Imaging in scattering media using correlation image sensors and sparse convolutional coding

Felix Heide,[1,2,*] **Lei Xiao,**[1,2] **Andreas Kolb,**[3]
Matthias B. Hullin,[4] **and Wolfgang Heidrich**[1,2]

[1] *King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia*
[2] *University of British Columbia, 2329 W Mall, Vancouver, BC V6T 1Z4, Canada*
[3] *Universität Siegen, Hölderlinstraße 3, 57068 Siegen, Germany*
[4] *Universität Bonn, Regina-Pacis-Weg 3, 53113 Bonn, Germany*

[*]fheide@cs.ubc.ca

Abstract: Correlation image sensors have recently become popular low-cost devices for time-of-flight, or range cameras. They usually operate under the assumption of a single light path contributing to each pixel. We show that a more thorough analysis of the sensor data from correlation sensors can be used can be used to analyze the light transport in much more complex environments, including applications for imaging through scattering and turbid media. The key of our method is a new convolutional sparse coding approach for recovering transient (light-in-flight) images from correlation image sensors. This approach is enabled by an analysis of sparsity in complex transient images, and the derivation of a new physically-motivated model for transient images with drastically improved sparsity.

© 2014 Optical Society of America

References and links

1. B. Das, K. Yoo, and R. Alfano, "Ultrafast time-gated imaging in thick tissues: a step toward optical mammography," Opt. Lett. **18**, 1092–1094 (1993).
2. P. Han, G. Cho, and X.-C. Zhang, "Time-domain transillumination of biological tissues with terahertz pulses," Opt. Lett. **25**, 242–244 (2000).
3. J. B. Schmidt, Z. D. Schaefer, T. R. Meyer, S. Roy, S. A. Danczyk, and J. R. Gord, "Ultrafast time-gated ballistic-photon imaging and shadowgraphy in optically dense rocket sprays," Appl. Opt. **48**, B137–B144 (2009).
4. A. Kirmani, T. Hutchison, J. Davis, and R. Raskar, "Looking around the corner using transient imaging," in "Computer Vision, 2009 IEEE 12th International Conference on," (IEEE, 2009), pp. 159–166.
5. A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," Nat. Commun. **3**, 745 (2012).
6. R. Schwarte, Z. Xu, H. Heinol, J. Olk, R. Klein, B. Buxbaum, H. Fischer, and J. Schulte, "New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device," in "Proc. SPIE," , vol. 3100 (1997), vol. 3100, pp. 245–253.
7. R. Lange and P. Seitz, "Solid-state time-of-flight range camera," IEEE J. Quantum Electron. **37**, 390–397 (2001).
8. M. Lindner, I. Schiller, A. Kolb, and R. Koch, "Time-of-flight sensor calibration for accurate range sensing," Computer Vision and Image Understanding **114**, 1318–1328 (2010).
9. F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, "Low-budget transient imaging using photonic mixer devices," ACM Trans. Graph. (Proc. SIGGRAPH 2013) **32**, 45:1–45:10 (2013).
10. D. Freedman, E. Krupka, Y. Smolin, I. Leichter, and M. Schmidt, "SRA: Fast removal of general multipath for tof sensors," arXiv preprint arXiv:1403.5919 (2014).

11. D. L. Donoho, "Compressed sensing," IEEE Trans. Inform. Theory **52**, 1289–1306 (2006).
12. E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Comm. Pure Appl. Math. **59**, 1207–1223 (2006).
13. Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications* (Cambridge University, 2012).
14. A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar, "Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles," ACM Trans. Graph. (Proc. SIGGRAPH Asia 2013) **32**, 167 (2013).
15. A. Bhandari, A. Kadambi, R. Whyte, C. Barsi, M. Feigin, A. Dorrington, and R. Raskar, "Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization," Opt. Lett. **39**, 1705–1708 (2014).
16. M. O'Toole, F. Heide, L. Xiao, M. B. Hullin, W. Heidrich, and K. N. Kutulakos, "Temporal frequency probing for 5d transient analysis of global light transport," ACM Trans. Graph. (Proc. SIGGRAPH) **33** (2014).
17. A. Velten, R. Raskar, and M. Bawendi, "Picosecond camera for time-of-flight imaging," in "Imaging and Applied Optics," (Optical Society of America, 2011), p. IMB4.
18. E. Grushka, "Characterization of exponentially modified gaussian peaks in chromatography," Anal. Chem. **44**, 1733–1738 (1972).
19. P. Datte, A. M. Manuel, M. Eckart, M. Jackson, H. Khater, and M. Newton, "Evaluating radiation induced noise effects on pixelated sensors for the national ignition facility," (2013), vol. 8850, pp. 885003–885003–9.
20. M. S. Lewicki and T. J. Sejnowski, "Coding time-varying signals using sparse, shift-invariant representations," in "Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II," (MIT, Cambridge, MA, USA, 1999), pp. 730–736.
21. M. Mørup, M. N. Schmidt, and L. K. Hansen, "Shift invariant sparse coding of image and music data," Tech. rep., Technical University of Denmark, Richard Petersens Plads bld. 321, 2800 Kgs. Lyngby, Denmark (2008).
22. M. D. Zeiler and R. Fergus, "Learning image decompositions with hierarchical sparse coding," Tech. Rep. TR2010-935, Courant Institute of Mathematical Science, New York University (2010).
23. M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in "Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on," (IEEE, 2010), pp. 2528–2535.
24. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in "Adv. Neural Inf. Process. Syst.", (2012), pp. 1097–1105.
25. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," Foundations and Trends® in Machine Learning **3**, 1–122 (2011).
26. H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," in "Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on," (IEEE, 2013), pp. 391–398.
27. J. Lin, Y. Liu, M. B. Hullin, and Q. Dai, "Fourier analysis on transient imaging with a multifrequency time-of-flight camera," in "IEEE Conference on Computer Vision and Pattern Recognition (CVPR)," (2014).
28. R. Tibshirani, "Regression shrinkage and selection via the lasso," J. R. Stat. Soc. Ser. B Stat. Methodol. pp. 267–288 (1996).
29. J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," IEEE Trans. Inform. Theory **53**, 4655–4666 (2007).
30. S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in "Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on," (IEEE, 2011), pp. 2168–2172.
31. F. Heide, L. Xiao, W. Heidrich, and M. B. Hullin, "Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors," in "IEEE Conference on Computer Vision and Pattern Recognition (CVPR)," (2014).

1. Introduction

Imaging through scattering media has recently received a lot of attention. While many works have considered microscopic settings such as imaging in biological tissue [1, 2], we consider here the macrosopic problem, with ultimate target applications such as underwater imaging or imaging through fog.

Traditional approaches such as imaging individual femtosecond laser pulses with either fast gated cameras [1, 3] or streak cameras [4, 5] do not scale well to this macroscopic setting, since ambient illumination can easily overpower the laser pulse, and even in the best case the limited laser power results in a low signal-to-noise ratio (SNR).

In this work we show that correlation image sensors such as photonic mixer devices [6, 7] can be effectively used for imaging in scattering and turbid media, when combined with computational analysis based on sparse coding (Fig. 1). Correlation image sensors are widespread on the consumer market in the form of range cameras that employ the time-of-flight (ToF)

principle. We use here the term "correlation image sensor" to avoid confusion with other ToF technologies. The primary advantage of correlation image sensors is that they have extended exposure intervals, much like regular cameras, and they integrate over many (ca. 10^4–10^5) pulses of a modulated light source, instead of a single pulse, resulting in a significant improvement in SNR, as well as slower, less expensive electronics.

Fig. 1. Example of imaging in scattering media using our approach. Left: Original scene with objects submerged in water-filled glass tank. Center: 160 ml milk added. Right: Objects that are "invisible" due to strong scattering, like the structured object on the top right or parts of the circular object become detectable using our approach.

2. Imaging with correlation sensors

Before introducing our new approach, in this section we summarize the operating principles of correlation imagers, the most common form of time-of-flight technology on the market. We will use the term "correlation image sensor" to avoid confusion with fast gated cameras, streak cameras with pulsed lasers, and other technical approaches.

2.1. Time-of-flight imaging using temporal coding

To explain the principle of a ToF camera based on temporal coding, let us assume a setup with a single diffuse reflector as shown in Fig. 2.
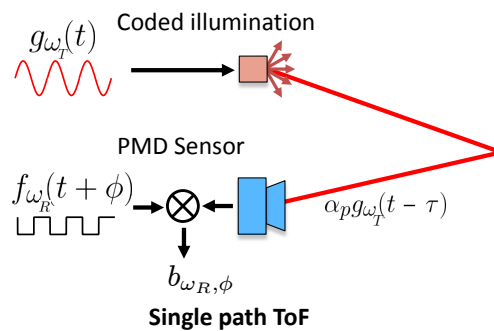

Fig. 2. Time-of-flight depth imaging assuming a single light path.

The light source emits intensity-modulated light $g(t)$ with a periodic modulation signal into the scene. For a single camera pixel, and under the assumption of a single light path with travel time τ contributing to this pixel, the illumination at the camera is a phase shifted signal $s(t)$ with a reduced amplitude due to surface albedo and geometric falloff. A correlation image sensor such as a photonic mixer device integrates (exposes) over a large number ($\approx 10^4$–10^5) of periods of this incident illumination, while modulating it with a periodic reference signal $f(t)$, thus essentially computing a correlation between s and f:

$$g(t) = \sum_{k=0}^{\infty} g_k \cos(\omega_T t)$$

$$s(t) = \alpha_p \, g(t - \tau) + I \qquad (1)$$

$$f(t) = \sum_{k=0}^{\infty} f_k \cos(\omega_R t + \phi).$$

Here ω_T is the base frequency of the illumination, ω_R is the base frequency of the sensor reference signal, ϕ is a programmable phase offset for the reference signal, and I is an offset to account for ambient illumination. The harmonics of the base frequencies are either assumed to be zero (i.e. $g_k = f_k = 0$ for $k > 1$), or can be obtained through calibration. Usually, correlation image sensors are operated in homodyne mode, i.e., $\omega_T = \omega_R$, in which case the measured correlation is given as

$$p(\phi) = f \star s = \mathfrak{F}^{-1}\Big((\overline{\mathfrak{F}(s)}(\xi) \cdot \mathfrak{F}(f)(\xi)) \Big) = \frac{\alpha_p}{2} \sum_k f_k g_k \, \cos(\phi - \omega_T \tau) \qquad (2)$$

This function is sampled by each correlation image pixel for different relative phase shifts ϕ. The travel time τ can then be reconstructed using these samples. Usually the sampling pattern consists of four phase offsets $\phi_0 = 0, \phi_1 = \pi/2, \phi_2 = \pi, \phi_3 = 3/2\pi$ which allows for a solution using simple trigonometric identities, see for example Lindner et al. [8]. Having recovered the travel time τ, a distance estimate for the surface can be obtained simply as $d_{est} = \frac{c}{2} \cdot \tau$.

2.2. Transient imaging imaging using correlation sensors

ToF imaging with correlation sensors as described above makes strong assumptions on the scene that do not hold for realistic environments. As shown in Fig. 3, realistic scenes do not only contain single point-scatterers. Many objects with different reflection or scattering properties can exist, causing multiple path contributions to be linearly combined in a sensor pixel.
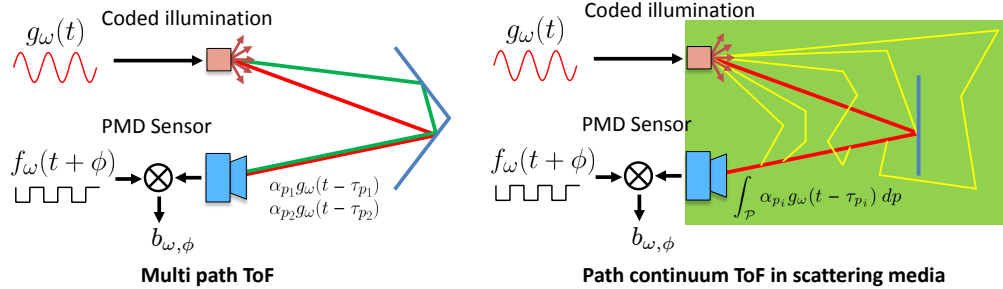

Fig. 3. Realistic scenes contain multiple light paths between source and sensor, and time-of-flight imaging has to account for them (left). Imaging in scattering media results in a continuum of paths of different lengths, resulting in a loss of sparsity in the time domain.

Heide et al. [9] first established a relation between the transient pixel $\alpha(\tau)$ (subsuming all multi-path contributions) and the resulting correlation measurement, as given in Eq. 3. For a given relative shift ϕ_i and frequency $\omega_i = \omega_T = \omega_R$, a correlation image pixel measures

$$b_{\omega_i, \phi_i} = \int_0^{\infty} \alpha(\tau) \cdot \underbrace{\int_0^T g_{\omega_i}(t - \tau) f_{\omega_i}(t + \phi_i) \, dt}_{= \mathbf{c}_{\omega_i, \phi_i}(\tau)} \, d\tau, \quad \text{with} \quad \alpha(\tau) = \int_{\mathscr{P}} \alpha_p \delta\left(|p| = \tau\right) dp, \quad (3)$$

where T is the integration time (which we assume to be an integer multiple of the modulation period) and $\alpha(\tau)$ physically represents the integral of all contributions from different light paths p that correspond to the same travel time τ. If we now discretize the travel time τ into M temporal bins and measure N different ϕ_i, ω_i pairs, we can formulate the measurement as a linear operator

$$\mathbf{C} = \left(\mathbf{c}_{\omega_i, \phi_i}(\tau_j)\right)_{i,j}, \tag{4}$$

where the correlation matrix $\mathbf{C}$ is obtained through a straightforward calibration step, imaging a planar target while cycling through the relative phase ϕ [9]. With $\alpha(\tau)$ discretized for the considered time steps, we obtain for a single pixel the measurement vector

$$\mathbf{b} = \mathbf{Ci} \quad \text{with} \quad \mathbf{i} = [\alpha(\tau_1), \ \alpha(\tau_2), \ \ldots, \ \alpha(\tau_M)]^T. \tag{5}$$

The unknown latent vector $\mathbf{i}$ is then the intensity profile of a pixel as a function of time, when the scene is illuminated by a Dirac pulse of light. This generalization of this vector to all pixels in an image is called the "transient image".

2.3. Transient imaging in scattering media

Now coming back to the multi-path problem illustrated in Fig. 3, in contrast to conventional intensity imaging, a transient image resolves the path travel time in an additional dimension. Thus, by measuring $\mathbf{b}$ and by inverting Eq. 5 for $\mathbf{i}$, one can untangle path contributions with different path length and thus remove unwanted multi-path contributions.

However, it is known that inverting Eq. 5 results in an ill-posed problem [9, 10]. This becomes obvious when choosing sinusoidal f and g, as then $\mathbf{C}$ becomes a truncated Fourier matrix, whose maximum temporal frequency is limited by the maximum achievable modulation frequency of typically 25–100 MHz for recent correlation-based camera systems. For imaging in scattering media multi-path contributions are even stronger than for regular ToF imaging as illustrated in Fig. 3 on the right. At a pixel a continuum of path lengths is measured, where only few ballistic photons directly hit objects submerged in the scattering media. This strong scattering, which makes traditional imaging very challenging, can only be handled if multi path contributions are removed effectively.

3. Method

3.1. Are transient images sparse?

The idea of using sparse representations for signal recovery was first analyzed extensively by Donoho and Candes et al. [11, 12] and since then has found use in many domains [13]. Several recent works [10, 14, 15] attempt to resolve multi-path interference by assuming sparsity of $\mathbf{i}$ in the Dirac basis. The vastly popular idea of compressed sensing [11] is applied in a straightforward manner, by trying to solve

$$\mathbf{i}_{\text{opt}} = \underset{\mathbf{i}}{\text{argmin}} \ \|\mathbf{i}\|_1 \quad \text{subject to} \quad \|\mathbf{Ci} - \mathbf{b}\|_2^2 < \varepsilon, \tag{6}$$

which is a convex relaxation of the sparsity requirement $\|\mathbf{i}\|_0 < K$, where K determines the sparsity, and hence can be solved using convex optimization methods. Commonly the ℓ_1 norm is used, which leads to solving for the MAP estimate $\mathbf{i}$ given a Laplacian prior distribution.

However, the basic assumption of $\mathbf{i}$ being sparse is violated for many realistic environments, and in particular in the case of scattering media. For regular scenes the (intensity modulated) radiosity leaving a single scene point integrates over a continuum of scene points, and thus in general cannot be sparse. For example, any concave object can be expected to deliver a non-sparse response, as shown by O'Toole et al. [16]. Especially for imaging in scattering media, the
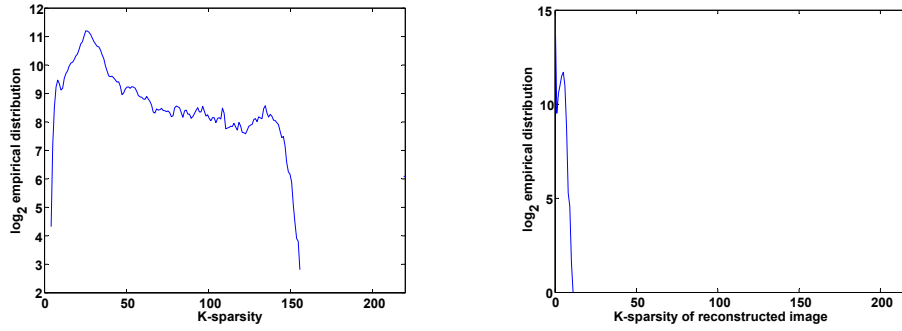
Fig. 4. Sparsity of 40K signals of a transient image measured image (left). Sparsity after fitting to the convolutional basis proposed in this paper (right).

assumption of temporal sparsity breaks down. Figure 4 on the left demonstrates this empirically by plotting the histogram of the ℓ_0 of a high-resolution transient image captured by Velten et al. [17].

The image depicts a scene composed of scattering objects (a tomato) and diffuse surfaces. However, note that even pixels that do not view a scattering object are not necessarily completely sparse, as we will show later in Section 3.5. The signal has been thresholded for 0.1 of the peak signal value along each pixel so as not to interpret sensor noise as sparse components. A large number of pixel signals have more than $K > 100$ components, which cannot be considered sparse, given a time discretization of $M = 220$ in this example.

3.2. A sparse local model for transient light interaction

Having shown that the popular sparse model does not apply in the Dirac basis for realistic scenes (and especially not for scattering media), we now introduce a model for local light transport interaction. This model leads to an overcomplete basis transforming the signal into a space where it is sparse/compressible.

We model the temporal point spread function (PSF) of direct local surface reflection from a single point as a Dirac peak, while the temporal structure of scattering processes is best represented by an exponential decay. In both cases, this PSF is convolved with a Gaussian that models the temporal PSF and resolution of the correlation image sensor. It is therefore plausible to describe a transient pixel by a mixture of exponentially modified Gaussians. Originally developed to describe chromatographic peaks [18], this versatile model has since been adopted in different fields such as high-energy physics, where it is used to model the response of pixel detectors [19]. A single exponentially modified Gaussian can be defined as

$$h(\tau; a, \sigma, \rho, \mu) = a \cdot \exp\left(\frac{1}{2}\left(\frac{\sigma}{\rho}\right)^2 - \frac{\tau - \mu}{\rho}\right) \cdot \left(1 + \mathrm{erf}\left(\frac{(\tau - \mu) - \sigma^2/\rho}{\sqrt{2}\sigma}\right)\right), \tag{7}$$

where here a (amplitude), σ (Gaussian width), ρ (skew) and μ (peak center) are the parameters of the exponentially-modified Gaussian function, and τ is the travel time at which it is evaluated. If we stack the parameters for a single exponentially-modified Gaussian in a vector $\mathbf{u} = [a, \sigma, \rho, \mu]^T$, then we can model the transient time-profile i as the mixture:

$$i(\tau) = \sum_{i=1}^{n} h(\tau; \mathbf{u}_i), \tag{8}$$

where n is here the number of mixture components. Figure 5 shows a few samples of the signals. We note that the joint modeling of the exponential and Gaussian nature of our signals

has a key advantage over previous models: the basis functions and hence the reconstructions are inherently smooth. The parametric model used by Heide et al. [9], on the other hand, produces discontinuous solutions whenever exponential components are being used.
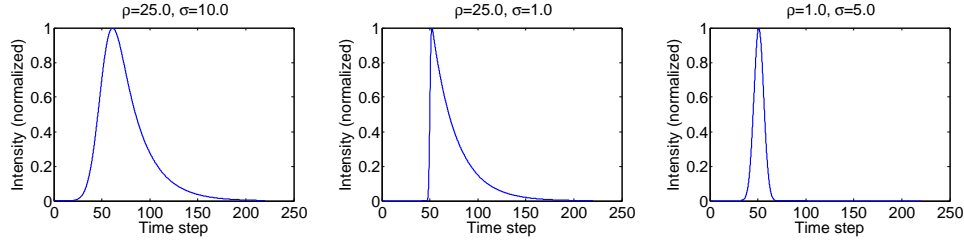

Fig. 5. A few samples of the exponentially modified Gaussian signals located at $\mu = 50$. Curves are normalized by adjusting amplitude a.

Now, fitting this model to the data from Fig. 4 (using the method described below), we can see that it defines a basis that transforms the transient image signals into a space that is significantly sparser than the signal itself (see Fig. 4 on the right).

3.3. Efficient optimization to find the mixture parameters

In order to find the optimal mixture parameters, commonly a non-convex non-linear problem is solved, which is prone to local minima and expensive to solve. Furthermore it leaves the number of mixing components open, which Heide et al. [9] solve in an alternate scheme that also offers no guarantee of global convergence.

We follow here a different approach and first linearize the above model by sampling the parameter space and later use it in a sparse convolutional basis pursuit fashion in a convex reconstruction problem. Linearizing the basis functions gives

$$\mathbf{H} = [h(\mathbf{s}; \mathbf{u}_0), \ldots, h(\mathbf{s}; \mathbf{u}_n)] \tag{9}$$

where $\mathbf{s}$ is the vector of all the sampling positions of the signal and n is the number of samples in the set containing all $\mathscr{C} = \{\sigma, \rho, \mu\}$ with $a = 1$. Using this (massively overcomplete) basis $\mathbf{H}$ we can formulate the reconstruction given a correlation image measurement $\mathbf{b}$ as the following basis-pursuit problem:

$$\mathbf{v}_{\text{opt}} = \underset{\mathbf{v}}{\arg\min} \; \frac{1}{2} \|\mathbf{CHv} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{v}\|_1 \tag{10}$$

Thus, the overcompleteness is handled by using the sparsity constraint. The values of a have now become the vector of basis coefficients $\mathbf{v}$. For simplicity, we consider only one pixel, although the approach can be trivially extended to multiple pixels/measurements.

However, this leads to an extremely large basis $\mathbf{H}$ (due to the large set $\mathscr{C}$), which makes the optimization inefficient even for problems of moderate size. One significant improvement to this situation is to exploit the convolutional nature of the basis. Instead of sampling the parameter space for μ (i.e., the translation of the peak along the time axis), one can fix μ and reformulate the problem as the following optimization problem:

$$\mathbf{v}_{\text{opt}} = \underset{\mathbf{v}}{\arg\min} \; \frac{1}{2} \left\| \mathbf{C} \sum_{i=0}^{k} \mathbf{H}_i \otimes \mathbf{v}_i - \mathbf{b} \right\|_2^2 + \lambda \|\mathbf{v}\|_1$$
$$\text{subject to} \quad \mathbf{H}_i \in h(\mathbf{s}, \mathscr{C}') \quad \forall i \in \{1, \ldots, k\} \tag{11}$$

where $\otimes$ is the convolution operator. The basis vectors are now sampled from the space of pulse *shapes*, $\mathscr{C}' = \{\sigma, \rho\}$, and invariant to translation. Therefore the size of the basis is drastically reduced from previously $n = k \cdot dim(\mathbf{v})$ to just k, i.e., by typically around 2 to 3 orders of magnitude.

This approach is motivated by the signal processing of the acoustic nerve, where *shift-invariant* sparse coding has first been proposed by Lewicki and Sejnowski [20]. Recently, sparse convolutional coding has also been used for audio and image detection tasks [21–23] motivated by the recent success of convolutional deep neural networks in image classification and detection [24].

3.4. Reformulation in the spectral domain

The problem from Eq. 11 can be defined even more compactly in the Fourier domain. The convolution reduces to pointwise multiplication, here expressed by the operator $\odot$:

$$
\begin{aligned}
\mathbf{v}_{\text{opt}} = \underset{\mathbf{v}}{\arg\min} \; & \frac{1}{2} \left\| \mathbf{C} \mathfrak{F}^{-1} \left(\sum_{i=0}^{k} \hat{\mathbf{H}}_i \odot \hat{\mathbf{v}}_i \right) - \mathbf{b} \right\|_2^2 + \lambda \sum_{i=0}^{k} \|\mathbf{t}_i\|_1 \\
\text{subject to} \quad & \mathbf{H}_i \in h\left(\mathbf{s}, \mathscr{C}'\right) \quad \forall i \in \{1, \ldots, k\} \;\; \wedge \\
& \mathbf{t}_i = \mathbf{v}_i \quad \forall i \in \{1, \ldots, k\}
\end{aligned}
\tag{12}
$$

where $\mathbf{t}_i$ is the dual variable. Due to the frequency domain formulation, this linear inverse problem can finally be efficiently solved with the Alternating Direction Method of Multipliers [25]. For the sake of brevity, we omit a derivation of the subproblems. However, we note that unlike previous work (e.g., by Bristow et al. [26]), we do not solve a simple fitting problem, but we include the measurement matrix $\mathbf{C}$ in our reconstruction problem. This leads to an inherently different algorithm than in previous approaches.

3.5. Synthetic evaluation of reconstruction code

To generate ground-truth pixel profiles, we sampled a high resolution "ground truth" transient image measured by Velten et al. using the direct sampling method [17]. The observations are then generated by assuming a typical sinusoidal measurement matrix $\mathbf{C}$ where f, g are defined as in Eq. 1. We sample ω evenly spaced in 100 steps from 10 MHz to 120 MHz, which is a realistic range for recent correlation image camera systems [9] and ϕ as $0, \pi/2$, giving exactly $N = 100 \cdot 2$ measurements $\mathbf{b}$ per pixel. The measurements are normalized and then corrupted with 1% Gaussian noise.

Figure 6 shows synthetic results for two different pixel profiles. We compare our method ("Reconstruction Gauss-Exp") to the ground-truth signal ("Original") as well as two recent methods from transient imaging literature: Lin et al.'s smooth frequency-domain interpolation ("FFT model") [27], the non-linear non-convex model-fit by Heide et al. ("Sequential model") [9]. For the sake of completeness, we further add comparisons to various state-of-the-art sparse reconstruction techniques, namely LASSO ("Sparse reconstruction LASSO") [28], OMP ("Sparse reconstruction OMP") [29] as well as Generalized Approximate Message Passing [30] using Donoho/Maleki/Montanari-style thresholding ("GAMP-DMM") and assuming a Laplacian signal in an MAP formulation ("GAMP-MAP").

One can see that, although these pixels are dominated by direct reflections, the signals are not sparse at all, and time-domain sparse backscatter models as used by Freedman et al.'s SRA [10], or Bhandari et al. [15] are not capable of close reconstructions. Our method, on the other hand, produces solutions that follow the ground-truth distributions more closely than any of the competing models. In particular, the exponentially modified Gaussian basis outperforms recent ap-
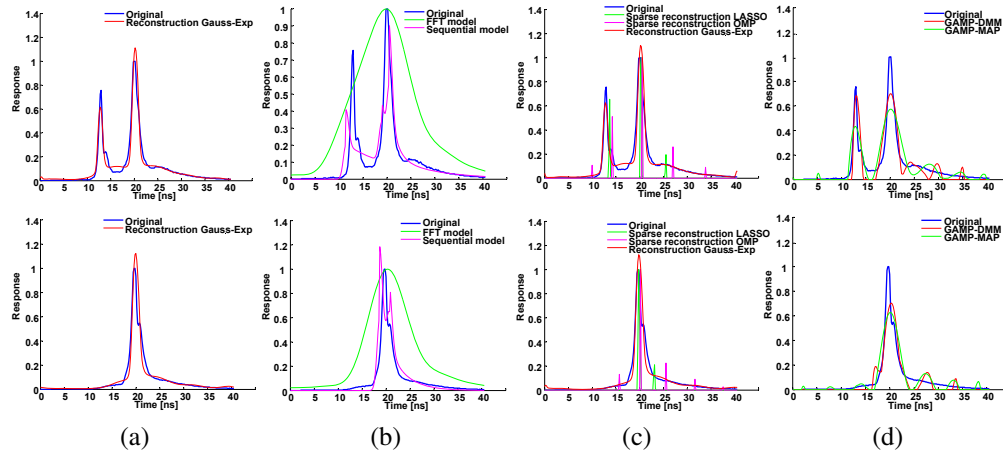
Fig. 6. Example showing the effect of our sparse coding optimization on two pixels from the "Tomato" dataset. From left to right: (a) Proposed new model, (b) FFT and sequential models, (c) state-of-the-art sparse reconstruction (LASSO and OMP), (d) two state-of-the-art compressed sensing models (GAMP).

proaches that solve non-linear non-convex optimization [9]. Out of all the methods tested, the one delivering the poorest fit is a recent approach [27], which imposes the weakest prior by just interpolating smoothly in the Fourier domain.

4. Analysis and evaluation

In this section we show results for imaging in scattering media using our reconstruction method proposed in the previous section. The measurement setup is explained first. After that the results are analysed qualitatively and quantitatively.
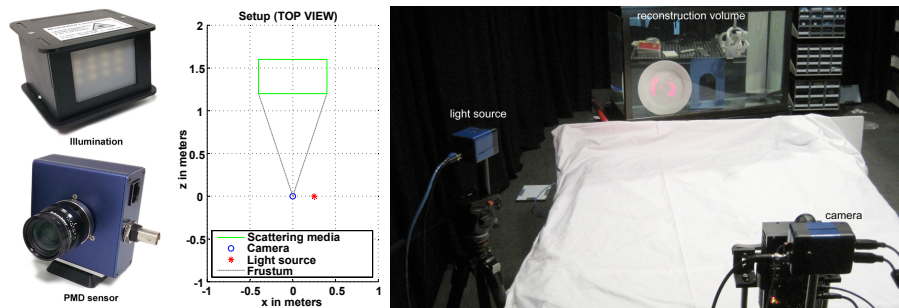

Fig. 7. Camera prototype and setup: We use a modified PMDTechnologies CamBoard nano using an array of red laser diodes for illumination (left) as described by Heide et al. [31]. In our setup we image a tank filled with a scattering medium of different concentrations frontal with the cameras. The spatial dimensions and arrangement of the setup is shown in the center. The setup is shown on the right.

4.1. Imager

Camera We use a correlation image ToF camera prototype for our experiments using red laser light as illumination, see Fig. 7 on the left.

The ToF camera consists of a modulated light source and a correlation image detector, as described by Heide et al. [31]. The sensor is a PMD CamBoard nano, modified to allow for external control of the modulation signal (since access to its FPGA configuration was not available). The light source consists of an array of 650 nm, 250mW laser diodes which are operated in pulse mode, also described in the same source.

We measure ω evenly spaced in 3 steps from 20 MHz to 60 MHz and ϕ evenly spaced between 0 and 5 m$/c\omega$ in 201 steps and with an additional shift of $(0, \pi/2)$, resulting in a measurement vector $\mathbf{b}$ with exactly $N = 201 \cdot 2 \cdot 3$ samples per pixel.

Calibration In agreement with Heide et al. [9], we calibrated the matrix $\mathbf{C}$ by measuring a diffuse planar target that is mounted on a translation stage perpendicular to the z-axis. The target is translated along this axis at positions according to different travel times $\tau_i, i \in \{1 \ldots M\}$. We can then populate $\mathbf{C}$ column by column for each τ_i.

4.2. Setup

Measurement setup An image of our measurement setup is shown in Fig. 7 on the right. We placed a water tank with glass hull at a distance of 1.2 m in front of our camera, so that the optical axis intersects the center of the largest planar side. The light source is placed at a slight offset to the right to eliminate direct reflections of the air-to-glass interface on the camera-facing wall. See the schematic in Fig. 7 in the center for the exact spatial alignment.

Scattering media We filled the tank with 80 liters of water and submerged objects at different positions in the water (and in particular at different distances to the camera). We then evaluated our approach on two different scattering materials with a series of different concentrations. First, we conducted a sequence of 100 experiments using homogenized milk from 0 ml to 500 ml in 5 ml steps. Second, we conducted a sequence of 50 experiments using Gypsum plaster with continuum of particle sizes ≤0.125 mm. We used 0 oz to 50 oz in 1 oz steps.

For each of the 150 experiments we take a full measurement with the measurement parameters as discussed in Section 4.1.

4.3. Qualitative results

Figures 8 and 9 show qualitative results for imaging through scattering media of increasing density. Objects are immersed in a tank filled with water, to which increasing concentrations of milk (Fig. 8) or plaster (Fig. 9) are added. With increasing concentration, visibility through the scattering medium quickly drops off for a conventional camera (left column). On the other hand, a light transport analysis based on correlation image sensors, not only increases the ability to detect objects in highly turbid solutions, but also allows for a simultaneous estimation of distance (color coded images in the right column of each figure).

4.4. Quantitative results

To quantitatively analyze our results, we measure the error of the depth estimate for three distinct camera pixels with respect to the measured ground truth depths. The three camera pixels 'plate', 'holder' and 'bar' are chosen as representatives for objects located at different depths in the reconstruction volume. Their spatial positions are shown in Fig. 10 (left). All objects have approximately diffuse reflectance except for the 'plate' which does have a specular component.

Ground truth depth measurements for all pixels were acquired manually. The pixel 'plate' is located at 5.5 cm behind front facing glass wall of the tank, pixel 'holder' at 21.5 cm into the tank, and 'bar' at 40.0 cm, touching the rear-facing wall of the tank.
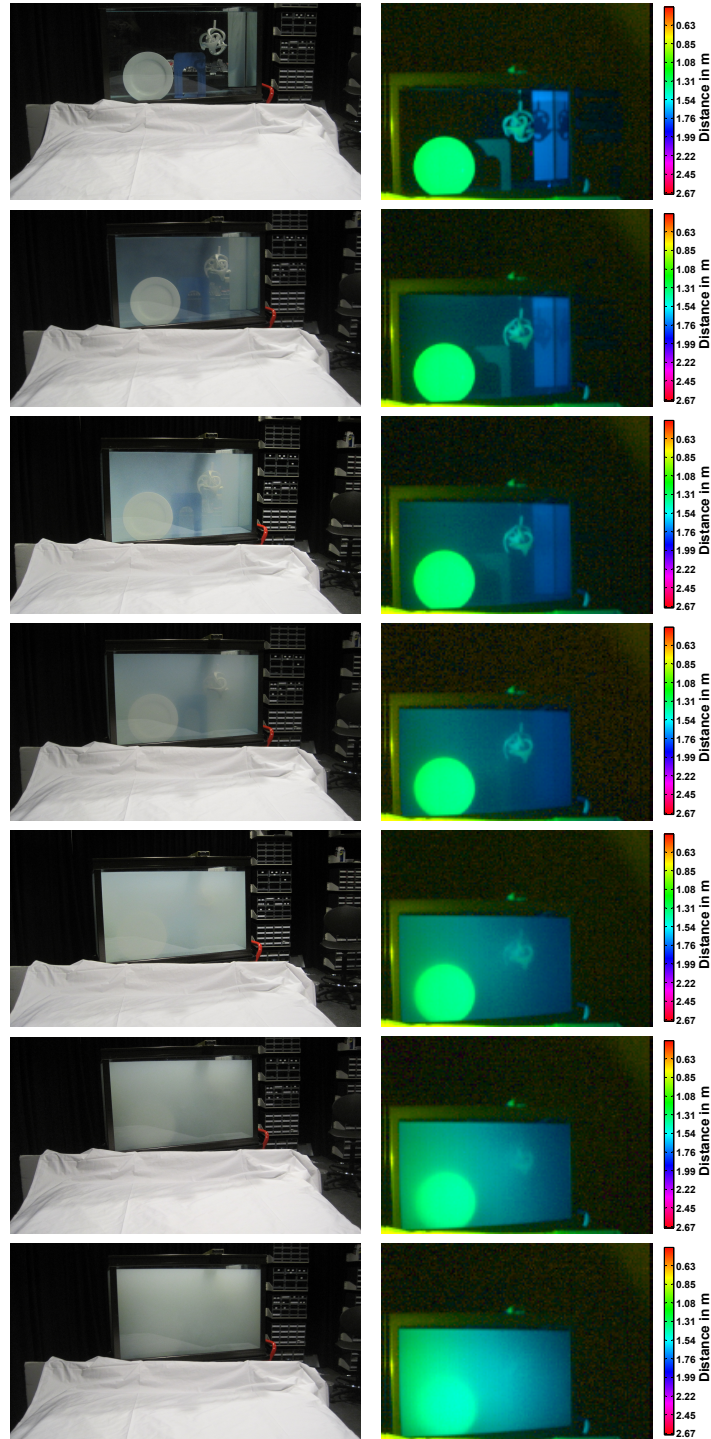
Fig. 8. Qualitative results for the milk experiment described in Section 4.2. Experiments in each row from top down: 0 ml, 10 ml, 20 ml, 40 ml, 80 ml, 160 ml, 300 ml of milk in water. Each row shows the regular camera image (left), peak image for red camera reconstruction (right). Peak images are here a parabola fit through the two nearest neighbor points of the strongest peak, where the position is encoded as as hue and the intensity is encoded as value in the HSV color model.
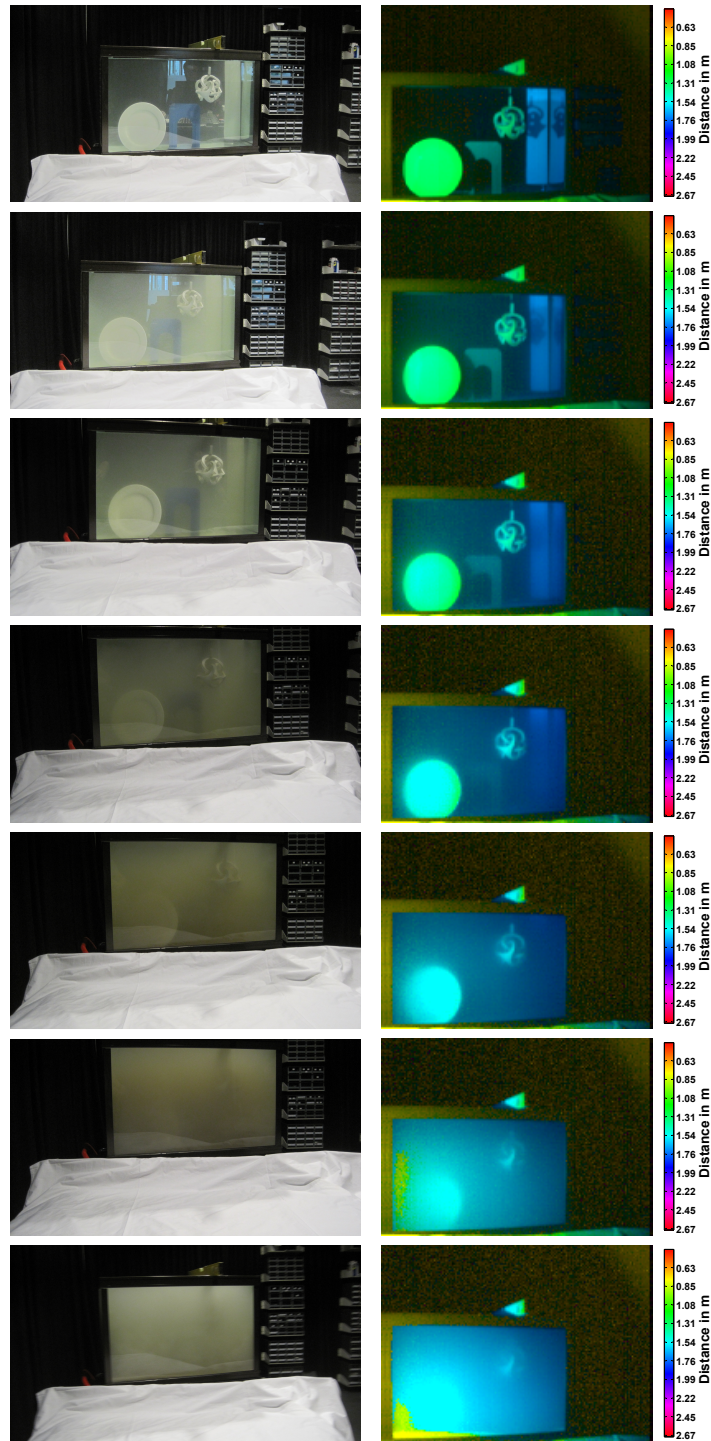
Fig. 9. Qualitative results for the plaster experiment described in Section 4.2. Experiments in each row from top down: 0 oz, 2 oz, 4 oz, 8 oz, 16 oz, 32 oz, 59 oz of plaster in water. Each row shows the regular camera image (left), peak image for red camera reconstruction (right). Peak images are here a parabola fit through the two nearest neighbor points of the strongest peak, where the position is encoded as as hue and the intensity is encoded as value in the HSV color model.
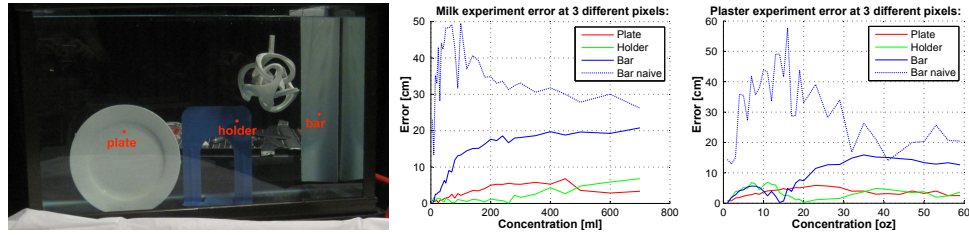
Fig. 10. Error for three specific pixels shown on the left. The error of the position is shown for all 100 experiments using milk as scattering medium and all 50 experiments using plaster as scattering medium. The error is measured with respect to 0% concentration and corrected for the speed of light in water. Standard time-of-flight depth reconstruction (arctan solution) breaks down even at very low concentrations of scattering agent.

The center and right portions of Fig. 10 show the error with respect to the ground truth pixel depth for *all 100 experiments with milk as scattering media and all 50 experiments with plaster as scattering media* as described in Section 4.2. We can see that the error of pixels 'plate' and 'holder' has a fairly low slope and remains almost flat around 1 cm–5 cm even for strong concentrations. These pixel depths are located close to the front and in the middle of the reconstruction volume, so the scattering is reduced in comparison to pixel 'bar' which is at the very rear of the reconstruction volume. Its error is significantly larger; around 10 cm–20 cm due to the increased scattering. However, performing a naïve reconstruction as described in Section 2.1 on the same pixel resulted in even significantly larger errors around 30 cm–60 cm for both the plaster and the milk sequence of experiments.

5. Discussion

We demonstrated that correlation image sensors can be used for imaging in scattering and turbid media. Unlike alternative methods such as fast gated imaging, this approach is robust under ambient illumination (all measurements were taken with room lighting switched on), and works with higher light levels than approaches based on single light pulses.

The key to using correlation image sensors for this problem lies in the realization that transient light transport is sparse, but only in a well-chosen representation. We develop here a model based on exponentially modified Gaussians, which is tailored towards representing combinations of surface reflection and volumetric scattering. Fast convolutional coding is employed to solve the resulting optimization problem.

The depth resolution of the current cameras is limited by the modulation frequency of the correlation image sensor, currently up to 100 MHz. Under ideal circumstances, a 10-bit correlation imager can distinguish 1000 distinct phases within the 3 m wavelength, corresponding to a resolution of ca. 3 mm. These numbers, as well as our results should scale up well with future increases in modulation frequency.

Acknowledgments