Hill Education, 2014).
28. Goral, C. M., Torrance, K. E., Greenberg, D. P. & Battaile, B. Modeling the interaction of light between diffuse surfaces. *SIGGRAPH Computer Graphics* **18**, 213–222 (1984).

Acknowledgements

This work was supported by the X-Rite Chair for Digital Material Appearance, and the German Research Foundation (HU 2273/2-1). We thank Frank Christnacher and Emmanuel Bacher for their help with the setup and discussion.

Author Contributions

The method was conceived by M.B.H. and refined by M.B.H., J.K., C.P., J.K. and M.L. implemented and optimized the system and performed the experiments based on prototypes by M.B.H. and J.M., M.B.H. and J.K. wrote the paper.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Klein, J. *et al.* Tracking objects outside the line of sight using 2D intensity images. *Sci. Rep.* **6**, 32491; doi: 10.1038/srep32491 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016