

# Pose estimation using time-resolved inversion of diffuse light

Dan Raviv,\* Christopher Barsi, Nikhil Naik, Micha Feigin,  
and Ramesh Raskar

Media Lab, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA  
02139, USA

[\\*darav@mit.edu](mailto:darav@mit.edu)

**Abstract:** We present a novel approach for evaluation of position and orientation of geometric shapes from scattered time-resolved data. Traditionally, imaging systems treat scattering as unwanted and are designed to mitigate the effects. Instead, we show here that scattering can be exploited by implementing a system based on a femtosecond laser and a streak camera. The result is accurate estimation of object pose, which is a fundamental tool in analysis of complex scenarios and plays an important role in our understanding of physical phenomena. Here, we experimentally show that for a given geometry, a single incident illumination point yields enough information for pose estimation and tracking after multiple scattering events. Our technique can be used for single-shot imaging behind walls or through turbid media.

© 2014 Optical Society of America

**OCIS codes:** (110.1758) Computational imaging; (320.7100) Ultrafast measurements.

---

## References and links

1. M. C. Roggeman, B. M. Welsh, and B. R. Hunt, *Imaging Through Turbulence* (CRC Press, 1996).
2. S. R. Arridge, P. van der Zee, M. Cope, and D. T. Delpy, "Reconstruction methods for infra-red absorption imaging," *Proc. SPIE* **1431**, 204–215 (1991).
3. E. Abraham, A. Younus, J. C. Delagnes, and P. Mounaix, "Non-invasive investigation of art paintings by terahertz imaging," *Appl. Phys. A* **100**, 585–590 (2010).
4. J. Dik, K. Janssens, G. Van der Snickt, L. Van der Loeff, K. Rickers, and M. Cotte, "Visualization of a lost painting by Vincent van Gogh using synchrotron radiation based x-ray fluorescence elemental mapping," *Anal. Chem.* **80**, 6436–6442 (2008).
5. P. Heckman Jr. and R. T. Hodgson, "Underwater optical range gating," *IEEE J. Quantum Electron.* **3**, 445–448 (1967).
6. I. Freund, "Looking through walls and around corners," *Physica A* **168**, 49–65 (1990).
7. I. M. Vellekoop and A. P. Mosk, "Focusing coherent light through opaque strongly scattering media," *Opt. Lett.* **32**, 2309–2311 (2007).
8. S. Popoff, G. Lerosey, M. Fink, A. C. Boccara, and S. Gigan, "Image transmission through an opaque material," *Nat. Commun.* **1**, 81 (2010).
9. H. W. Babcock, "The possibility of compensating astronomical seeing," *Publ. Astron. Soc. Pac.* **65**, 229–236 (1953).
10. A. Buffington, F. S. Crawford, R. A. Muller, A. J. Schwemin, and R. G. Smits, "Correction of atmospheric distortion with an image-sharpening telescope," *J. Opt. Soc. Am.* **67**, 298–303 (1977).
11. D. M. Mittleman, M. Gupta, R. Neelamani, R. G. Baraniuk, J. V. Rudd, and M. Koch, "Recent advances in terahertz imaging," *Appl. Phys. B* **68**, 1085–1094 (1999).
12. D. Brown, V. B. Fleurov, P. Carroll, and C. M. Lawson, "Coherence-based imaging through turbid media by use of degenerate four-wave mixing in thin liquid-crystal films and photorefractives," *Appl. Opt.* **37**, 5306–5312 (1998).
13. H. Chen, Y. Chen, D. Dilworth, E. Leith, J. Lopez, and J. Valdmann, "Two-dimensional imaging through diffusing media using 150-fs gated electronic holography techniques," *Opt. Lett.* **16**, 487–489 (1991).

14. B. Danielson and C. Boisrobert, "Absolute optical ranging using low coherence interferometry," *Appl. Opt.* **30**, 2975–2979 (1991).
15. J. C. Hebden and R. A. Kruger, "Transillumination imaging performance: a time-of-flight imaging system," *Med. Phys.* **17**, 351–356 (1990).
16. M. R. Hee, J. A. Izatt, J. M. Jacobson, J. G. Fujimoto, and E. A. Swanson, "Femtosecond transillumination optical coherence tomography," *Opt. Lett.* **18**, 950–952 (1993).
17. Y. Pan, R. Birngruber, and R. Engelhardt, "Contrast limits of coherence-gated imaging in scattering media," *Appl. Opt.* **36**, 2979–2983 (1997).
18. A. Liutkus, D. Martina, S. Popoff, G. Chardon, O. Katz, G. Lerosey, S. Gigan, L. Daudet, and I. Carron, "Imaging with nature: A universal analog compressive imager using a multiply scattering medium," arXiv:1309.0425 (2013).
19. A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, "Recovering three dimensional shape around a corner using ultra-fast time-of-flight imaging," *Nat. Commun.* **3**, 745 (2012).
20. O. Gupta, T. Willwacher, A. Velten, A. Veeraraghavan, and R. Raskar, "Reconstruction of hidden 3D shapes using diffuse reflections," *Opt. Express* **20**, 19096–19108 (2012).
21. N. Naik, A. Velten, C. Barsi, and R. Raskar, "Estimating wide-angle, spatially varying reflectance using time-resolved inversion of backscattered light," *J. Opt. Soc. Am. A* **31**, 957–963 (2014).
22. Q. Wang, L. Wang, and J. F. Sun, "Rotation-invariant target recognition in lidar range imagery using model matching approach," *Opt. Express* **18**, 15349–15360 (2010).
23. R. D. Nieves and W. D. Reynolds, "Three-dimensional transformation for automatic target recognition using lidar data," *Proc. SPIE* **7684**, 76840Y (2010).
24. D. Lv, J.-F. Sun, Q. Li, and Q. Wang, "3D pose estimation of ground rigid target based on lidar range image," *Appl. Opt.* **52**, 8073–8081 (2013).
25. S.-H. Hong and B. Javidi, "Distortion-tolerant 3D recognition of occluded objects using computational integral imaging," *Opt. Express* **14**, 12085–12095 (2006).
26. C. S. L. Chun and F. A. Sadjadi, "Target recognition study using polarimetric laser radar," *Proc. SPIE* **5426**, 274–284 (2004).
27. A. T. N. Kumar, J. Skoch, B. J. Bacskaï, D. A. Boas, and A. K. Dunn, "Fluorescence lifetime based tomography for turbid media," *Opt. Lett.* **30**, 3347–3349 (2005).
28. J. C. Hebden, F. E. W. Schmidt, M. E. Fry, M. Schweiger, E. M. C. Hillman, and D. T. Delpy, "Simultaneous reconstruction of absorption and scattering images by multichannel measurement of purely temporal data," *Opt. Lett.* **24**, 534–536 (1999).
29. P. J. Besl and N. D. MacKay, "A method for registration of 3-D shapes," *Trans. Pattern Anal. Mach. Intell.* **14**, 239–256 (1992).
30. Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," in *International Conference on Robotics and Automation* (1991), Vol. 3, pp. 2724–2729.
31. L. Bottou and Y. Le Cun, "Large scale online learning," in *Advances in Neural Information Processing Systems* (2004), Vol. 16.
32. S. S. Gorthi, D. Schaak, and E. Schonbrun, "Fluorescence imaging of flowing cells using a temporal coded excitation," *Opt. Express* **21**, 5164–5170 (2013).
33. S. Herbert, H. Soares, C. Zimmer, and R. Henriques, "Single-molecule localization super-resolution microscopy: deeper and faster," *Microsc. Microanal.* **18**, 1419–1429 (2012).
34. R. Pandharkar, A. Velten, A. Bardagjy, E. Lawson, M. Bawendi, and R. Raskar, "Estimating Motion and size of moving non-line-of-sight objects in cluttered environments," in *International Conference on Computer Vision and Pattern Recognition (CVPR)* (2011), pp. 265–272.
35. M. Laurenzis and A. Velten, "Non-line-of-sight active imaging of scattered photons," *Proc. SPIE* **8897**, 889706 (2013).
36. L. F. Gillespie, "Apparent illumination as a function of range in gated, laser night-viewing systems," *J. Opt. Soc. Am.* **56**, 883–887 (1966).
37. M. Laurenzis and A. Velten, "Nonline-of-sight laser gated viewing of scattered photons," *Opt. Eng.* **53**(2), 023102 (2014).

## 1. Introduction

Reconstruction of three-dimensional (3D) objects through scattering media is a long-sought goal in many applications involving, e.g., long-range imaging through turbulence [1], deep-tissue biological imaging [2], non-invasive inspection of artwork [3,4], and underwater detection [5]. In these cases, scattering corrupts an optical signal and precludes direct image formation. Consequently, several modalities, including coherent wavefront correction [6–8], adaptive optics [9,10], terahertz imaging [11], and coherence- and time-gating [12–17], have been devel-

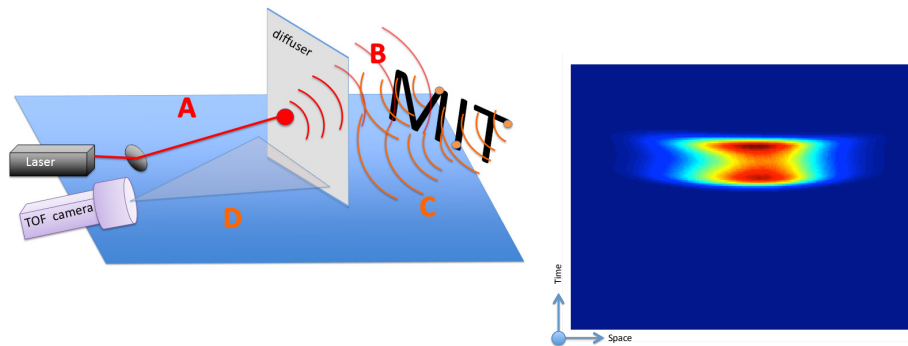


Fig. 1. Light propagation (left) and synthesized streak camera image (right). Coherent light (A) hits the diffuser in one position, scattered behind the diffuser (B) hits the object, scattered once again from many positions (C), hits the diffuser a second time, and reaches the streak camera (D) which measures one spatial line over time, generating a 2D image. Window:  $1\text{ ns} \times (\approx)10\text{ cm}$ .

oped to mitigate scattering effects. Usually, for strong scattering or volumetric imaging, these methods require multiple measurements, raster scanning, or time scanning. The consequent increase in acquisition time is often impractical for dynamic scenes of interest.

On the other hand, scattering is a natural “information mixer,” with each scattered wave containing information about the entire scene [18]. In this respect, scattering can be exploited (rather than mitigated) for object estimation using far fewer measurements than needed in existing techniques. This intuition is especially useful when only recovery of low-dimensional data is required. In such cases, the problem is simplified to estimation of several unknown parameters, such as position and orientation, and has many applications in pose estimation, 3D tracking, and target recognition. However, relatively little work has studied pose estimation through turbid layers.

Here, we demonstrate a computational method for estimating the six degrees of freedom (DOF), the 3D location and the orientation, of rigid objects hidden behind a diffuse layer. The method relies on non-invasive, time-resolved sensing of scattered light. We extend previous time-resolved imaging methods using diffuse reflections [19, 20] or scattering [21] using a streak camera. Coupling prior information about the object geometry with the inherent mixing that scattering provides, we show that the acquisition ultimately requires only a single illumination point, rather than a raster scan. Our work here extends the technique by recognizing that a full image reconstruction is often unnecessary, i.e., that recovery of fewer unknowns (e.g., location and orientation) in a low-dimensional data system is sufficient for many applications. This allows for (1) a simplified image acquisition without raster scanning and (2) a different numerical optimization algorithm for robust recovery.

This scenario is applicable in several areas where geometry either is captured through other modalities or is known *a priori*. Therefore, one-shot localization has potential in locating individuals in hazardous environments (such as natural disasters or war zones), detecting objects remotely [22–26], tracking tumors or organs over time with less radiation exposure, or improving inversion in time-resolved diffuse optical tomography or fluorescence tomography [27, 28].

Schematically, the method is shown in Fig. 1. An ultrafast laser pulse (A) is directed toward a

diffuse layer, which scatters light (B) toward an object (C). Subsequently, light is scattered from the object back toward the diffuser, where it is then imaged (D) onto a time-resolved detector. The orientation and location of the object determines the path length of the scattered light and hence the intensity time profile of each sensor pixel. Therefore, as the object moves and rotates, the time-resolved scattered light image will change accordingly. The question we answer is the following: given a single illumination point, is it possible to recover the orientation of the object and track its motion through space? By considering the six-dimensional space consisting of 3D translations and 3D rotations, our problem thus changes from object reconstruction to optimization over the space of possible transformations.

The rest of the paper is organized as follows. Section 2 describes the forward light transport model, and Section 3 outlines the associated optimization method. The numerical algorithm is detailed in Section 4, followed by a presentation of results in Section 5. Features and extensions of the method are discussed in Section 6. Section 7 concludes the paper.

## 2. Forward light model

We model scattering and propagation in the geometric approximation, assuming no occlusions, with the geometric factors shown in Fig. 2 and outlined in Table 1. A pulsed laser ( $L$ ) is focused onto a thin diffuser ( $D$ ) to a point  $d_l$ , scatters towards a given object point  $w$  and back towards the diffuser at point  $d_c$ . A one-dimensional, time-resolved sensor observing the diffuser will record an image  $I_l(d_c, t)$ . For each laser position  $d_l$ , we have [21]

Table 1. Symbols used in Fig. 2.

Symbol	Description
$L$	Laser origin
$C$	Streak camera
$D$	Diffuser
$w$	Object point
$d_l$	Incident laser position
$d_c$	Position on $D$ where one ray from $w$ strikes
$\mathbf{1}_d$	Diffuser surface normal
$\mathbf{1}_w$	Object surface normal at $w$
$\overrightarrow{Ld_l}$	Light ray from $L$ to $d_l$
$\overrightarrow{d_l w}$	Light ray from $d_l$ to $w$
$\overrightarrow{wd_c}$	Light ray from $w$ to $d_c$
$\overrightarrow{cd_c}$	Light ray from $C$ to $d_c$
$\alpha$	Angle between $\overrightarrow{Ld_l}$ and $\mathbf{1}_d$
$\beta$	Angle between $\overrightarrow{wd_c}$ and $\mathbf{1}_d$
$\gamma$	Angle between $\overrightarrow{d_l w}$ and $\mathbf{1}_w$
$\delta$	Angle between $\mathbf{1}_w$ and $\overrightarrow{wd_c}$
$\eta$	Angle between $\overrightarrow{d_l w}$ and $\overrightarrow{Ld_l}$
$\zeta$	Angle between $\overrightarrow{wd_c}$ and $\overrightarrow{cd_c}$
$r_l$	Euclidean distance between $d_l$ and $w$
$r_c$	Euclidean distance between $d_c$ and $w$

$$I_l(d_c, t) = \int g(d_l, w, d_c) f_a(w) \delta(ct - |\vec{d_l w}| - |\vec{w d_c}|) dw \quad (1)$$

$$g(d_l, w, d_c) = \frac{I_0 N(\eta) N(\zeta)}{\pi} \frac{\cos \alpha \cos \gamma \cos \beta \cos \delta}{r_l^2 r_c^2},$$

where  $c$  is the speed of light, and  $g(d_l, w, d_c)$  is a physical factor that depends on only the system geometry and the diffuser properties.  $f_a(w)$  is the albedo of the object point  $w$  and  $I_0$  is the laser intensity. Angles  $(\alpha, \beta, \gamma, \delta, \eta, \text{ and } \zeta)$  and distances  $(r_l \text{ and } r_d)$  are defined by the coordinates  $w, d_l$ , and  $d_c$ , as shown in Fig. 2.  $N(\cdot) = \exp(-\mu^2/\sigma^2)$  is the Gaussian scattering profile of the diffuser, for scattering angle and mean  $\sigma$  and  $\mu$ , respectively. The delta function restricts all recorded light propagating along the path  $d_l \rightarrow w \rightarrow d_c$  to arrive at the same instant. Note here that, although we treat scattering in a single layer, multiple scattering primarily affects the signal-to-noise (SNR), rather than the time resolution, for the parameter space considered here [21].

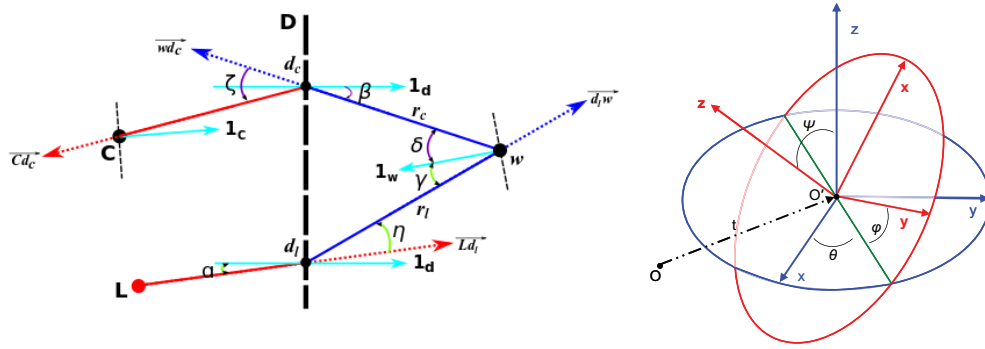


Fig. 2. Left: Pattern of light ray traveling from the laser  $L$ , through a diffuser  $D$  at point  $d_l$ , striking an object at point  $w$ , bouncing towards the diffuser at point  $d_c$ , and captured by the camera. Right: Spatial orientation of objects for the three Euler's angles.

### 3. Pose estimation

For a given experimental geometry, we model a rigid object  $M$  by a set of  $|M|$  uniformly spaced points  $w \in M \subset \mathbb{R}^3$ . The recorded image is determined by the location and orientation of  $M$ , which can be parametrized by six unknowns,  $\Theta$ :

$$\Theta = (t_x, t_y, t_z, \theta, \phi, \psi), \quad (2)$$

where  $t_*$  is the relative translation in the  $*$  direction, and  $\theta, \phi, \psi$  are the three Euler angles, measured about the center of mass  $\mathcal{C} = \frac{1}{|M|} \sum_{w \in M} w$ . We define the repositioning function  $\mathcal{L}_\Theta : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  to be

$$\mathcal{L}_\Theta(w) = R \times (w - \mathcal{C}) + \mathcal{C} + \vec{t}, \quad (3)$$

where  $R$  is the rotation matrix constructed from the Euler angles, and  $\vec{t} = (t_x, t_y, t_z)$  is the translation vector. Similarly,

$$\mathcal{L}_\Theta(M) = \{\mathcal{L}_\Theta(w)\}_{w \in M}. \quad (4)$$

Because not all points  $w$  are illuminated, we define the *frontal projection*  $\pi_l(M) \subseteq M$  as the points in space that are illuminated directly by diffuse light from laser position  $d_l$ . Every point  $p \in \pi_l(M)$  generates a hyperbolic curve in the streak image, and their sum is the total expected image,  $I_l(d_c, t)$ . We define the image of a point  $w$  generated from laser position  $l$  as  $\mathcal{S}_l(w)$ .

Our goal is to recover the pose that best agrees with the measured data. Hence, given a set of streak images  $I_l$  and an a prior on the geometry  $M$  we search for the unknown pose  $\hat{\Theta}$

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum_l \rho \left( I_l(d_l, t), \sum_{w \in \pi_l(\mathcal{L}_\Theta(M))} \mathcal{S}_l(\mathcal{L}_\Theta(w)) \right), \quad (5)$$

where  $\rho(A, B)$  is the “stress,” a measure of agreement between two quantities  $A$  and  $B$ . Here, we used the  $L_1$  metric between images, i.e.,  $\rho(A, B) = \|A - B\|_1$ .

#### 4. Optimization scheme

Equation (5) resembles the Iterative Closest Points (ICP) paradigm [29,30], where rigid motion is estimated in alternating steps of correspondence search and energy minimization to find the optimal transformation of an estimated point set into a reference one. However, our system is not amenable to the ICP algorithm because of the added complexity due to the time dimension: each object point generates a space-time hyperbola ( $\mathcal{S}_l(w)$  above) [19]. Thus, overlapping space-time curves and noise pose new challenges than those encountered in ICP. Consequently, we do not minimize the spatial location between matching points, but instead compare their nonlinear space-time functions.

To solve the Eq. (5), we use a stochastic gradient descent numerical optimization approach [31], in which each parameter is minimized separately. For each of the six parameters, we synthesize, according to our forward model, the expected streak images for two optional positions: one above and one below the current value. The step size is dynamic and is reduced for each subsequent iteration. In each step, 13 images, two options per unknown plus the identity, are synthesized, and the best option that reduces the stress is chosen. The procedure was terminated when the stress reached a threshold, or when the change in parameters was infinitesimal. In each iteration,  $\pi_l$  should be re-evaluated, but in our experiments there was no need due to the narrow geometric structure of the objects in usage. The procedure is summarized in Alg. 1.

---

##### Algorithm 1: Localization algorithm

---

**Input:** Scene geometry. ( A model  $M$  / Laser point locations  $L$  / inner parameters of the diffuser, camera, and laser )

**Output:**  $\hat{\Theta} = (t_x, t_y, t_z, \theta, \phi, \psi)$ . Position and orientation of the model in space.

- 1 Set  $\Theta^k = (0, 0, 0, 0, 0, 0)$ .
  - 2 Set  $sT = 30$ . Set  $sA = \frac{30\pi}{180}$ . (Chosen arbitrarily)
  - 3 Calculate  $\pi_l(\mathcal{L}_{\Theta^k}(M))$  for every  $l \in L$  by choosing only the visible points with relation to laser position  $l$ .
  - 4 Calculate  $\rho$  for 13 options:  $\Theta^k$  and  $\Theta^k + (\pm sT, \pm sT, \pm sT, \pm sA, \pm sA, \pm sA)$ .
  - 5 Choose  $\Theta^{k+1}$  as the minimiser of all  $\rho$ 's.
  - 6 If  $\Theta^k = \Theta^{k+1}$  then  $sT = \frac{sT}{2}$  and  $sA = \frac{sA}{2}$ .
  - 7 If  $\rho \leq \varepsilon_1$  or  $sT \leq \varepsilon_2$  or  $sA \leq \varepsilon_3$  for predetermined  $\varepsilon_*$ , set  $\hat{\Theta} = \Theta^{k+1}$  and exit.
  - 8 Goto step 3
-

## 5. Results

### 5.1. Synthetic experiments

#### 5.1.1. Pose estimation

We performed 2352 synthetic experiments for different spatial positions and orientations, and we evaluated the success rate and convergence quality. For an object, we chose three characters, “M,” “I,” and “T,” all lying in the same plane. See Fig. 1. For this object we generated 7 different spatial translations and rotated each of the Euler angles 7 times in increments of  $20^\circ$ , for a total of 49 poses. We evaluated each location and orientation using a single laser position, sets of three positions (per row and per column), and all 9 positions together, as can be seen in Table 2. The set of laser positions determines the summation over  $l$  in Eq. (5). We repeated each experiment three times with different values of SNR (100, 75, 50 using awgn MATLAB code), which corresponds to both spatial and temporal intensity noise.

Table 2. Depth (X), in planar (YZ), and angular convergence for the synthesized MIT banner given 2352 experiments. Each column in the table represent a different translation, and each row a different orientation. In each cell, we averaged the mean error of 48 evaluated distances or angles (16 different laser spots locations times 3 different SNR levels), and the percentage of experiments that converged to a minima near the correct solution (right member of each cell). Depth (X) can be recovered with high accuracy, and in-plane translations (YZ) can be recovered in most circumstances. Angular changes which have strong spatial shift ( $\psi$  for the MIT banner) are found with high accuracy, while other rotations are harder to evaluate.

Depth (X) convergence in mm / success rate							
	0	X+20	Y+20	Z+20	X-20	Y-20	Z-20
0	0.00 / 100.00	0.00 / 85.42	0.00 / 100.00	0.00 / 100.00	0.00 / 100.00	0.00 / 100.00	0.01 / 100.00
$\theta + 20$	0.00 / 100.00	2.49 / 100.00	2.49 / 100.00	0.84 / 100.00	1.78 / 91.67	1.94 / 93.75	2.36 / 97.92
$\phi + 20$	0.00 / 100.00	0.00 / 100.00	0.00 / 100.00	10.00 / 100.00	0.00 / 100.00	0.00 / 100.00	9.94 / 100.00
$\psi + 20$	0.00 / 100.00	0.00 / 77.08	0.00 / 100.00	0.00 / 100.00	0.00 / 100.00	0.01 / 100.00	0.00 / 100.00
$\theta - 20$	2.50 / 100.00	2.50 / 100.00	2.49 / 100.00	2.34 / 97.92	2.48 / 100.00	2.53 / 100.00	1.85 / 100.00
$\phi - 20$	0.00 / 100.00	0.01 / 87.50	0.00 / 100.00	4.19 / 100.00	0.02 / 100.00	0.00 / 100.00	0.00 / 100.00
$\psi - 20$	0.00 / 100.00	0.00 / 91.67	0.00 / 100.00	0.00 / 100.00	0.00 / 100.00	0.00 / 100.00	0.02 / 100.00
In-plane (YZ) norm ( $L_2$ ) convergence in mm / success rate							
	0	X+20	Y+20	Z+20	X-20	Y-20	Z-20
0	0.00 / 100.00	0.89 / 93.75	0.00 / 100.00	0.02 / 100.00	0.00 / 100.00	0.00 / 100.00	2.86 / 100.00
$\theta + 20$	0.00 / 100.00	1.82 / 83.33	3.56 / 100.00	2.01 / 83.33	0.49 / 77.08	5.84 / 100.00	6.47 / 77.08
$\psi - 20$	0.40 / 100.00	0.70 / 100.00	1.36 / 100.00	20.00 / 4.17	0.00 / 100.00	0.78 / 100.00	17.27 / 72.92
$\psi + 20$	0.15 / 100.00	1.94 / 85.42	0.22 / 100.00	3.75 / 100.00	0.63 / 100.00	6.58 / 100.00	6.24 / 100.00
$\theta - 20$	3.22 / 100.00	3.22 / 100.00	3.28 / 100.00	3.99 / 75.00	3.81 / 100.00	8.26 / 97.92	4.99 / 58.33
$\phi - 20$	6.74 / 87.50	2.34 / 93.75	9.48 / 100.00	10.34 / 33.33	10.07 / 100.00	4.49 / 100.00	10.17 / 8.33
$\psi - 20$	0.01 / 100.00	1.39 / 100.00	2.47 / 100.00	1.83 / 100.00	0.00 / 100.00	0.26 / 100.00	6.23 / 75.00
Angular norm ( $L_2$ ) convergence in degrees / success rate							
	0	X+20	Y+20	Z+20	X-20	Y-20	Z-20
0	0.00 / 100.00	0.00 / 93.75	0.00 / 100.00	0.00 / 100.00	0.00 / 100.00	0.00 / 100.00	0.00 / 75.00
$\theta + 20$	0.00 / 100.00	0.00 / 75.00	0.00 / 41.67	0.01 / 75.00	0.00 / 75.00	0.02 / 33.33	0.00 / 41.67
$\phi + 20$	0.00 / 89.58	0.00 / 72.92	0.00 / 75.00	0.06 / 41.67	0.00 / 100.00	0.00 / 62.50	0.06 / 39.58
$\psi + 20$	0.00 / 100.00	0.00 / 60.42	0.01 / 100.00	0.00 / 66.67	0.01 / 89.58	0.00 / 31.25	0.00 / 50.00
$\theta - 20$	0.00 / 75.00	0.00 / 75.00	0.01 / 72.92	0.00 / 58.33	0.00 / 75.00	0.05 / 31.25	0.00 / 33.33
$\phi - 20$	0.00 / 18.75	0.00 / 37.50	— / 0.00	— / 0.00	— / 0.00	0.00 / 39.58	— / 0.00
$\psi - 20$	0.00 / 100.00	0.00 / 83.33	0.00 / 75.00	0.00 / 75.00	0.00 / 100.00	0.00 / 93.75	0.01 / 41.67

Table 2 shows the mean norm error between the correct data and the calculated one, after removing outliers that converged to local minima far ( $5^\circ$ , 1 cm in the X direction (depth), and 2 cm in YZ plane) from the real shift. Each column in the table represents a different translation, and each row a different orientation. Each cell contains the averaged mean error of 48 instances of the corresponding parameter estimate (the 16 different sets of laser positions times the 3

different SNR levels above) and the percent of experiments that converged to a minima near the correct solution. The table shows that depth (X) can be recovered with high accuracy, and in-plane translations (YZ) can be recovered in most circumstances. Angular changes which translates to a strong spatial shift ( $\psi$  for the MIT banner) are found with high accuracy, whereas other rotations are harder to evaluate.

### 5.1.2. Tracking

We synthesized 180 time-sequential frames of the MIT banner rotating in  $2^\circ$  increments of  $\phi$  (out of plane angle). We used the same algorithm as before and initialized each step with its predecessor outcome. In Fig. 3, we can see that the rotating banner has slight changes for different angles, yet we were able to track the rotating with very high accuracy. Notice that in several angles we encounter ambiguity in the streak images. Even though there is no true symmetry with relation to the laser position, the images are numerically similar. We can see a decrease in accuracy for those angles in Fig. 3. Notice that noise was added to the synthesized data (35 SNR using awgn MATLAB code) in both space and time.

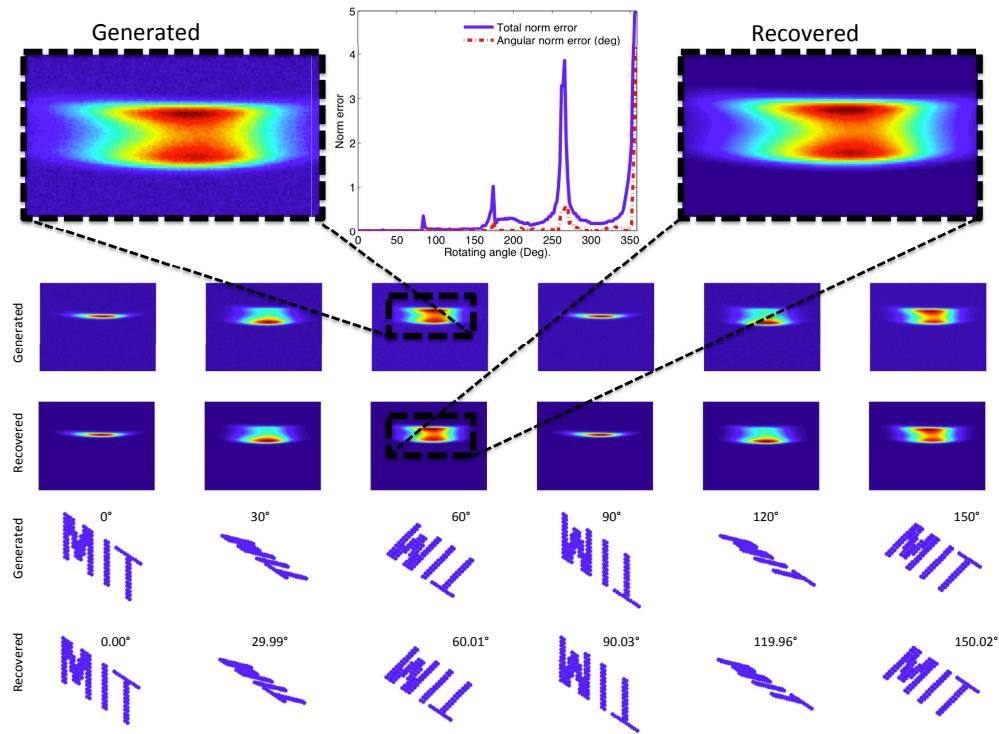


Fig. 3. Tracking of a rotating MIT banner from scattered light. Row 1: enlarged images. Row 2: synthesized noisy data, Row 3: recovered images, Row 4: real orientation of the banner, Row 5: calculated orientation. 180 images with 2 degrees change in  $\psi$  (in-plane angle) are synthesized and tracked using the proposed algorithm. Even though, the synthesized images have slight changes for different angles and high noise ratio (35 SNR), we were able to track the rotation with very high accuracy. Notice the ambiguity in the current setup which relates to higher error norm shown in the first row. The two graphs measure different quantities in different units but overlaid for compactness. Errors appear when the laser position does not hold enough meaningful information to infer the pose. Window:  $1\text{ ns} \times (\approx)10\text{ cm}$ .



Table 3. Translation (in mm) and angular (in degrees) of real experiments. Distances and angles are measured with reference to one spatial setup (R). While not mandatory, it provides an additional alignment. In the top table, we see that one laser position is sufficient for angle evaluation to within  $1\text{-}2^\circ$ , and from the tables below, we learn that depth (X) can be evaluated with high accuracy, with in-plane movements (YZ) are harder to find.

Angles A and B with relation to the angle in position R  
Real angle A: 16.84, Real angle B: 32.79 in degrees

Pos	Calc angle A / Diff	Calc angle B / Diff
1	16.17 / 0.67	30.93 / 1.85
2	15.46 / 1.38	29.79 / 2.99
3	14.23 / 2.61	30.05 / 2.73
1,2	15.82 / 1.02	30.93 / 1.85
2,3	14.85 / 1.99	29.79 / 2.99
1,3	15.55 / 1.29	30.76 / 2.03
1,2,3	15.64 / 1.20	30.32 / 2.47

Position C with relation to position R:  
Real X -25.85 / Y 0.697 / Z 0.733 in mm

Pos	Calc X / Diff	Calc Y / Diff	Calc Z / Diff
1	-24.68 / 1.16	3.35 / 2.266	-3.98 / 4.71
2	-24.92 / 0.93	5.27 / 4.57	-8.51 / 9.24
3	-25.00 / 0.85	-0.07 / 0.77	-5.46 / 6.20
1,2	-24.72 / 1.12	3.59 / 2.89	-5.93 / 6.67
2,3	-24.92 / 0.93	2.38 / 1.68	-7.34 / 8.07
1,3	-24.68 / 1.16	5.19 / 4.49	-10.27 / 11.00
1,2,3	-25.00 / 0.85	2.89 / 2.19	-4.68 / 5.42

Position D with relation to position R:  
Real X 0.06 / Y 25.23 / Z 0.43 in mm

Pos	Calc X / Diff	Calc Y / Diff	Calc Z / Diff
1	-0.39 / 0.45	18.75 / 6.48	1.56 / 1.12
2	-0.15 / 0.21	32.81 / 7.57	-25.54 / 25.98
3	-1.25 / 1.31	27.34 / 2.10	-9.21 / 9.65
1,2	0.27 / 0.21	30.89 / 5.66	-23.04 / 23.48
2,3	-0.62 / 0.68	29.14 / 3.90	-16.71 / 17.15
1,3	-0.31 / 0.37	27.92 / 2.69	-12.53 / 12.97
1,2,3	-0.07 / 0.14	27.53 / 2.30	-15.85 / 16.29

## 5.2. Laboratory experiments with streak camera

The experimental setup is photographed in Fig. 4 and sketched in Fig. 5. A mode-locked Titanium:Sapphire laser emits pulses of duration 50 fs at a repetition rate of 80 MHz with average power of 1 W. The pulse train is focused and directed via two computer-controlled galvo mirrors onto a  $10 \times 10 \text{ cm}^2$  holographic diffuser (Edmund Optics, 65-886), which has a scattering angle of approximately  $\sigma = 8.8^\circ$ . The objects behind the diffuser are Lambertian. The front

side of the diffuser is imaged onto a streak camera (Hamamatsu C5680), which records a horizontal line view of the diffuser with a time resolution of 2 ps. The galvo mirrors relocate the incident laser to different spatial positions, and a streak image is recorded for each one, though for most experiments, we use only one laser position. To improve SNR, the exposure time for each streak image is 10 s (so that each image averaged  $10\text{ s} \times 80\text{ MHz} = 8 \cdot 10^8$  pulses). The estimated pose is compared to the object's ground truth pose, which is measured with a FARO digitizer arm. We measure three points on the base of the shape and calculate the Euler angles and positions accordingly. The object volume was approximately  $10 \times 5 \times 1\text{ cm}^3$  (height width depth) and was located 18 cm from the diffuse layer. The diffuser itself was placed about 0.5 m from the streak camera. The space-time (x-t) pixel resolution of the streak camera is  $672 \times 512$ . In addition to the galvo-controlled signal beam, a beamsplitter was used to create a separate calibration beam, which was focused onto the diffuser in the field of view of the camera. This allows a continual reference measurement to correct for any timing or intensity noise in the laser itself. Because a first-principles calculation of the optimal point source location is beyond the scope of the manuscript, we measure streak images from multiple incident laser spots. But, we show that only using one in the algorithm allows for successful pose estimation.

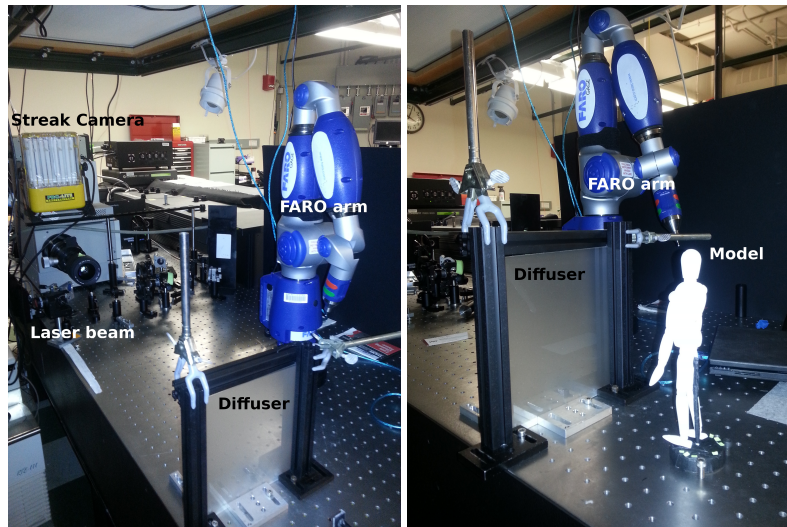


Fig. 4. The human model behind the diffuser. A photograph of the experiment from the model's point of view.

We performed two separate experiments using a human-like figure. The first was dedicated to estimating angular changes and the second for estimating translations. For the angle estimation, we searched for 3 different angles rotating around the vertical axis, orthogonal to the table. We used 3 different laser positions and evaluated the angle with 7 different sets of measurements (3 individual laser positions, all pairs of laser positions, and all three laser positions). We compared our results with the object's ground truth measurement, which has a precision of approximately 3 mm. We rotated the object clock-wise by 16 and 32 degrees and tried to recover those angles.

For translation measurements, we used a similar procedure for 3 different locations of the object. We moved the object by 25 mm, both in-plane and out-of-plane. We searched for only the unknown parameters (one rotating angle for the first experiment, and three translation vectors (one for each location) for the second).

A summary of estimation of the angular parameter is shown in in Table 3. We used the angle measured in one orientation as a reference (R) and measured the angles of the second

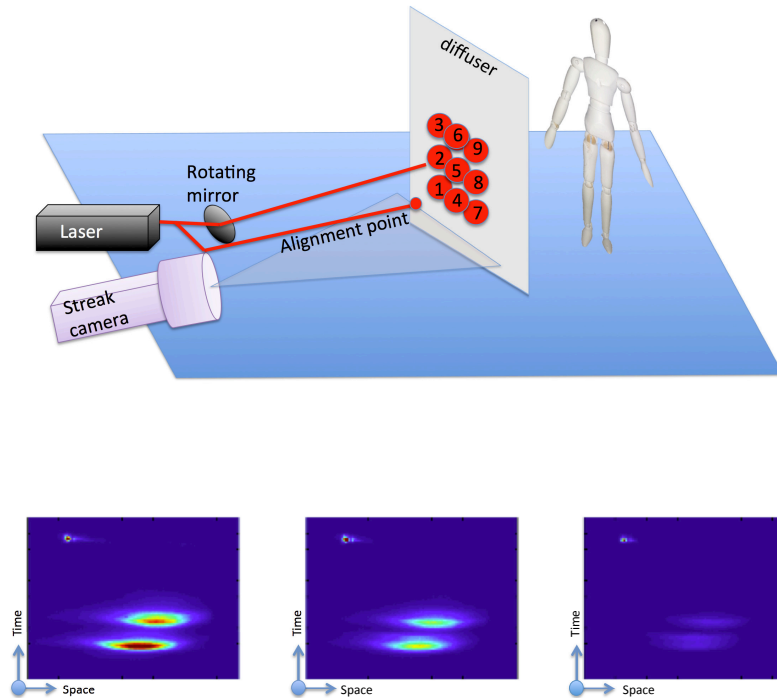


Fig. 5. Scene setup for a human-like model. We used several laser spots for different positions and orientations of the model in space. Bottom row: raw images captured by the streak camera (positions 1,2 and 3 from right to left). The top-left bright spot is a physical reference beam used for spatio-temporal alignment. Window:  $1\text{ ns} \times (\approx)10\text{ cm}$ . The source is split into two beams with a beamsplitter. The small focused spot in the upper-left corner of each measurement is the calibration beam, used to normalize the global intensity of the data and correct for any timing jitter. The second beam scatters through the diffuser toward the object, and is scanned across the diffuser with a galvo mirror to generate multiple streak images.

and third orientations (A and B, respectively) relative to it. The reference orientation provides a calibration measurement to compensate our light transport model for small changes ( $\approx 1^\circ$ ).

The bottom row of laser positions (1,4,7 in Fig. 5) provided the best results, less than a  $1^\circ$  error for the first comparison and less than a  $2^\circ$  error for the second. This is within the error bounds of our true data measurements using the robotic arm. The top positions gave the worst results but still less than a  $3^\circ$  error. Options 4 through 7 in the table show different variations of laser positions and error values. Because we do not know *a priori* which location of laser spot will provide the best results, we conclude that using all three is the best approach. Figure 6 shows the acquired streak images and steps 1, 5, 10, and the last step (from right to left) in the optimization procedure. In the first row, we see the convergence for position A, and in the second row for position B, both using only laser position 1 (see Fig. 5). We recorded 50 sample points of the object with the robotic arm. Because it is a coarse, noisy version of the model, it yields slight differences between the synthesized images and the measured ones.

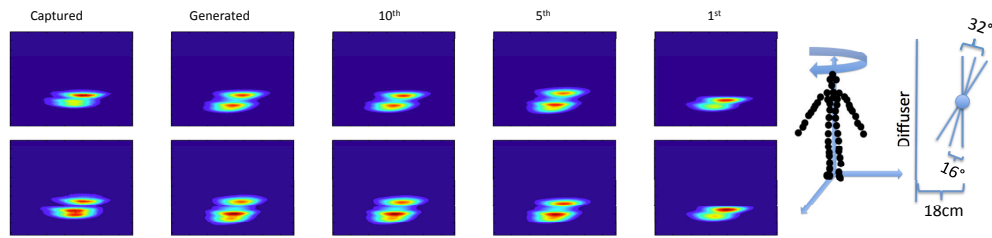


Fig. 6. Convergence of the algorithm for laser position 3 (see Fig. 5). First row: angle A. Second row: angle B. From right to left: initial iteration, 5th iteration, 10th iteration, final generated model, and real captured image. A threshold was added to remove noise. Window:  $1\text{ns} \times (\approx)10\text{ cm}$ .

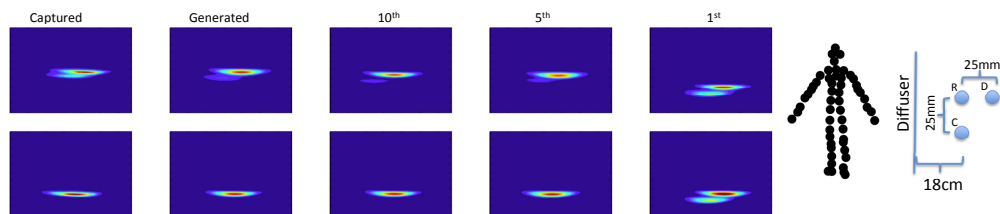


Fig. 7. Convergence of the algorithm for laser position 1 (see Fig. 5). First row: translation C. Second row: translation D. From right to left: initial iteration, 5th iteration, 10th iteration, final generated model, and real captured image. A threshold was added to remove noise. Window:  $1\text{ns} \times (\approx)10\text{ cm}$ .

The translation results are shown in Table 3. Once more, we used one location as a reference R, and measured the other two (C and D) accordingly. The depth (X axis) is calculated with high accuracy, below 1 mm, which is within the error bound of our ground truth data. Movement parallel to the diffuser is harder to evaluate and differs for each laser position. With all three positions used, we achieved approximately 2 mm accuracy along the Y axis (orthogonal to the table) and between 0.5 cm and 1.5 cm along the Z axis (parallel to the table). In Fig. 7, we present the corresponding convergence images. The first row corresponds to position C, and the second row to position D, both given laser position 1 (see Fig. 5).

## 6. Discussion

Generally, our pose estimation method yields good results for a single incident laser position, though more positions provide angular diversity and improves the result. Interestingly, stronger/multiple scattering is a benefit, because the increasingly scattered light provides wider and more uniform illumination compared to that from a diffuse layer. Nevertheless, a first-principles derivation of the minimum number of streak images must be performed to understand the role of noise in the optimization problem. The generalization to thick turbid media is also possible, provided that time resolution of detector can overcome any resulting temporal blurring. For biological samples of moderate thickness, the streak camera resolution will suffice [21]. This can be useful for dynamic scenes, such as tracking the flow of blood cells [32], and can be integrated with super-resolution fluorescence microscopy techniques in thick media [33], assuming the temporal profile still holds enough information.

Additionally, because no raster scanning is needed, the acquisition time is fast. Although we

average over a pulse train for improved SNR, the method is amenable for a single incident pulse if, for example, the laser source is amplified, or the camera sensitivity is increased. Coupled with scaled versions of the method [35, 37], this technique is thus readily applicable for long-range imaging after appropriate calibration [36].

Interestingly, recovery is unique up to certain ambiguous symmetries. In Fig. 3 we see degraded accuracy near the symmetric axes (every 90 degrees). However, the symmetry can be easily broken by adding a second illumination source. Moreover, note that for tracking, we utilize prior information on the pose (calculated from the previous step) for convergence to the correct solution (cf. the second column from the left versus the second column from the right in Fig. 3 ).

## 7. Conclusions

In conclusion, the spatial location and angular orientation of objects hidden behind turbid layers can be estimated given a known prior of the shape and a single illumination point, with accuracy below 3 degrees, about 1mm in depth, and several millimeters in-plane, for the provided setup. Unlike previous work in the time-resolved inversion, we consider here a small number of incident illumination sources coupled to a mathematical optimization technique to retrieve the low-dimensional structure such as location and orientation of objects. In practice, even a single illumination point provides enough information, up to certain symmetries, for successful recovery. The method relies on time-resolved sensing and treats scattering as a benefit, rather than as a hindrance, and it can be integrated with other time-resolved methods of motion detection in cluttered environments. Future work includes: improving robustness to handle larger displacements, extending the method for integration in remote sensing applications, and considering volumetric scattering for biological applications.

## Acknowledgment

D.R. was partially funded by the Technion-MIT fellowship.