Compressive Structured Light for Recovering Inhomogeneous Participating Media

Jinwei Gu, *Member, IEEE,* Shree K. Nayar, *Member, IEEE,*
Eitan Grinspun, Peter N. Belhumeur, Ravi Ramamoorthi, *Member, IEEE*

Abstract—We propose a new method named compressive structured light for recovering inhomogeneous participating media. Whereas conventional structured light methods emit coded light patterns onto the surface of an opaque object to establish correspondence for triangulation, compressive structured light projects patterns into a *volume* of participating medium to produce images which are *integral measurements* of the volume density along the line of sight. For a typical participating medium encountered in the real world, the integral nature of the acquired images enables the use of *compressive sensing* techniques that can recover the entire volume density from only a few measurements. This makes the acquisition process more efficient and enables reconstruction of dynamic volumetric phenomena. Moreover, our method requires the projection of multiplexed coded illumination, which has the added advantage of increasing the signal-to-noise ratio of the acquisition. Finally, we propose an iterative algorithm to correct for the attenuation of the participating medium during the reconstruction process. We show the effectiveness of our method with simulations as well as experiments on the volumetric recovery of multiple translucent layers, 3D point clouds etched in glass, and the dynamic process of milk drops dissolving in water.

Index Terms—compressive sensing, volume reconstruction, structured light, time-varying

✦

1 INTRODUCTION

Structured light has a long history in the computer vision community [Salvi et al., 2004]. It has matured into a robust and efficient method for recovering the surfaces of objects. By projecting coded light patterns on the scene, and observing it using a camera, correspondences are established and the 3D structure of the scene is recovered by triangulation. Over the years, researchers have developed various types of coding strategies, such as binary codes, phase shifting, spatial neighborhood coding, *etc.*. All structured light range finding approaches are based on a common assumption: Each point in the camera image receives light reflected from a single *surface* point in the scene.

However, many real-world phenomena can only be described by volume densities rather than boundary surfaces. Such phenomena are often referred to as *participating media*. Examples include translucent objects, smoke, clouds, mixing fluids, and biological tissues. Consider an image acquired by photographing a volume of a participating medium. Unlike in the case of an opaque object, here each pixel receives scattered light from *all points along the line of sight* within the volume. Narasimhan et al. [2005] have addressed the problem of recovering opaque surfaces immersed in

a participating medium (rather than a clear medium) using structured light range finding. The goal in this case is to make existing structured light range finding methods robust to light scattering by the medium, rather than recover the medium itself.

The problem of recovering the volume density of a participating medium (in particular, smoke), was addressed by Hawkins et al. [2005]. They used a high-powered laser sheet and a high-speed camera (5000fps) to measure thin slices of a smoke density field via *scanning*, which is similar to a technique termed laser-induced fluorescence (LIF) in the fluid imaging community [Deusch and Dracos, 2001]. Fuchs et al. [2007] proposed the idea of shooting a set of static laser rays into the volume and using spatial *interpolation* to reconstruct the volume. However, the measurements are inherently sparse in this case and hence the recovered information is low in resolution.

In this paper, we show that by using coded light patterns, one can make the measurement of a participating medium highly efficient in terms of acquisition time as well as illumination power. In particular, we exploit the fact that the brightness measurements made at image pixels correspond to true line-integrals through the medium (see Fig. 1a), and then solve for its volumetric density. We consider both spatially- and temporally-coded light patterns. Because the patterns are predetermined, measurement and reconstruction time are decoupled. We target low-density inhomogeneous media, for which the density function is *sparse* in an appropriately-chosen basis; this allows us to harness compressive sensing techniques [Candes et al., 2006; Donoho, 2006] that accurately reconstruct

- *J. Gu is with the Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, 14623. E-mail: jwgu@cis.rit.edu.*
- *S. K. Nayar, E. Grinspun, and P. N. Belhumeur are with Columbia University, New York, NY 10027. E-mail: {nayar,eitan,belhumeur}@cs.columbia.edu.*
- *R. Ramamoorthi is with University of California at Berkeley, Berkeley, CA 94720. E-mail: ravir@cs.berkeley.edu.*
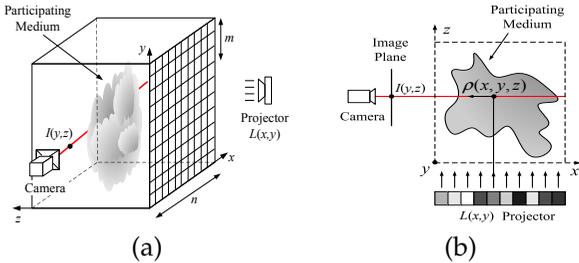
Fig. 1. (a) Compressive structured light for recovering inhomogeneous participating media. Coded light is emitted along the z-axis to the volume while the camera acquires images as line-integrated measurements of the volume density along the x-axis. The light is coded in either the spatial domain or temporal domain with a predetermined (non-adaptive) sequence. We reconstruct the volume density from the acquired measurements by using *compressive sensing* techniques. (b) Image formation model for participating medium under single scattering. The image irradiance at one pixel, $I(y, z)$, depends on the integral along the x-axis of the projector's radiance, $L(x, y)$, and the medium density, $\rho(x, y, z)$, along a ray through the camera center; refer to (3)–(4).

a signal from only a few measurements. We refer to our approach as *compressive structured light*.

We show that compressive structured light is more economical than a straightforward sequential scanning of a volumn. Whereas the sampling rate of the latter is limited by the desired spatial resolution, the sampling rate of the former is restricted by the sparsity of the data–a considerably more relaxed constraint for low-density phenomena. Since our approach requires fewer measurements, it naturally enables the recovery of dynamic participating media. An added advantage of compressive structured light is that it requires the projection of multiplexed coded illumination which results in measurements with higher signal-to-noise ratio [Schechner et al., 2007]. An important practical consequence is that light sources of significantly lower power than in the case of sequential scanning can be used.

We have implemented our approach using a digital projector and a camera as shown in Fig. 9. The projector and the camera are synchronized and both operate at 360fps. Using 24 coded light patterns, we are able to recover a 128^3 volume at 15fps. Using this system, we have recovered various types of inhomogeneous participating media, including multiple translucent layers, a 3D point cloud of a face etched in a glass cube, and the dynamic process of milk dilution.

2 RELATED WORK

Compressive Sensing Compressive sensing [Candes et al., 2006; Donoho, 2006] is a nascent field of applied mathematics with a variety of successful applications including imaging [Takhar et al., 2006; Willett et al., 2007], medical visualization [Lustig et al., 2007], and

face recognition [Wright et al., 2009]. Recently, compressive sensing has also been widely used to solve many computer vision and computer graphics problems, such as high speed imaging [Veeraraghavan et al., 2011; Sankaranarayanan et al., 2010; Hitomi et al., 2011], image restoration and denoising [Mairal et al., 2009; Elad and Aharon, 2006; Protter and Elad, 2009], and light transport measurement [Peers et al., 2009; Sen and Darabi, 2009]. It offers a theoretical framework to reconstruct "sparse" signals from far fewer samples than required by the conventional Shannon sampling theorem. Our work builds on the basic formulation of compressive sensing, which we augment with auxiliary terms specific to the reconstruction of volume density.

Reconstruction of Volumetric Phenomena There are several recent works in reconstruction of volumetric phenomena from multiple views. Hasinoff and Kutulakos [2007] used two views to reconstruct flames by assuming flames are surfaces in the volume. Based on tomographic algorithms, Ihrke and Magnor [2004, 2006] and Trifonov et al. [2006] used eight views and 72–360 views, respectively, for recovering flames (and smoke) as well as transparent objects. We mentioned light-based methods [Hawkins et al., 2005; Deusch and Dracos, 2001; Fuchs et al., 2007] earlier in the previous section. For a comprehensive survey of works in this area, see Ihrke et al. [2008].

Multiplexed Illumination Our work is also related to multiplexed illumination [Schechner et al., 2007] in that both use coded light as illumination. However, there is a fundamental difference: Whereas the conventional multiplexing aims at increasing signal-to-noise ratio of the measurements, our work aims at increasing the efficiency of the acquisition process, *i.e.*, to reconstruct high dimensional signals from a few measurements. In summary, both the coding strategies and the reconstruction algorithms are different.

This paper builds upon and extends our previous work [Gu et al., 2008].

3 COMPRESSIVE SENSING: BACKGROUND

We give a brief introduction on compressive sensing. In its simplest form, compressive sensing seeks a solution of the underdetermined linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{x} \in \mathbb{R}^n$ is a sparse signal, $\mathbf{A}$ is an $m \times n$ matrix called the "measurement ensemble", and $\mathbf{b}$ is the vector of m measurements, with $m < n$.

Compressive sensing theory asserts that one can recover the signal from far fewer measurements than the dimension of the signal, if the signal is *sparse*—it is represented with few non-zero coefficients in a suitable basis—and the measurements are uncorrelated, in the sense that each measurement is an inner product of the signal with a *test function* that has a necessarily *dense* representation in the chosen basis. Equivalently, the measurement ensemble $\mathbf{A}$ needs to satisfy the *restrictive isometry condition (RIC)*. Strict definitions of sparsity and RIC can be found in Candes

et al. [2006]. Given a measurement ensemble matrix $\mathbf{A}$, compressive sampling theory predicts that $\mathbf{x}$ is the minimizer of $||\mathbf{x}||_1$, subject to $\mathbf{Ax} = \mathbf{b}$.

The above reconstruction strategy has been shown to work well for sparse signal estimation, even from a noisy measurement [Candes et al., 2006]. In our work, we augment the basic problem above with auxiliary terms that enforce the nonnegative constraint for the signal, and that exploit the sparsity not only of the signal value but also its gradient. Indeed, our evaluation (§7, §8) indicates that using the sparsity of the gradient is important for accurate and efficient reconstruction.

4 IMAGE FORMATION MODEL

Let us first derive the relationship between the volume density $\rho(x, y, z)$ and the image irradiance $I(y, z)$ of the camera under our camera/projector setting. We focus on non-emissive, diluted participating media, in which multiple scattering is assumed to be negligible.

As shown in Figure 1b, each camera pixel receives light scattered from a row of voxels along the line of sight in the volume (*i.e.*, the red line in Figure 1b). For simplicity, we assume the camera and the projector are placed sufficiently far from the working volume, and thus they are orthographic projection. The distortion caused by perspective projection can be corrected with a calibration step, if needed.

Consider one voxel $\rho(x, y, z)$ in the row. Light emitted from the projector, $L(x, y)$, is first *attenuated* as it travels from the projector to the voxel, *scattered* at the voxel, and then *attenuated* as it travels from the voxel to the camera. Assuming single scattering, the radiance sensed by the camera from this particular voxel is [Ishimaru, 1978]

$$L(x, y) \cdot \exp(-\tau_1) \cdot \sigma_s \cdot \rho(x, y, z) \cdot p(\theta) \cdot \exp(-\tau_2), \quad (1)$$
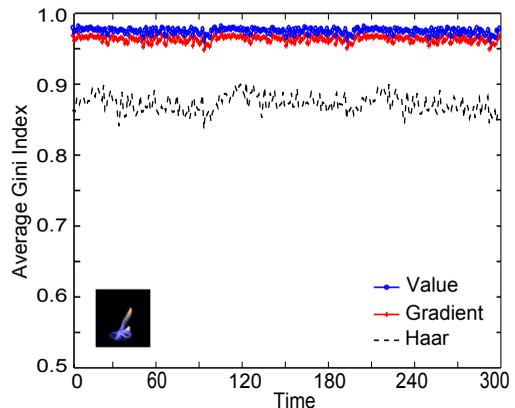
where $\rho(x, y, z)$ is the volume density (*i.e.*, density of particles) at the voxel, $p(\theta)$ is the phase function ($\theta = \pi/2$ since the camera and the projector are perpendicularly placed), and τ_1 and τ_2 are the optical thicknesses from the projector to the voxel and from the voxel to the camera; σ_s is the scattering cross section of the participating medium. Since σ_s and $p(\theta = \pi/2)$ are the same for all voxels, the above formula can be simplified to (up to a scale $\sigma_s \cdot p(\theta = \pi/2)$)

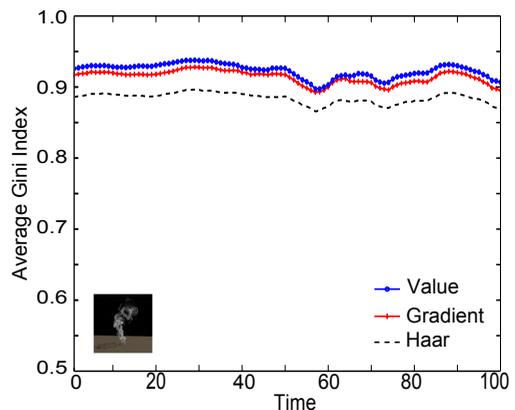$$L(x, y) \cdot \exp\left(-(\tau_1 + \tau_2)\right) \cdot \rho(x, y, z). \quad (2)$$

The image irradiance, $I(y, z)$, which is the integral of the scattered light from all the voxels along the line, is therefore

$$I(y, z) = \int_x L(x, y) \cdot \exp\left(-(\tau_1 + \tau_2)\right) \cdot \rho(x, y, z) dx. \quad (3)$$

For highly diluted media (*i.e.*, $\rho \to 0$), because the optical thicknesses τ_1 and τ_2, which are proportional to the density ρ, are close to 0, the attenuation term usually can also be ignored (*i.e.*, $\exp\left(-(\tau_1 + \tau_2)\right) \approx 1$)for the recovery of volume densities [Hawkins et al.,

(a)

(b)

Fig. 2. Sparsity (*i.e.*, the Gini index) of two measured dynamic participating media: (a) flame [Ihrke and Magnor, 2004]; (b) smoke [Hawkins et al., 2005]. For each volume, we compute the Gini index for each row of the volume and average them as a sparsity measure of the volume. Three types of bases are tested: the value of the volume density itself, the gradient of the volume density, and the Haar wavelet transform of the volume density. As shown, these two dynamic phenomena have consistently large sparsity over time, especially in their values and gradients.

2005; Fuchs et al., 2007]. In this case, Equation (3) is reduced to a linear projection of the illumination and the volume density,

$$I(y, z) \approx \int_x \rho(x, y, z) \cdot L(x, y) \, dx. \quad (4)$$

For denser media, we have to consider the attenuation term. We present a iterative method to correct for the attenuation in Section 6.3.

5 SPARSITY OF PARTICIPATING MEDIA

As mentioned in Section 1, the sparsity in the volume densities of diluted participating media (*e.g.* smoke, mixing liquids) has been employed for efficient acquisition. Qualitatively, the assumption of sparsity is reasonable since often only a small portion of the entire volume has non-zero density.
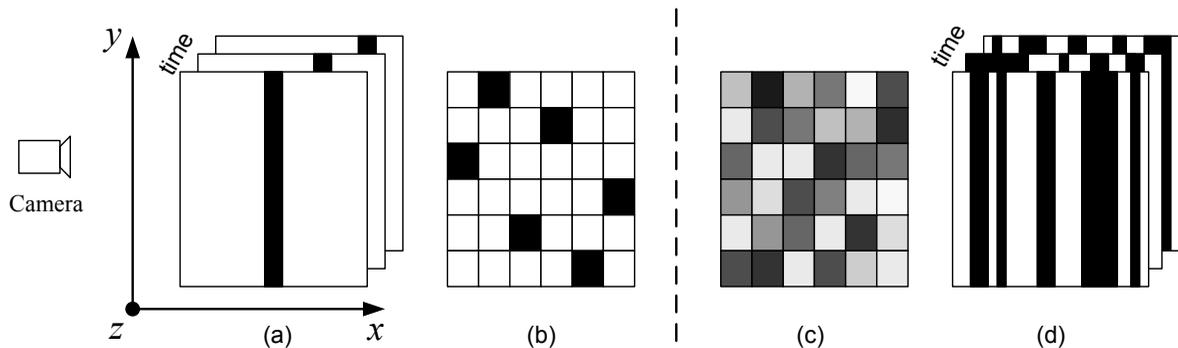
Fig. 3. Coding strategies of the light $L(x, y)$ at time t for recovering dynamic volumes: (a) scan (one stripe turned on) [Deusch and Dracos, 2001; Hawkins et al., 2005]; (b) laser-lines interpolation (one pixel turned on per one row) [Fuchs et al., 2007]; (c) Spatial coding of compressive structured light (all pixels are turned on with random values per time frame); (d) Temporal coding of compressive structured light (random binary stripes are turned on per time frame). Compressive structured light, shown in (c) and (d), recovers the volume by reconstructing the 1D signal along x-axis from a few integral measurements. It utilizes the light more efficiently and thus makes the measurement process highly efficient both in acquisition time and illumination power.

To quantitatively justify the use of the sparsity for reconstruction, in this section, we compute the sparsity of two sequences of time-varying participating media measured by other researchers. One sequence is flame measured by Ihrke and Magnor [2004] and the other sequence is smoke measured by Hawkins et al. [2005]. Although these are not the volumetric phenomena we recovered in our experiments, they have similar characteristics in the time-varying volume densities, and thus can give us some estimate about the sparsity of our subjects. Moreover, we compute the sparsity of the same signals at different bases — this will also give us insights of which space is the best for sparse reconstruction.

The key is how to compute sparsity for a given signal. While the definition of sparsity is the number of non-zero elements (*i.e.*, ℓ-0 norm), this is not suitable for signals with noise. While ℓ-1 norm has been used for sparse reconstruction, it is not normalized and thus not good for comparing the sparsity of two signals or two sets of bases. A thorough discussion about the measures for sparsity can be found in [Hurley and Rickard, 2009], in which they suggested Gini index as a good measure for sparsity. Gini index was originally proposed in economics as a measure of the inequality of wealth [Gini, 1921]. It is normalized (its value is between 0 and 1), and it is robust to noise. Given a vector, $\mathbf{x} = [x_1, x_2, x_3, \cdots, x_N]$, we take its absolute value and sort the elements from smallest to largest, $|x|_{(1)} \leq |x|_{(2)} \leq \cdots \leq |x|_{(N)}$. The Gini index of the signal $\mathbf{x}$ is defined as:

$$G(\mathbf{x}) = 1 - \frac{2 \sum_{k=1}^{N} |x|_{(k)} \cdot (N - k + 0.5)}{N \sum_{k=1}^{N} |x|_{(k)}}. \quad (5)$$

The higher $G(\mathbf{x})$ is, the more sparse the signal is. More properties of Gini index measure can be found in [Hurley and Rickard, 2009].

We compute the Gini indices for all the rows of each volume and average them as the sparsity of

the volume. Figure 2 shows the sparsity for the two data sets (Figure 2(a) for the flame sequence [Ihrke and Magnor, 2004] and Figure 2(b) for the smoke sequence [Hawkins et al., 2005]). Three types of bases are used to compute the Gini index — the value of the volume density itself, the gradient of the volume density, and the Haar wavelet transform of the volume density. As shown in Figure 2, we found that: (1) over the entire sequence, the two dynamic participating media has consistently high sparsity (≥ 0.95). (2) Compared with the Haar wavelet, the volume densities have higher sparsity in their value and gradients. These observations help us design the acquisition system.detailed below.

6 COMPRESSIVE STRUCTURED LIGHT

Unlike the conventional structured light methods for surface recovery where each camera pixel receives light reflected from one point, for participating media, each camera pixel receives light from all points along the line of sight within the volume. Thus each camera pixel is an integral measurement of one row of the volume density. Whereas conventional structured light range finding methods seek to triangulate the position of a single point, compressed structured light seeks to reconstruct the 1D density "signal" from a few measured integrals of this signal.

This is clearly a more difficult problem. One way to avoid this problem is to break the integrals into pieces which can be measured directly. The price, however, is the deterioration of either spatial resolution or temporal resolution of the acquisition. Existing methods either illuminate a single slice at a time and scan the volume (see Figure 3(a) and [Deusch and Dracos, 2001; Hawkins et al., 2005]), thus sacrificing temporal resolution, or they illuminate a single pixel per row and use interpolation to reconstruct the volume (*e.g.*, Figure 3(b) and [Fuchs et al., 2007]), sacrificing spatial resolution.

TABLE 1
Objective functionals used for volume reconstruction

Method	Optimization Functional	Constraints
Least Square (LS)	$\|\mathbf{Ax} - \mathbf{b}\|_2$	
Nonnegative LS	$\|\mathbf{Ax} - \mathbf{b}\|_2$	$\mathbf{x} \geq 0$
CS-Value	$\|\mathbf{x}\|_1$	$\mathbf{Ax} = \mathbf{b},\ \mathbf{x} \geq 0$
CS-Gradient	$\|\mathbf{x}'\|_1$	$\mathbf{Ax} = \mathbf{b},\ \mathbf{x} \geq 0$
CS-Both	$\|\mathbf{x}\|_1 + \lambda\|\mathbf{x}'\|_1$	$\mathbf{Ax} = \mathbf{b},\ \mathbf{x} \geq 0$

In contrast, the proposed compressive structured light method uses the light much more efficiently, projecting coded light patterns that yield "signatures," or integral measurements, of the unknown volume density function.

The didactic illustration in Figure 1(a) depicts a simple lighting/viewpoint geometry under orthographic projection, with the camera viewpoint along the x-axis, and the projector emitting along the z-axis. Consider various coding strategies of the 3D light function $L(x, y, t)$: *Spatial* codes (Figure 3(c)) recover the volume from a single image by trading spatial resolution along one dimension; *Temporal* codes (Figure 3d) trade temporal resolution by emitting a sequence of vertical binary stripes (with no coding along y-axis), so that full spatial resolution is retained. All of the four methods shown in Figure 3 can be equally improved using color channels.

In the following, we will see that these compressive structured light codes yield high efficiency both in acquisition time and illumination power; this comes at the cost of a more sophisticated reconstruction process, to which we now turn our attention.

6.1 Coding and Formulation

To better visualize our formulation, consider first the case of spatial coding. Suppose we want to reconstruct a volume at the resolution $n \times n \times n$ (*e.g.*, $n = 100$). The camera and the projector have the resolution of $M \times M$ pixels (*e.g.*, $M = 1024$). Therefore, one row of voxels along the x-axis (refer to the red line in Figure 1a) will receive light from $m = M/n$ (*e.g.*, $m = 1024/100 = 10$) rows of the projector's pixels. The light scattered by these voxels in the viewing direction will then be measured, at each z-coordinate, by a vertical column of m camera pixels. Without loss of generality, we use $\mathbf{l}_1 = L(x, 1), \cdots, \mathbf{l}_m = L(x, m)$ to denote the m rows of pixels from the projector, and $b_1 = I(1, z), \cdots, b_m = I(m, z)$ to denote the image irradiance of the m pixels in the camera image. Let $\mathbf{x} = [\rho_1, \cdots, \rho_n]^T$ be the vector of the voxel densities along the row. Assuming no attenuation for now, the image irradiance for each of these m pixels is a linear projection of the light and the voxels' density from Equation (4):

$$b_i = \mathbf{l}_i^T \mathbf{x}, \qquad i = 1, \cdots, m. \qquad (6)$$

Rewriting these m equations in matrix form, we have:

$$\mathbf{Ax} = \mathbf{b}, \qquad (7)$$

where $\mathbf{A} = [\mathbf{l}_1, \cdots, \mathbf{l}_m]^T$ is a $m \times n$ matrix, $\mathbf{b} = [b_1, \cdots, b_m]^T$ is a $m \times 1$ vector.

Thus, if attenuation is not considered, the problem of recovering the volume is formulated as the problem of reconstructing the 1D signal $\mathbf{x}$ given the constraints $\mathbf{Ax} = \mathbf{b}$. To retain high spatial and temporal resolution, we often can only afford far fewer measurements than the number of unknowns, *i.e.*, $m < n$, which

means the above equation is an underdetermined linear system and optimization is required to solve for the best $\mathbf{x}$ according to certain priors.

One benefit of this optimization-based reconstruction is high efficiency in acquisition, which we quantify using the *measurement cost*, m/n, where m is the number of the measurements and n is the number of unknowns (*i.e.*, the dimension of the signal). For example, the measurement cost of the scanning method [Deusch and Dracos, 2001; Hawkins et al., 2005] is one. We show that by exploiting the sparsity of the signal, we can reconstruct the volume with much lower measurement cost (about $1/8$ to $1/4$).

6.2 Reconstruction via Compressive Sensing

Solving the underdetermined linear system requires some prior (assumed) knowledge of the unknown signal, which can be represented as optimization functionals or constraints on the data. We consider several alternatives, as listed in Table 1. In addition to the commonly-used Least Square (LS) and Nonnegative Least Square (NLS) approaches, we also consider functionals using ℓ_1-norms, as these bias toward sparse representation, based on our observations on the sparsity of some dynamic participating media (Section 5).

First, we observe that for many natural volumetric phenomena, often only a small portion of the entire volume is occupied by the participating media. For example, consider the beautiful ribbon patterns generated by smoke; similarly, sparsity was implicitly used to reconstruct (surface-like) flames [Hasinoff and Kutulakos, 2007]). This suggests the use of the ℓ_1-norm of the signal value (CS-Value).

Furthermore, from Figure 2 we also observe high sparsity in the gradients of the volume densities. The sparsity of *gradients* of natural images is well studied [Olshausen and Field, 1996; Simoncelli, 1997] and shown to work well for image restoration [Rudin et al., 1992]. In this vein, we consider the use of ℓ_1-norm on the signal's gradient (CS-Gradient).

Finally, consider a dynamic process, such as milk dissolving in water: here diffusion decreases the signal value's sparsity over time, but it increases the gradient sparsity. Motivated by this observation, we consider the linear combination of ℓ_1-norms of both the value and the gradient (CS-Both). The objective function is thus defined as $|\mathbf{x}|_1 + \lambda|\mathbf{x}'|_1$, where λ is the weight. In order to obtain a one-to-one mapping
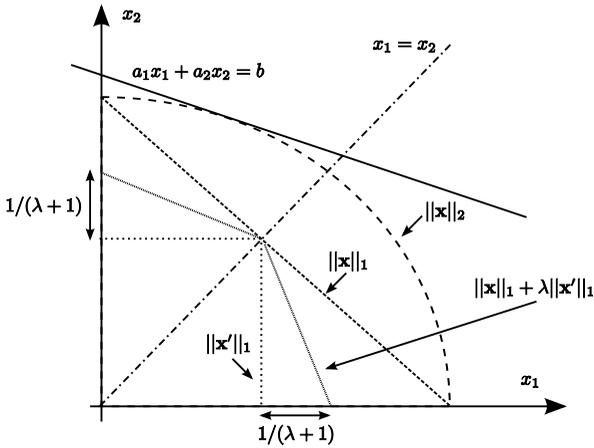
Fig. 4. Geometric interpretation of the objective functionals for a 2×1 vector $\mathbf{x} = [x_1, x_2]$. Please refer to text for details.

from the gradient domain to the original signal, the gradient $\mathbf{x}'$ is defined as a $(n+1) \times 1$ vector consisting of both the differentials and the first and the last element of $\mathbf{x}$, *i.e.*, $\mathbf{x}' = [\rho_1, \rho_2 - \rho_1, \rho_3 - \rho_2, \cdots, \rho_n - \rho_{n-1}, \rho_n]$, where $\mathbf{x} = [\rho_1, \cdots, \rho_n]$. Please note that λ is a dimensionless quantity because $\mathbf{x}'$ is represented as a vector of the finite difference of $\mathbf{x}$.

To understand these objective functionals intuitively, let us consider a simple example where $\mathbf{x}$ is a 2×1 vector, $\mathbf{x} = [x_1, x_2]$, and assume $x_1 \geq 0$, $x_2 \geq 0$. As shown in Figure 4, suppose we only have one measurement $a_1 x_1 + a_2 x_2 = b$ which tells us the solution of $\mathbf{x}$ should be on a line. The least square solution minimizes $\|\mathbf{x}\|_2$ and it will be the tangential point between the line $a_1 x_1 + a_2 x_2 = b$ and the circle, which is less likely to be sparse. The CS-Value solution minimizes $\|\mathbf{x}\|_1$, and it will be the intersection point between the line $a_1 x_1 + a_2 x_2 = b$ and the x_2 axis, which is sparse (*i.e.*, one element is zero). The CS-Gradient solution minimizes $\|\mathbf{x}'\|_1$, and in this case it will be the intersection point between the line $a_1 x_1 + a_2 x_2 = b$ and the diagonal line $x_1 = x_2$, which is sparse in the gradient. Finally, the CS-Both solution finds a trade-off between CS-Both and CS-Gradient. It allows sparse solutions both on the two axes and the diagonal line. Its exact shape depends on the weight λ, as shown in Figure 4. The optimal weight λ should be set to the ratio between the sparsity in the signal's gradient and the sparsity in the signal's value.

To find the optimal λ, we run simulation on the smoke sequence [Hawkins et al., 2005]. For each simulation, we fix the measurement cost $m/n = 1/4$ and generate the matrix $\mathbf{A}$ with Gaussian random variables. We then reconstruct the volume densities by minimizing the objective functional $|\mathbf{x}|_1 + \lambda |\mathbf{x}'|_1$ at multiple λ values. This simulation is performed for all rows of each volume, and the averaged reconstruction error is computed for each λ value. We found $\lambda = 1$ gives the minimal reconstruction error overall, and thus we set $\lambda = 1$ in all the experiments.

6.3 Attenuation Correction

Until now, we have not yet considered the attenuation in the image formation model in Equation (3). To take into account attenuation, we use a simple iterative linearization algorithm as follows:

1) Assume no attenuation, solve the optimization problem with techniques from Section 6.2 to get the initial reconstruction of the volume density $\rho^{(0)}(x, y, z)$.
2) At iteration k, compute the attenuated light as:
$$L^{(k)}(x, y, z) = \exp\left(-(\tau_1 + \tau_2)\right) \cdot L(x, y),$$
where τ_1 and τ_2 are the optical thicknesses which can be computed using the volume density from the previous iteration $\rho^{(k-1)}(x, y, z)$ as
$$\tau_1 = (\sigma_a + \sigma_s) \int_z \rho^{(k-1)}(x, y, z)\, dz,$$
$$\tau_2 = (\sigma_a + \sigma_s) \int_x \rho^{(k-1)}(x, y, z)\, dx,$$
where σ_a and σ_s are the absorption cross section and the scattering cross section of the medium, which are assumed to be known They can be measured by a second camera taking the shadowgraph of the volume.
3) With the attenuated light $L^{(k)}(x, y, z)$, Equation (3) becomes a linear equation. We solve for $\rho^{(k)}(x, y, z)$ and go to next iteration until it converges. In practice, we found that the algorithm usually converges within 3-4 iterations.

Since our method accommodates the scanning method [Deusch and Dracos, 2001; Hawkins et al., 2005] and the interpolation method [Fuchs et al., 2007] as special cases, the iterative algorithm could be directly applied to these prior methods as well.

7 VALIDATION WITH SIMULATION

In this section, we perform simulations in order to compare the several objective functionals shown in Table 1 and validate their accuracy. Comparison of these reconstruction methods is first performed on 1D synthetic signals. These signals are randomly sampled rows from the volume density of smoke acquired in [Hawkins et al., 2005]. We restrict the measurement cost, m/n, to be $1/4$. The measurement ensemble, $\mathbf{A}$, is generated in a way that each element is drawn independently from a normal distribution and each column is normalized to 1, which is effectively a white noise matrix and is known to be good for compressive sensing [Donoho, 2006]. Normalized Root Mean Squared Error (NRMSE) is used as the measure of error, which is defined as:

$$\text{NRMSE} = \frac{1}{x_{\max} - x_{\min}} \sqrt{\frac{\sum_{i=1}^{n} (\hat{x}_i - x_i)^2}{n}}, \qquad (8)$$

where $\hat{x}_i$ and x_i are the i-th element of the reconstructed signal and the original signal, respectively, and $x_{\max} - x_{\min}$ is the range of the signal.

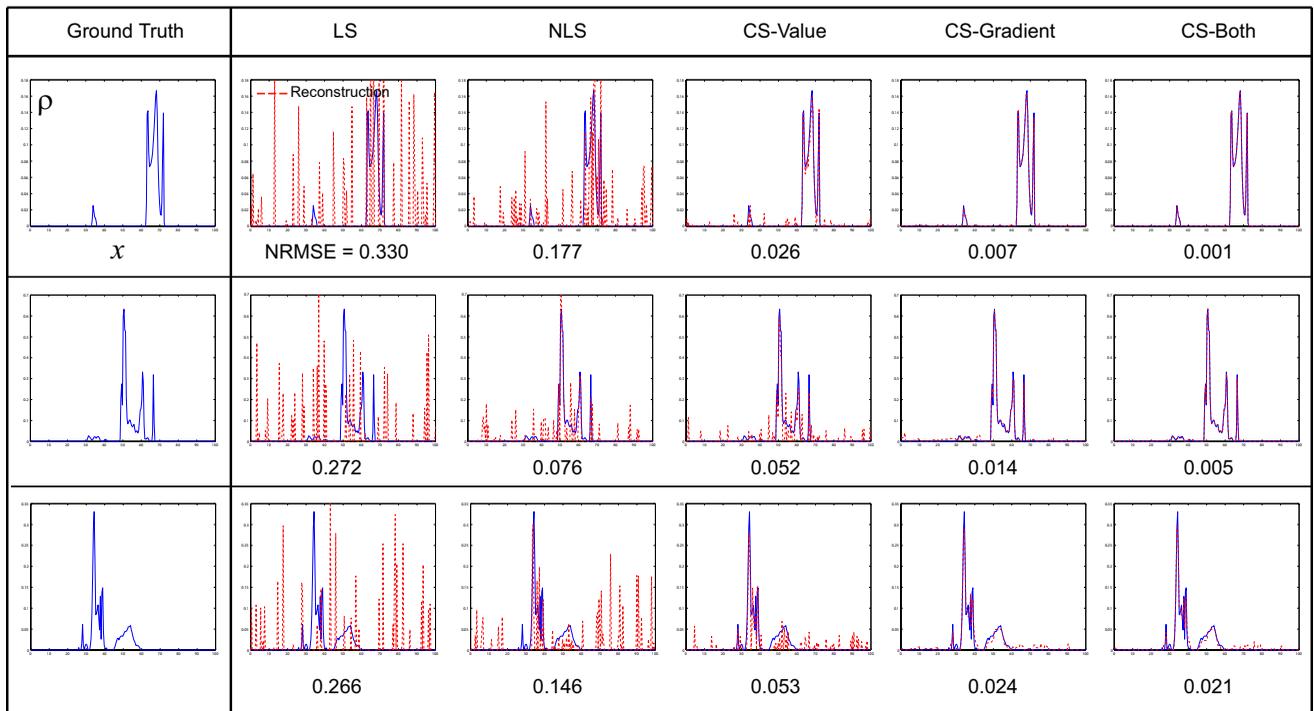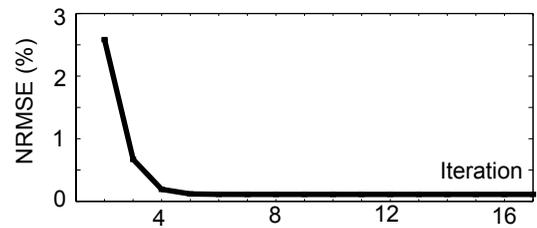| Ground Truth	LS	NLS	CS-Value	CS-Gradient	CS-Both
ρ / x	NRMSE = 0.330	0.177	0.026	0.007	0.001
	0.272	0.076	0.052	0.014	0.005
	0.266	0.146	0.053	0.024	0.021

Fig. 5. Comparison of the objective functionals for volumetric reconstruction. The first column is the original signal. The remaining columns show reconstruction results (red dashed lines) for different methods, given the measurement cost, m/n, is equal to $1/4$. The value below each plot is the NRMSE (Normalized Root Mean Squared Error) of reconstruction.
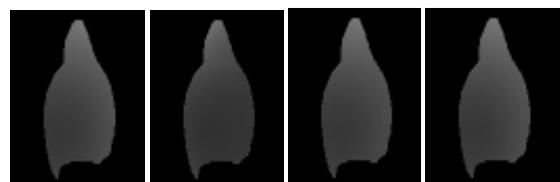
The reconstruction results are shown in Figure 5. The commonly-used LS performs the worst, since it merely minimizes the errors without using any prior on the data. With the nonnegative constraint added, NLS has better performance. CS-Value and CS-Gradient are better than NLS given that both use one more prior—the sparsity on the signal value or on the signal gradient. CS-Both($\lambda = 1$) outperforms other methods due to its adaptive ability. In our trials, the favorable performance of CS-Both was not very sensitive to changes of λ.

These observations carry over to the 3D setting (see Figure 6), where we reconstruct a $128 \times 128 \times 128$ volume; note that this requires 128×128 independent 1D reconstructions. The volume is generated from a triangular mesh of a horse and it is divided into $128 \times 128 \times 128$ voxels. For each voxel, if it is inside the mesh, the density is designed to be proportional to the distance from the center of the voxel to the center of the mesh, otherwise the density is 0 — we intentionally design the volume densities to be non-sparse (smooth-varying) in order to test the algorithm. Figure 6(a) shows the volume where blue corresponds to the lowest density while yellow corresponds to the highest density. A slice of the volume is shown in Figure 6(b).

Both spatial coding and temporal coding of compressive structured light are tested. The measurement cost, m/n, is fixed to $1/4$. For spatial coding, we use a random color image with resolution of 1280×1280 as the coded light from the projector. This gives us

(a)

Ground Truth   Iteration 1   Iteration 2   Iteration 3

(b)

Fig. 7. Simulation results for iterative attenuation correction. (a) Reconstruction errors and (b) slices with iterative attenuation correction.

$m = 1280/128 \times 3 = 30$ measurements to recover densities of 128 voxels on one row of the volume. Based on Equation (3), a single image is generated from the camera view and used for reconstruction. For temporal coding, we use random binary stripes as illumination and generate 32 images for reconstruction. CS-Both is used to reconstruct the volume for both cases. As shown in Figure 6, both methods accurately reconstruct the volume. Moreover, Figure 7 shows the reconstruction errors and reconstructed slices at different iterations of attenuation correction,
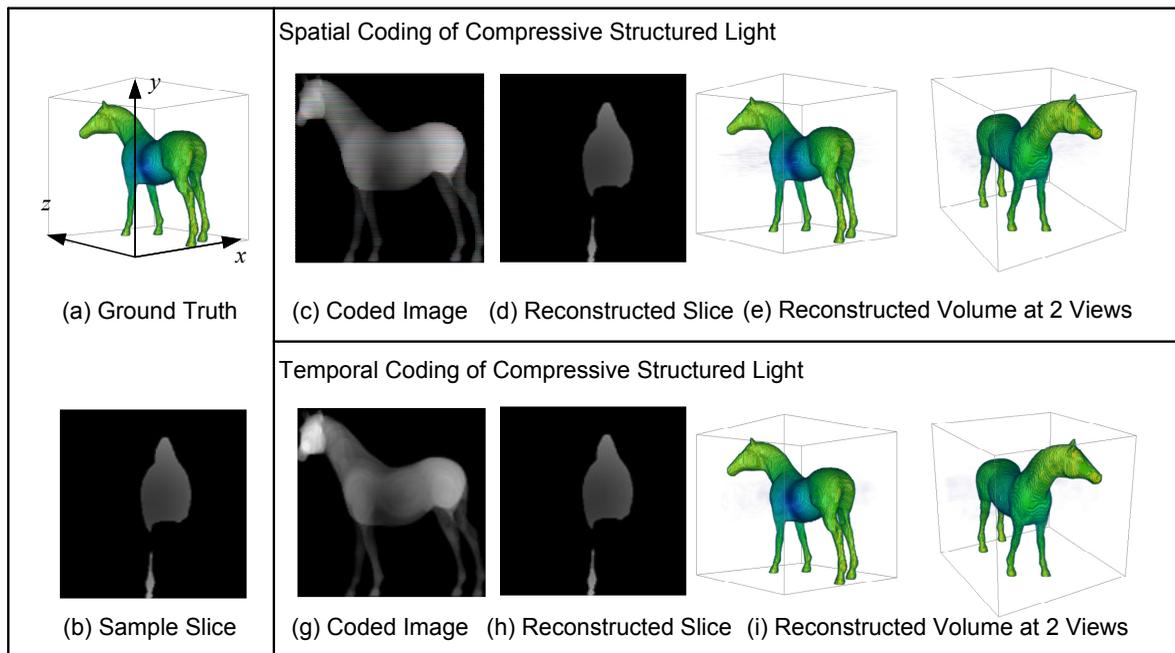
Fig. 6. Simulation results of volumetric reconstruction using compressive structured light. (a) The original volume where blue means the lowest density and yellow means the highest density. (b) A slice of the volume. On the right, the top and the bottom row shows the reconstruction results for spatial coding and temporal coding, respectively. For each row, from left to right are the coded image acquired by the camera, the reconstruction of the slice, and the reconstructed volume under two different views.

TABLE 2
NRMSE for two coding patterns with noise

Noise	Pattern	Measurement Cost (m/n)			
		1/8	1/4	1/2	1
$\sigma = 0.001$	Hadamard	0.041	0.021	0.0007	2.47e-05
	Random	0.0063	0.0010	8.91e-05	3.13e-05
$\sigma = 0.005$	Hadamard	0.042	0.023	0.0014	0.0011
	Random	0.008	0.0026	0.0017	0.0013
$\sigma = 0.01$	Hadamard	0.045	0.022	0.0026	0.0020
	Random	0.011	0.0042	0.0031	0.0023

which demonstrates the effectiveness of the iterative algorithm.

We also evaluate different reconstruction methods at various measurement costs from 1/16 to 1. The results are shown as a table in Figure 8. Conclusions similar to the ones from the previous 1D signal simulation can be drawn from these results: (1) As expected, all methods have improvements as the measurement cost increases. (2) Without using any prior of the data, LS is the worst for reconstruction with insufficient measurements. (3) CS-Gradient and CS-Both largely outperform other methods, especially for low measurement cost, which indicates strong sparsity in the signal's gradient. (4) CS-Both is better than CS-Gradient, especially at low measurement cost (*e.g.*, as shown in Figure 8 at $m/n = 1/16$). Based on these preliminary simulations, we chose to run our actual acquisition experiments with a measurement cost of 1/4 and the CS-Both optimization functional.

Table 2 shows the reconstruction error (NRMSE) for the CS-Both method for two coding patterns (*i.e.*, partial Hadamard code and random binary code) at different noise levels. As shown, the CS-Both method is robust across multiple noise levels. When the measurement cost is low, random binary code has better performance than the Hadamard code. If we have enough measurement (*i.e.*, the measurement cost is 1), the Hadamard code has better performance, as expected [Schechner et al., 2007]. In our experiments, we used random binary coding pattern.

8 EXPERIMENTAL RESULTS

We have implemented the temporal coding of compressive structured light for recovering inhomogeneous participating media. Our system consists of a 1024×768 DLP projector and a 640×480 Dragonfly Express 8-bit camera, positioned at right angles, both viewing the inhomogeneous participating medium (milk drops in water). The projector and the camera are synchronized and both operate at 360fps. The camera's resolution is set to 320×140 in order to achieve 360fps. Using 24 coded light patterns, we are able to recover a $128 \times 128 \times 128$ volume at 15fps. These light patterns consist of 128 vertical stripes. Each stripe is assigned 0 or 1 randomly with the probability of 0.5. In this way, about half the amount of the light is turned on for each measurement. We also tried alternative light patterns such as Hadamard codes, and found the random binary codes have better performance. The 24 light patterns correspond to 24

m/n	1/16	1/8	1/4	1/2	1
LS					
NLS					
CS-Value					
CS-Gradient					
CS-Both					

Fig. 8. Comparison of the objective functionals at different measurement costs m/n. CS-Both outperforms other methods.

randomly chosen rows from a 127×127 Hadamard matrix.

We used this system to recover several types of inhomogeneous participating media, including, multiple translucent layers (Figure 10), a 3D point cloud of a face etched in a glass cube (Figure 11), and the dynamic process of milk mixing with water (Figure 12). The reconstructed volumes are visualized with the ray casting algorithm [Schroeder et al., 2006] in which the opacity function is set to the volume density.

8.1 Recovery of Static Volumes

We first perform reconstruction on static volumes. Figure 10 shows the results of an object consisting of two glass slabs with powder on both. The letters "EC" are drawn manually on the back plane and "CV" on the front plane by removing the powder. Thus we create a volume in which only two planes have non-zero density. A photograph of the object is shown in Figure 10a. We then project coded light patterns on the object and reconstruct the volume using the proposed method. Figure 10 shows one of
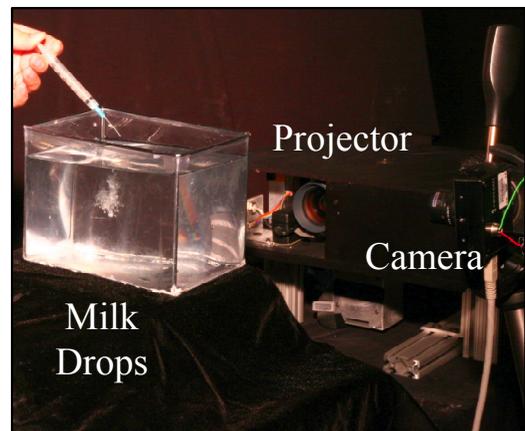

Fig. 9. Experimental setup for compressive structured light. The projector and the camera are synchronized.

the 24 captured images as well as the reconstructed volume at different views. We show the reconstructed volume with and without attenuation correction. It shows that attenuation correction improves the results by increasing the density on the back plane.
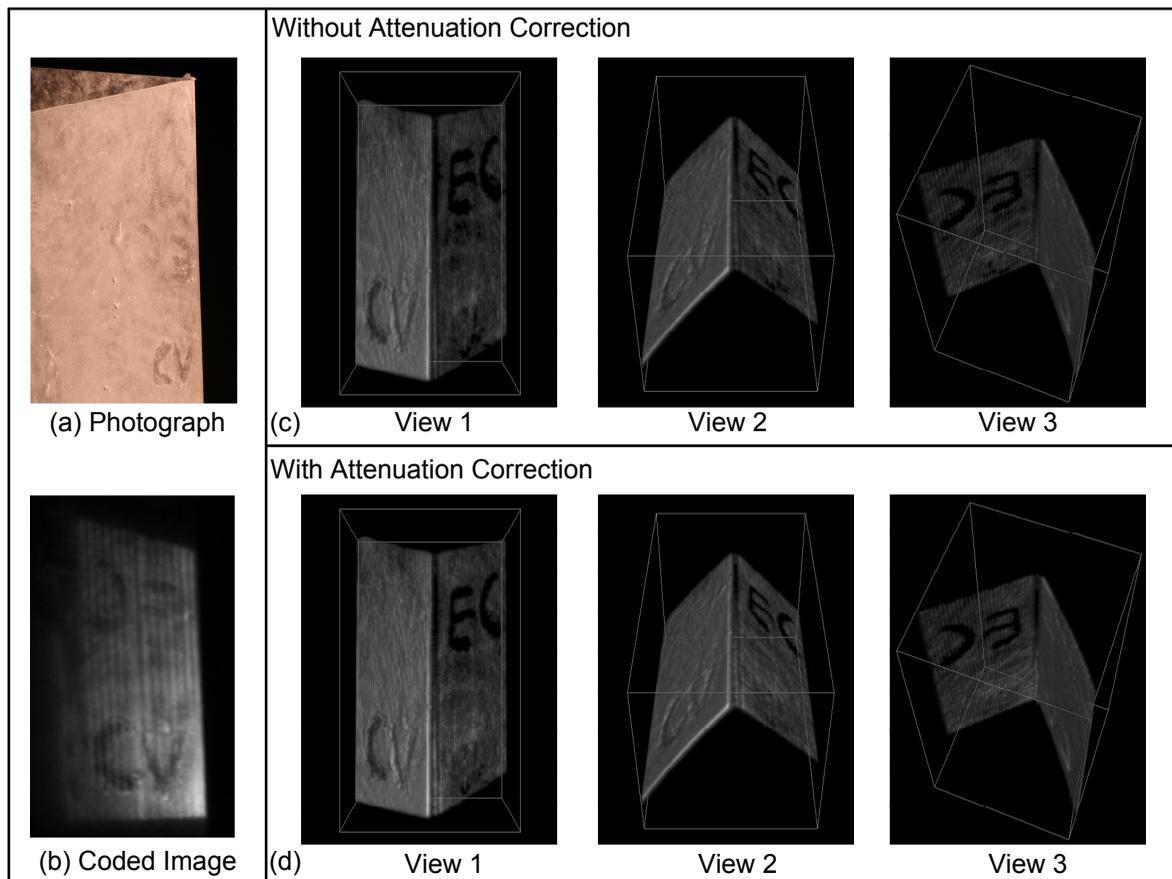
Fig. 10. Reconstruction results of two planes. (a) A photograph of the object consisting of two glass slabs with powder. The letters "EC" are on the back slab and "CV" on the front slab. (b) One of the 24 images captured by the camera. Reconstructed volume at different views with (c) and without (d) attenuation correction.

Figure 11 shows the reconstruction for a 3D point cloud of a face etched in a glass cube. As seen, our method achieved good reconstruction of the volume. In this example, multiple scattering and attenuation within the point cloud are much stronger than the previous example. As a result, in the reconstructed volume, the half of the face not directly visible to the camera has a lower estimated density (*e.g.*, the relative darker area of the right eye and right forehead in Figure 11).

8.2 Recovery of Dynamic Volumes

Finally, we use our system to reconstruct time-varying volumes. We take the dynamic process of milk drops dissolving in water as an example. We use a syringe to drip milk drops into a water tank as shown in the adjacent figure. With the proposed method, we are able to reconstruct time-varying volumes with high spatial resolution ($128 \times 128 \times 250$) at 15fps, which recovers the interesting patterns of the dynamic process (see Figure 12).

9 SUMMARY AND DISCUSSION

We proposed compressive structured light for recovering the volume densities of inhomogeneous participating media. Unlike conventional structured light range finding methods where coded light patterns are used to establish correspondence for triangulation, compressive structured light uses coded light as a way to generate measurements which are line-integrals of volume density. By exploiting the sparsity of the volume density, the volume can be accurately reconstructed from a few measurements. This makes the acquisition highly efficient both in acquisition time and illumination power, and thus enables the recovery of time-varying volumetric phenomena.

There are several limitations to our proposed compressive structured light formulation, which need to be addressed in the future:

1) *Multiple Scattering*. Although utilizing more light elements increases the efficiency of the acquisition, it will increase multiple scattering as well, which will cause biased reconstruction, as the artifacts shown in Figure 11. One potential way to alleviate this problem is to separate multiple/single scattering by using more complex light codes in a similar way to [Nayar et al., 2006]. For example, instead of using vertical stripes, one could use vertical 0/1-interleaved stripes and thus estimate the global illumination from neighboring pixels. This, however, requires trading off spatial resolution.
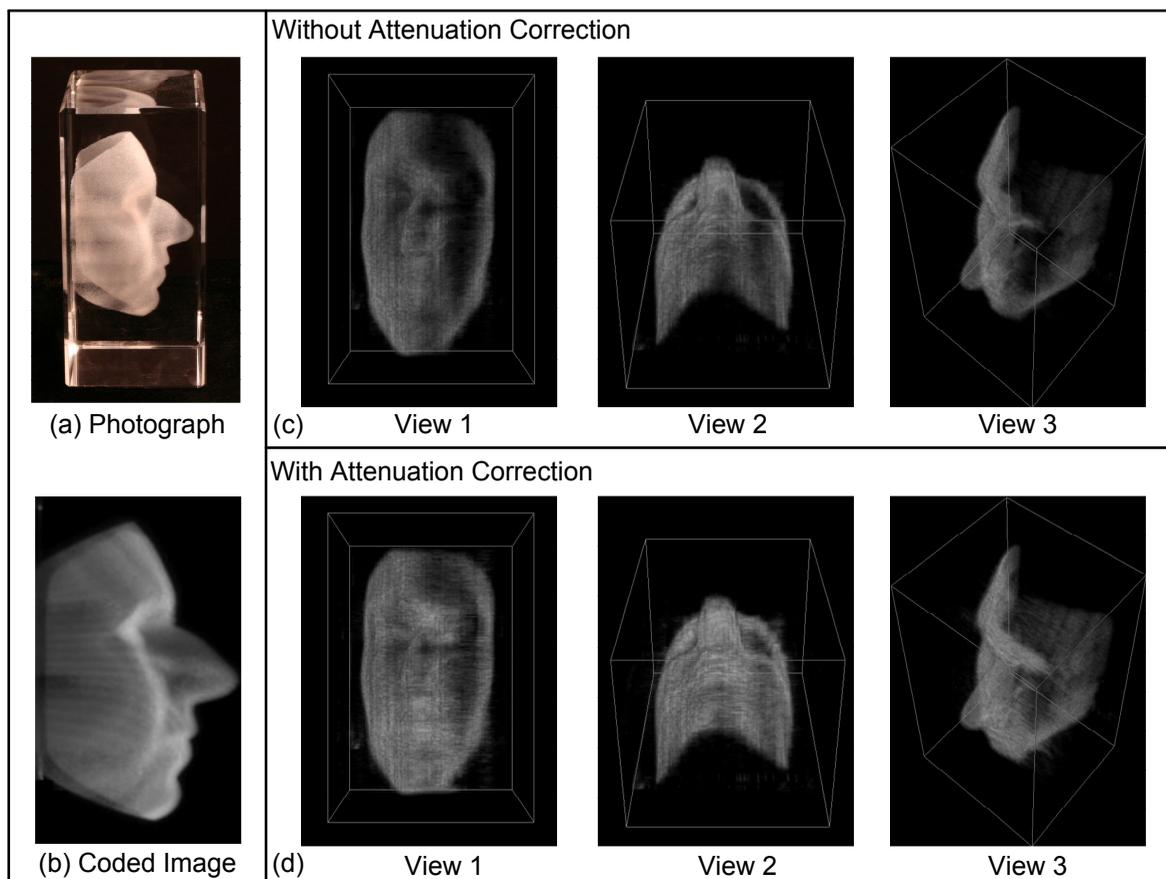
Fig. 11. Reconstruction results of a 3D point cloud of a face etched in a glass cube. (a) A photograph of the object. (b) One of the 24 images captured by the camera. Reconstruction results are shown on the right side in the same manner as Figure 10.

2) *Calibration for the Spatial Coding Method*. The spatial coding seems more desirable than the temporal coding due to its high temporal resolution (*i.e.,* volume reconstruction from one single image) and the easy access of high spatial resolution devices. However, it requires highly accurate calibration both geometrically and radiometrically. The defocus of both the projector and the camera needs to be considered as well. In contrast, the temporal coding method is more robust to noise and defocus and easy to calibrate.

We view compressive structured light as a general framework for coding the 3D light function $L(x, y, t)$ for reconstruction of signals from line-integral measurements. In this light, existing methods such as laser sheet scanning and laser line interpolation, as well as the spatial coding and temporal coding discussed in this work, can be considered as special cases. One interesting future direction is to design more complex coding strategies to improve the performance or apply the method to new problems.

ACKNOWLEDGMENTS

REFERENCES

Emmanuel J. Candes, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

S. Deusch and T. Dracos. Time resolved 3D passive scalar concentration-field imagingby induced fluorescence (LIF) in moving liquids. *Measurement Science and Technology*, 12(2):188–200, 2001.

David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

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

Christian Fuchs, Tongbo Chen, Michael Goesele, Holger Theisel, and Hans-Peter Seidel. Density estimation for dynamic volumes. *Computers and Graphics*, 31(2):205–211, 2007.
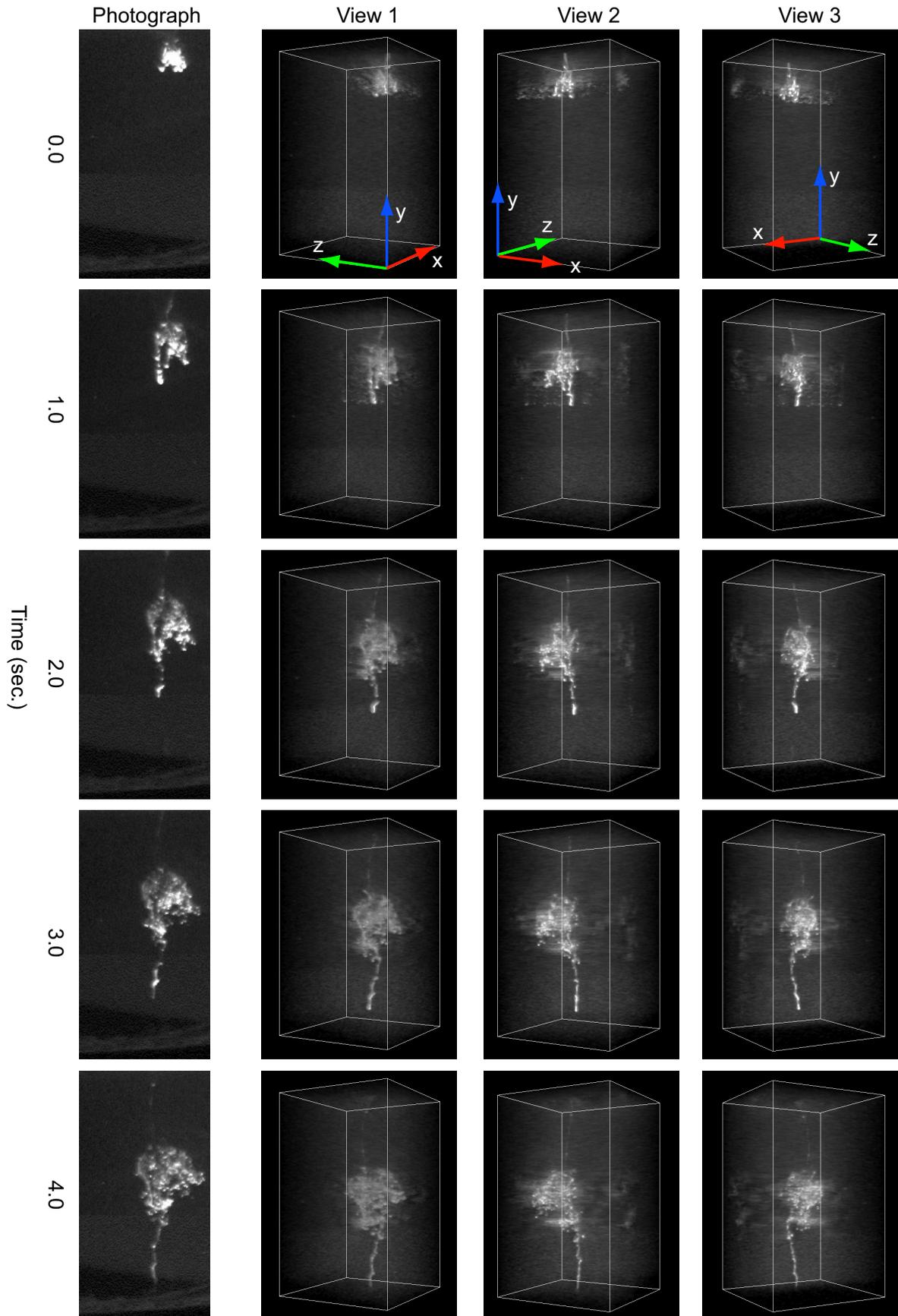
Fig. 12. Reconstruction of milk drops dissolving in water. 24 images are used to reconstruct a $128 \times 128 \times 250$ volume at 15fps. Three views are shown. Each row corresponds to one instance in time. The leftmost column shows the corresponding photographs.

C. Gini. Measurement of inequality of incomes. *The Economic Journal*, pages 124–126, 1921.

Jinwei Gu, Shree Nayar, Eitan Grinspun, Peter Belhumeur, and Ravi Ramamoorthi. Compressive structured light for recovering inhomogenous participating media. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 845–858, 2008.

Samuel W. Hasinoff and Kiriakos N. Kutulakos. Photo-consistent reconstruction of semi-transparent scenes by density sheet decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):870–885, 2007.

Tim Hawkins, Per Einarsson, and Paul Debevec. Acquisition of time-varying participating media. *ACM Transactions on Graphics (SIGGRAPH)*, 24(3):812–815, 2005.

Yasunobu Hitomi, Jinwei Gu, Mohit Gupta, Tomoo Mitsunaga, and Shree K Nayar. Video from a singe coded exposure photograph using a learned over-complete dictionary. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2011.

Niall P. Hurley and Scott T. Rickard. Comparing measures of sparsity. *IEEE Transactions on Information Theory*, 55(10):4723–4741, 2009.

Ivo Ihrke and Marcus Magnor. Image-based tomographic reconstruction of flames. In *ACM SIGGRAPH / Eurographics Symposium on Computer Animation (SCA)*, pages 367–375, 2004.

Ivo Ihrke and Marcus Magnor. Adaptive grid optical tomography. *Graphical Models*, 68(5):484–495, 2006.

Ivo Ihrke, Kiriakos N. Kutulakos, Hendrik P. A. Lensch, Marcus Magnor, and Wolfgang Heidrich. State of the art in transparent and specular object reconstruction. In *STAR Proceedings of Eurographics*, 2008.

Akira Ishimaru. *Wave Propagation and Scattering in Random Media*. IEEE/OUP Series on Electromagnetic Wave Theory. IEEE Press, New York, 1978.

Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MRI imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. Non-local sparse models for image restoration. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 2272–2279, 2009.

Srinivasa G. Narasimhan, Shree K. Nayar, Bo Sun, and Sanjeev J. Koppal. Structured light in scattering media. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 420–427, 2005.

Shree K. Nayar, Gurunandan Krishnan, Michael D. Grossberg, and Ramesh Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Transactions on Graphics (SIGGRAPH)*, 25(3):935–944, 2006.

Bruno A. Olshausen and David J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

Pieter Peers, Dhruv K. Mahajan, Bruce Lamond, Abhijeet Ghosh, Wojciech Matusik, Ravi Ramamoorthi, and Paul Debevec. Compressive light transport sensing. *ACM Transactions on Graphics*, 28(3:1-3:18): 1289–1306, 2009.

Matan Protter and Michael Elad. Image sequence denoising via sparse and redundant representations. *IEEE Transactions on Image Processing*, 18(1):27–35, 2009.

Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, 1992.

Joaquim Salvi, Jordi Pages, and Joan Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37:827–849, 2004.

Aswin C. Sankaranarayanan, Pavan K. Turaga, Richard G. Baraniuk, and Rama Chellappa. Compressive acquisition of dynamic scenes. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2010.

Yoav Y. Schechner, Shree K. Nayar, and Peter N. Belhumeur. Multiplexing for optimal lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1339–1354, 2007.

Will Schroeder, Ken Martin, and Bill Lorensen. *The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics*. Pearson Eduction, Inc., 4th edition, 2006.

Pradeep Sen and Soheil Darabi. Compressive dual photography. *Computer Graphics Forum*, 28(2), March 2009.

Eero P Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *Proceedings of Asilomar Conference on Signals, Systems and Computers*, pages 673–678, 1997.

D. Takhar, J. Laska, M. Walkin, M. Durate, and D.Baron. A new compressive imaging camera architecture using optical-domain compression. In *Computational Imaging IV at SPIE Electronic Imaging*, 2006.

Borislav Trifonov, Derek Bradley, and Wolfgang Heidrich. Tomographic reconstruction of transparent objects. In *Eurographics Symposium on Rendering (EGSR)*, pages 51–60, 2006.

Ashok Veeraraghavan, Dikpal Reddy, and Ramesh Raskar. Coded strobing photography: Compressive sensing of high-speed periodic events. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):671–686, 2011.

R. Willett, M. Gehm, and D. Brady. Multiscale reconstruction for computational spectral imaging. In *Computational Imaging V at SPIE Electronic Imaging*, 2007.

John Wright, Allen Y. Yang, Arvind Ganesh,

Shankar S. Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

Jinwei Gu is currently an assistant professor in the Munsell Color Science Laboratory in the Center for Imaging Science at Rochester Institute of Technology. He received his PhD degree from Columbia University in May 2010, and his bachelor and master degree from Tsinghua University, China in 2002 and 2005. His research interests are computer vision and computer graphics. His current research focuses on computational photography, physics-based computer vision, and data-driven computer graphics.

Shree K. Nayar received his PhD degree in Electrical and Computer Engineering from the Robotics Institute at Carnegie Mellon University in 1990. He is currently the T. C. Chang Professor of Computer Science at Columbia University. He co-directs the Columbia Vision and Graphics Center. He also heads the Columbia Computer Vision Laboratory (CAVE), which is dedicated to the development of advanced computer vision systems. His research is focused on three areas; the creation of novel cameras, the design of physics based models for vision, and the development of algorithms for scene understanding. His work is motivated by applications in the fields of digital imaging, computer graphics, and robotics. He has received best paper awards at ICCV 1990, ICPR 1994, CVPR 1994, ICCV 1995, CVPR 2000 and CVPR 2004. He is the recipient of the David Marr Prize (1990 and 1995), the David and Lucile Packard Fellowship (1992), the National Young Investigator Award (1993), the NTT Distinguished Scientific Achievement Award (1994), the Keck Foundation Award for Excellence in Teaching (1995), the Columbia Great Teacher Award (2006) and Carnegie Mellon University's Alumni Achievement Award. In February 2008, he was elected to the National Academy of Engineering.

Eitan Grinspun is Associate Professor of Computer Science at Columbia University, and Director of the Columbia Computer Graphics Group. His research seeks to discover connections between geometry, physics, and computation, typically with applications to computer graphics. He received his Ph.D. in Computer Science from the California Institute of Technology in 2003, and his B.A.Sc. in Engineering Science from the University of Toronto in 1997. He was Professeur d'Universit Invit in Paris at l'Universit Pierre et Marie Curie in 2009, and a Research Scientist at the Courant Institute of Mathematical Sciences in 2003-04, He is an Alfred P. Sloan Research Fellow and NSF CAREER Award recipient, and was previously an NVIDIA Fellow and a Caltech Everhart Distinguished Lecturer. The technologies developed by his laboratory are used in consumer software such as Adobe Photoshop & Illustrator, in film studios such as Disney, Pixar, and Weta Digital, and in physics laboratories at institutions such as MIT and the Universit Paris VI. His work has been profiled in The New York Times, Scientific American, and Popular Science (Brilliant 10).

Peter N. Belhumeur is currently a Professor in the Department of Computer Science at Columbia University and the Director of the Laboratory for the Study of Visual Appearance. He received a Sc.B. in Information Sciences from Brown University in 1985. He received his Ph.D. in Engineering Sciences from Harvard University under the direction of David Mumford in 1993. He was a postdoctoral fellow at the University of Cambridge's Isaac Newton Institute for Mathematical Sciences in 1994. He was made Assistant, Associate and Professor of Electrical Engineering at Yale University in 1994, 1998, and 2001, respectively. He joined Columbia University as a Professor of Computer Science in 2002. His research focus lies somewhere in the mix of computer vision, computer graphics, and the visual recognition of plant and animal species. His papers have more than 10,000 citations to date. His work has been featured multiple times in the New York Times, on National Public Radio, and the Associated Press. He is a recipient of the Presidential Early Career Award for Scientists and Engineers (PECASE) and the National Science Foundation Career Award. He won both the Siemens Best Paper Award at the 1996 IEEE Conference on Computer Vision and Pattern Recognition and the Olympus Prize at the 1998 European Conference of Computer Vision. In October 2011, he received the Edward O. Wilson Biodiversity Pioneer Award, given to individuals who have significantly contributed to the preservation of biodiversity on Earth.

Ravi Ramamoorthi received the BS degree in engineering and applied science and the MS degrees in computer science and physics from the California Institute of Technology in 1998. He received the PhD degree in computer science from Stanford University's Computer Graphics Laboratory in 2002, upon which he joined the Columbia University Computer Science Department. He is now an associate professor at the University of California, Berkeley since 2009. His research interests cover many aspects of computer vision and graphics, including mathematical foundations and sampling and reconstruction, real-time photorealistic rendering, data-driven appearance models, physics-based computer vision, and vision-graphics methods for image and video editing. He has published papers on all of these topics in leading graphics and vision conferences and journals. His work has been recognized by a number of awards from private foundations (Sloan Research Fellowship and Okawa Foundation Research Grant), federal agencies (NSF Career Award and ONR Young Investigator Award), professional organizations (ACM SIGGRAPH Significant New Researcher Award) and the white house (Presidential Early Career Award for Scientists and Engineers).