Differentiable Dynamic Visible-Light Tomography

Kaizhang Kang[*]
generous.kkz@gmail.com
State Key Lab of CAD&CG,
Zhejiang University
Hangzhou, China

Zoubin Bi[*]
bzb@zju.edu.cn
State Key Lab of CAD&CG,
Zhejiang University
Hangzhou, China

Xiang Feng
xfeng.cg@zju.edu.cn
State Key Lab of CAD&CG,
Zhejiang University
Hangzhou, China

Yican Dong
3190105140@zju.edu.cn
State Key Lab of CAD&CG,
Zhejiang University
Hangzhou, China

Kun Zhou[†]
kunzhou@acm.org
State Key Lab of CAD&CG, Zhejiang
University and ZJU-FaceUnity Joint
Lab of Intelligent Graphics
Hangzhou, China

Hongzhi Wu[†]
hwu@acm.org
State Key Lab of CAD&CG,
Zhejiang University
Hangzhou, China

ABSTRACT

We propose the first visible-light tomography system for real-time acquisition and reconstruction of general temporally-varying 3D phenomena. Using a single high-speed camera, a high-performance LED array and optical fibers with a total length of 5 km, we build a novel acquisition setup with no mechanical movements to simultaneously sample using 1,920 interleaved sources and detectors with a complete 360° coverage. Next, we introduce a novel differentiable framework to map both tomography acquisition and reconstruction to a carefully designed autoencoder. This allows the joint and automatic optimization of both processes in an end-to-end fashion, essentially learning to physically compress and computationally decompress the target information. Our framework can adapt to various factors, and trade between capture speed and reconstruction quality. We achieve an acquisition speed of up to 36.8 volumes per second at a spatial resolution of 32×128×128; each volume is captured with as few as 8 images. The effectiveness of the system is demonstrated on acquiring various dynamic scenes. Our results are also validated with the reconstructions computed from the measurements with one source on at a time, and compare favorably with state-of-the-art techniques.

CCS CONCEPTS

• **Computing methodologies → Volumetric models**; • **Hardware** → *Scanners*.

KEYWORDS

CT, dynamic acquisition, illumination multiplexing

ACM Reference Format:
Kaizhang Kang, Zoubin Bi, Xiang Feng, Yican Dong, Kun Zhou, and Hongzhi Wu. 2023. Differentiable Dynamic Visible-Light Tomography. In *SIGGRAPH Asia 2023 Conference Papers (SA Conference Papers '23), December 12–15, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3610548.3618166

1 INTRODUCTION

Computed Tomography (CT) is a fundamental imaging technique to obtain complete 3D structures from measurements of attenuated lights along different lines across an object [Hsieh 2003]. It has many important applications, including medical imaging, industrial inspection, aviation security, and cultural heritage. The external and internal structures revealed by CT considerably deepen our understanding of the object.

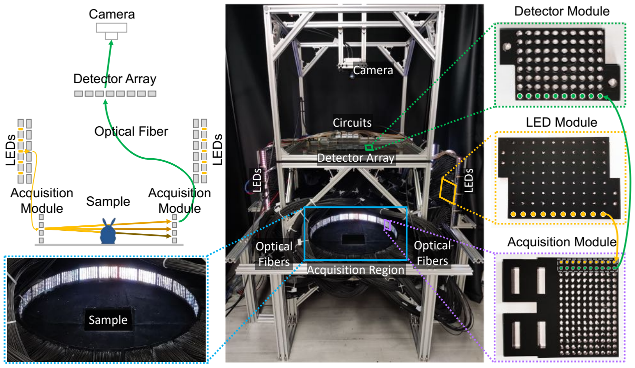


Figure 1: Using our novel stationary prototype (center) with 1,920 visible-light LEDs, 1,920 detectors and a single camera, we capture a 3D volume with as few as 8 photographs under optimized lighting patterns in 27.2 ms**. This performance allows us to acquire a variety of dynamic 3D scenes (Fig. 12). The top-left diagram illustrates the acquisition process. Please refer to Sec. 5 for more details on the prototype.**

While originally limited to scanning static scenes, it is of significant scientific and practical value to extend CT to capture dynamic 3D structures/phenomena for mechanics/biology research, medical

[*] The first two authors contributed equally.
[†] Corresponding authors: Kun Zhou & Hongzhi Wu.

This work is licensed under a Creative Commons Attribution-NoDerivs International 4.0 License.

SA Conference Papers '23, December 12–15, 2023, Sydney, NSW, Australia
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0315-7/23/12.
https://doi.org/10.1145/3610548.3618166

diagnosis, etc. However, one key difficulty arises in the extension to dynamic CT. In theory, CT requires a dense set of individual measurements along different lines intersecting a scene to reconstruct a single 3D volume. As the scene is changing temporally, this dense set of measurements must be repeatedly acquired within a short period of time to avoid ghosting artifacts and provide a sufficient temporal resolution. This translates to a requirement of considerably higher **sampling capability** beyond traditional work, making dynamic CT extremely challenging.

Over the past decades, substantial research efforts have been made towards this goal, but general dynamic CT remains difficult. One idea is to exploit specific properties of certain dynamic phenomena (e.g., periodic movements [Chen et al. 2011]); the number of required measurements can be reduced at the cost of generality. One can also capture a small set of the complete measurements, and rely on additional priors to fill in the information gap [Chen et al. 2008b; Zang et al. 2020]. Another idea is to employ multiplexing to turn on multiple sources at the same time; the exposure time can be reduced, due to the increased emitted energy. Recently, fully stationary geometry [Zhang et al. 2020] has been proposed to eliminate the time-consuming mechanical movements as in traditional CT [Hounsfield 1973]. Its sampling capability, however, is limited by the cost and space for placing multiple sources and detectors.

In this paper, we propose the first visible-light tomography system for real-time acquisition and reconstruction of general dynamic 3D phenomena. We first build a low-cost, stationary acquisition setup with 1,920 sources and 1,920 detectors, which are interleaved with each other and evenly placed on a cylinder. A total of 3.6 million rays can be sampled without any mechanical movements. The sources/sensors are connected via optical fibers to a high-performance LED array (48,000 fps)/a single high-speed camera (400 fps), respectively. This allows *rapid, concurrent* capture using all detectors with multiple sources on.

Next, we map both tomography acquisition and reconstruction to a carefully designed neural network. This allows the joint and automatic optimization of both processes in an end-to-end fashion, essentially learning to physically compress and computationally decompress the target information. The capture process is mapped to a linear fc (fully connected) layer, whose weights correspond to the lighting patterns for illumination multiplexing in the acquisition and can be automatically optimized in training. For reconstruction, our network demultiplexes the measurements and then outputs a final 3D volume via trainable filtered back projection (FBP). Similar to existing work, we use random combinations of various volumetric primitives as well as simulated 3D fluid sequences as training data.

We achieve an acquisition speed of 36.8 volumes per second at a spatial resolution of 32×128×128; each volume is captured with as few as 8 images. We also achieve a real-time reconstruction speed of 38.5 volumes per second on an 8-GPU server. The effectiveness of the system is demonstrated on 5 sequences of dynamic scenes. These scenes are captured with different numbers of pre-optimized lighting patterns, showing the flexibility of our framework to trade between acquisition speed and reconstruction quality. Our results are also validated with the reconstructions computed from the measurements with one light on at a time (OLAT). We compare with state-of-the-art techniques qualitatively and quantitatively, and evaluate the impact of various factors.

2 RELATED WORK

Below we review CT scanning geometry, acquisition and reconstruction with limited samples as well as multiplexed CT acquisition, three classes of work mostly related to our paper. Interested readers are referred to Hsieh [2003] for an excellent introduction on this extensively studied field. While our focus is on CT, it is not the only approach to reconstruct transparent/translucent objects with visible light. Similar results could be obtained with other setups and algorithms developed in computer graphics and vision (e.g., [Bemana et al. 2022; Hullin et al. 2008; Ihrke et al. 2005; Trifonov et al. 2006]). Please refer to related surveys for an overview [Ihrke et al. 2010; Zhou et al. 2021].

2.1 CT Scanning Geometry

The sampling capability of CT scanning geometry has been steadily increasing over time. Starting with a single source and detector in the seminal work of Hounsfield [1973], CT geometry has evolved from 2D parallel/fan beam to 3D cone beam (CBCT), which adds more detectors along the z axis [Arai et al. 1999; Mozzo et al. 1998]. However, the sampling rate of CBCT decreases when away from the sample plane, leading to inferior reconstructions along the z axis. This motivates inverse-geometry volumetric CT (IGCT) [Schmidt et al. 2004], which improves the sampling with more sources added to the z axis as well. Since the sampling coverage is incomplete, all the above work including their modern variants [De Man et al. 2016; Kim et al. 2021; Zhang et al. 2021] require mechanical movements to finish one scan.

Recently, fully stationary geometry [Schwoebel et al. 2014; Spronk 2021; Yao et al. 2021] are proposed, with linearly [Zhang et al. 2020] or helically interleaved sources and detectors [Chen et al. 2014]. However, due to the space conflict, only a limited number of sources and detectors can be put in place, leading to an undesired sparser sampling, compared with the counterparts with mechanical movements. In addition, since the placement of sources and detectors are derived from traditional scanning trajectories, the possibility of more efficient placement is not explored.

While vanilla CT employs X-ray, visible light is often used in computer graphics and vision for reconstructing phenomena like flames [Hasinoff and Kutulakos 2007; Ihrke and Magnor 2004] and fluids [Atcheson et al. 2008; Eckert et al. 2018]. The work here shares similar weaknesses with their X-ray counterparts, including "the small number of view points/projection images due to constraints in the hardware setup (e.g. cost of the cameras and space limitations)." as mentioned in Zang et al. [2020].

In comparison, our stationary setup is designed to achieve concurrent dense sampling capability with 1,920 sources and 1,920 detectors. Please refer Fig. 4 and 3 for a visualization. We tackle the challenge of cost and space limit, by using optical fibers to route the light from an LED array to the acquisition region and finally to a single camera.

2.2 Acquisition & Reconstruction with Limited Samples

Traditional CT reconstruction algorithms can be divided into analytical [Feldkamp et al. 1984] and algebraic methods [Gordon et al. 1970], usually taking as input a dense set of measurements.

For scenes with specific properties, dynamic CT can be achieved with limited samples. For periodic movements, one can scan different subparts at several cycles and assemble them as an effective fast scan in one cycle. The idea is successfully applied to imaging cardiac [Chen et al. 2011] and breathing motions [Sonke et al. 2005].

For general scenes, a common idea is to capture only a small set of the complete measurements and fill in the information gap with additional priors. There are two main classes based on the distribution of samples: limited angle [Anirudh et al. 2018; Huang et al. 2017] and sparse view [Chen et al. 2008b] reconstruction. Zang et. al [2018] design a low-discrepancy view-sampling strategy to acquire slowly deforming objects. Various forms of priors are proposed, including total variance (TV) [Huang et al. 2013], temporal coherence [Zang et al. 2019], subspaces spanned by different basis functions [Hasinoff and Kutulakos 2007; Ihrke and Magnor 2004], and spatial compactness [Atcheson et al. 2008].

Recently, deep learning is applied to build the reconstruction prior from data. Jin et al. [2017] train a neural network for reconstruction in an end-to-end fashion. Wang and Liu [2020] combine the FBP algorithm with a neural network to improve its generalization. A deep sinogram prediction module is proposed in Zang et al. [2021] to in-paint missing samples. Rückert et al. [2022] develop a hierarchical neural rendering pipeline for tomography reconstruction.

While our approach is also based on limited samples, we jointly optimize the acquisition along with reconstruction, to fully exploit the sampling capability of our setup.

2.3 Multiplexed CT Acquisition

Zhang et al. [2010; 2008] apply Hadamard multiplexing to the sources to decrease the exposure time of each image. It is not clear how to further reduce the required number of image, which is the same of the number of sources. While learned illumination multiplexing has made a considerable success in other contexts (e.g., reflectance acquisition [Kang et al. 2018]), it is challenging to apply the idea to CT due to the huge domain gap. In comparison, we automatically learn a compressive multiplexed scheme along with a corresponding CT reconstruction algorithm, using a highly tight budget on the number of measurements.

3 OVERVIEW

We build an acquisition setup with multiple interleaved sources and detectors on a cylinder, which are connected to an LED array and a camera via optical fibers, respectively. To capture a dynamic scene, we place it around the center of the cylinder. Pre-optimized lighting patterns (whose number is denoted as #) are cast from the LED array to all sources simultaneously, resulting in a dense set of rays that pass through the scene. The detectors pick up the rays and direct them to a common plane, where they are captured by the synchronized camera and then extracted as measurements. The measurements corresponding to lighting patterns are fed to our network for reconstructing a single 3D volume. Please refer to Fig. 2 for a visualization of our pipeline.

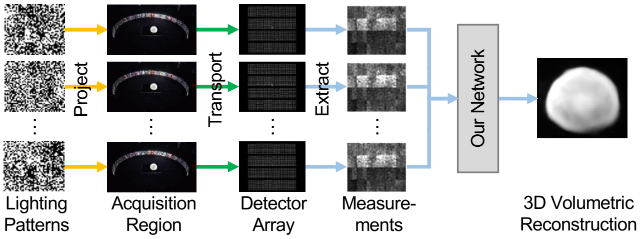


Figure 2: Acquisition pipeline. Pre-optimized lighting patterns are cast by LEDs to illuminate the scene via optical fibers. Attenuated lights are picked up by the detectors and directed via optical fibers to a detector array. Measurements are extracted from images taken at the array. The measurements from all # images are sent to our network to reconstruct a 3D density volume.

4 PRELIMINARIES

4.1 Assumptions

We assume that the scene remains static in the duration for capturing every # photographs. To support general dynamic scenes, no temporal coherence is exploited and each 3D volume is *independently reconstructed*. Similar to the majority of existing work, we do not consider refractions or reflections. In addition, only a single gray-scale channel is used throughout the paper.

4.2 Tomographic Imaging Model

We use a long vector $\mathbf{x}$ to represent a discrete 3D density volume with a spatial resolution of 32×128×128. If we set the source with an index of j to the maximum power, the corresponding readout value from the detector with an index of i is denoted as $\mathbf{I}_{ij}$. The ratio between $\mathbf{I}_{ij}$ and the counterpart $\tilde{\mathbf{I}}_{ij}$ on an empty scene is related to the accumulated density $\mathbf{D}_{ij}$ along the path from source j to detector i as:

$$\mathbf{D}_{ij} = -\log(\mathbf{I}_{ij}/\tilde{\mathbf{I}}_{ij}). \tag{1}$$

The collection of the accumulated densities at all source-detector pairs is called a **sinogram**, and can be stored in a 1920×1920 matrix $\mathbf{D}$.

In CT literature, $\mathbf{D}_{ij}$ is computed as a line integral over the density volume $\mathbf{x}$, which is called the Radon transform projection model. We use a matrix $\mathbf{K}$ to represent this procedure:

$$\mathbf{D} = \text{Reshape}(\mathbf{K} \cdot \mathbf{x}). \tag{2}$$

Here Reshape() reorganizes a vector to a matrix.

Combining Eq. 1 and 2, we derive the following relationship between detector readout values and the density volume:

$$\mathbf{I} = \exp(-\text{Reshape}(\mathbf{K} \cdot \mathbf{x})) \odot \tilde{\mathbf{I}}. \tag{3}$$

Here ⊙ represents element-wise product between two matrices, $\mathbf{I}$ is a 1920×1920 matrix called a **multi-view CT image**, and each column of $\mathbf{I}$ is referred to as a **CT image**.

Below we describe $s - \phi$ parameterization of 2D lines. For a 2D line, s is its distance to the origin, and ϕ is the angle between the x-axis and another line orthogonal to the current one and through the origin, in the range of $[-\pi, \pi)$. Every 2D line can be represented as a point in the $s - \phi$ plane, which is referred to as the **projection space**. For a line in the 3D space, it can be parameterized by two angles and two distances, resulting in a 4D projection space. Please refer to Fig. 3 for a visualization.

4.3 Multiplexing & FBP-based Reconstruction

We represent a **lighting pattern** as a vector $\mathbf{p}$ in $\mathbb{R}^{1920}$, in which $\mathbf{p}_j$ stores the intensity set to the source with an index of j. Due to the linearity in the physical domain, the readouts of all detectors under a lighting pattern $\mathbf{p}$ can be computed as:

$$\mathbf{m} = \mathbf{I} \cdot \mathbf{p} = (\exp(-\text{Reshape}(\mathbf{K} \cdot \mathbf{x})) \odot \tilde{\mathbf{I}}) \cdot \mathbf{p}. \tag{4}$$

Here $\mathbf{m}$ is a **multiplexed CT image**, represented as a vector in $\mathbb{R}^{1920}$. As we use multiple lighting patterns $\mathbf{P}$ to acquire a volume, the collection of all measurements $\mathbf{M}$ related to this volume can be expressed as:

$$\mathbf{M} = \mathbf{I} \cdot \mathbf{P} = (\exp(-\text{Reshape}(\mathbf{K} \cdot \mathbf{x})) \odot \tilde{\mathbf{I}}) \cdot \mathbf{P}, \tag{5}$$

where $\mathbf{M}$ and $\mathbf{P}$ are both $1920 \times \#$ matrices.

To reconstruct a density volume $\mathbf{x}$ from a sinogram $\mathbf{D}$, FBP/FDK (Feldkamp, Davis and Kress [1984]) algorithms are widely used in industry as:

$$\tilde{\mathbf{x}} = \mathbf{B} f(\mathbf{D}), \tag{6}$$

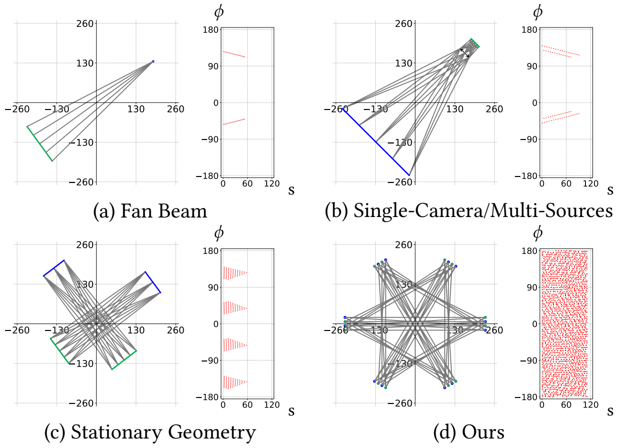
where f is a filtering function, and $\mathbf{B}$ is the back projection operation. More details can be found in Hsieh [2003].

5 ACQUISITION PROTOTYPE

5.1 Design Decisions

Our goal is to build a setup with a high sampling capability within a short period of time to support dynamic capture. The setup should be stationary, as the mechanical movement limits the temporal resolution of the reconstructions. Moreover, it is desirable to sample the projection space as densely and as completely as possible, which is directly related to the final quality (Sec. 4.2). Through trial-and-error, we examine and compare the sampling capabilities of different geometries (see Fig. 3 for a visualization of 2D examples). We find that an interleaved layout of many sources and detectors on a cylinder around the target allows a dense and complete sampling of the projection space, leading to a 10x increase in coverage compared with one state-of-the-art work [Zhang et al. 2020].

However, it is not straightforward to directly implement the above design with LEDs as sources and cameras as detectors, due to the cost and space limit. Therefore, we employ light routing [Pereira et al. 2014] to flexibly direct the light emitted by an LED array to the desired locations on the cylinder, and similarly direct the attenuated light collected along various directions to a common plane to be captured by a single camera, with the help of optical fibers. For convenience of installation, we group every 10×8 sources/detectors together as an LED/detector/acquisition module, made of laser-cut acrylic. Please refer to Fig. 1 and 4 for illustration.



(a) Fan Beam

(b) Single-Camera/Multi-Sources

(c) Stationary Geometry

(d) Ours

Figure 3: Different scanning geometries and their concurrent sampling capabilities on $s - \phi$ plane. For each pair of images, we draw representatives among all concurrent rays for a specific type of scanning geometry on the left. All sampled rays in the $s - \phi$ plane are marked as red dots. The blue dot represents a source, and green a detector. For a fair comparison, we use the same source/detector spacing. The geometry of Zhang et al. [2020] is shown in (c). According to the method in Luo et al. [2021], the area covered on the $s - \phi$ plane by sampled rays in (c) and (d) are 4,318 **and** 43,200(degree · mm)**, respectively.**

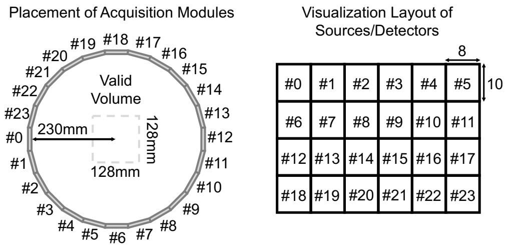


Figure 4: The placement of acquisition modules from the top view (left) and the visualization layout for light sources as well as detectors (right). Acquisition module IDs are also marked for reference.

5.2 Prototype Description

The intensity of each LED in the array is independently controlled by custom designed circuits with Pulse Width Modulation (PWM). The binary lighting pattern projection speed is 48,000 fps. As shown in Fig. 4, the emitted light is routed from an LED module to a corresponding acquisition module as effective sources. The remaining half holes on the acquisition module serve as detectors. They collect lights and direct via optical fibers to a detector array on a common plane for acquisition. We use a 3 MP machine vision camera Basler boA1936-400cm to capture the plane, with a maximum speed of 400 fps. The camera is precisely synchronized with the light casting of the LED array using our circuits.

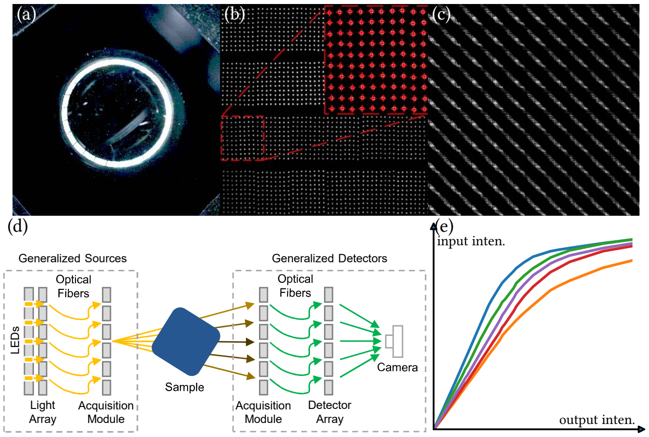


Figure 5: Calibrations. (a) The top-view of a calibration hemisphere in the center of the valid volume, reflecting all sources. (b) The photo of the detector array captured by our camera. The red crosses indicate the calibrated centers of each optical fiber. (c) A visualization of calibrated $\tilde{\mathbf{I}}$. Each row/column corresponds to a source/detector, respectively. (d) A visualization of generalized source and detector. (e) Representatives of calibrated response curves of generalized detectors.

We use Mitsubishi Eska MEGA SH4001 optical fibers, with a fiber diameter of 1 mm and a jacket diameter of 2.2 mm. The inner material is plastic designed for visible light routing. As aforementioned, we design 3 types of modules (LED/detector/acquisition), to ease the installation of optical fibers. A number of holes are pre-cut from the modules, and the diameter of each hole is slightly larger than an optical fiber. We first use a professional optical fiber cleaver to cut the optical fibers with a total length of about 5 km into 1920×2 segments. Next, we apply LOCTITE 406 glue to fit each end of a segment to a hole in a module. The acquisition module has 10×16 holes. The odd columns are for the fibers from an LED module, and the remaining for those from a detector module. It takes about 100 hours for a graduate student to finish the process. We place 24 acquisition modules to form a cylinder-like geometry with a radius of 230 mm. This ends up with 24×10×8=1,920 sources and detectors around a designated valid volume of 32 mm×128 mm×128 mm.

5.3 Calibration

Extrinsics of Acquisition Modules. We first mount an extra camera below the detector array to take a top view of the acquisition region (Fig. 5-a). Then a black calibration hemisphere with a known radius is placed in the center of the valid volume. Next, each LED is turned on one at a time for the extra camera to capture the reflection of the corresponding source on the hemisphere. The 2D location of the reflected source is estimated with sub-pixel accuracy by fitting a Gaussian. Finally, we jointly optimize the center of the calibration hemisphere and the extrinsic parameters of each acquisition module, by minimizing the reprojection errors. The extrinsic parameters of individual sources and detectors are computed using the corresponding transformations relative to an acquisition module, according to our 3D design.

Measurement Extraction from the Detector Array. First, our camera captures an image with all LEDs on. Then we extract filtered contours of fibers from the image after binarization. We take a 30×30 patch from the original image around the center of each contour and fit a 2D Gaussian. Finally, the measurement of each detector is computed as the product of the corresponding image patch and the pre-fit Gaussian. Please refer to Fig. 5 for a visualization.

Optical Fiber Characteristics. Technically, we need to know how the light intensity is changed after transmitting through each segment of optical fiber. To simplify the calibration, we first view the LED array, LED modules, sources on acquisition modules and optical fibers between the latter two as *generalized sources*, and the remaining parts as *generalized detectors* (Fig. 5-d). Note that generalized sources are linear with respect to a lighting pattern, due to our PWM control mechanism. So we only need to measure the characteristics of the fibers in the generalized detectors. To do so, we place an additional light, whose intensity can be continuously adjusted, in the valid volume. Then, we capture with the light set to different intensities to obtain the response curve for each detector (Fig. 5-e).

Calibrating $\tilde{\mathbf{I}}$. For the j-th column of $\tilde{\mathbf{I}}$, we turn on the j-th LED and fill in the extracted measurements from all detectors, after corrected with the corresponding response curves. We loop over all LEDs to complete $\tilde{\mathbf{I}}$ (Fig. 5-c).

Point Spread Function. We compute 2D point spread functions of our prototype according to a method similar to Chen et al. [2008a]. The estimated full width at half maximum (FWHM) for the horizontal and vertical directions are 1.3 mm and 1.8 mm, respectively.

6 ACQUISITION & RECONSTRUCTION ALGORITHM

6.1 Design Decisions

A naïve way to acquire with our prototype is to loop over all sources with OLAT, to obtain the complete multi-view CT image $\mathbf{I}$. However, this is highly inefficient, because (1) the power of a single source is limited, which results in a long exposure, and (2) the number of sources is large. To capture dynamic scenes, this approach has to sacrifice either sampling density or coverage for acquisition speed [Arai et al. 1999; Mohan et al. 2015; Mozzo et al. 1998; Zang et al. 2018], leading to suboptimal results. Multiplexed acquisition addresses the issue of long exposure by programming the intensities of multiple sources simultaneously. However, existing work is handcrafted and separately considers acquisition and reconstruction; the number of photographs is also the same as the number of sources. As a result, the sampling capability of our prototype is not fully exploited.

Therefore, we propose a novel differentiable framework to map CT acquisition and reconstruction to a deep neural network, to allow joint and automatic optimization of both processes. Here one straightforward way to perform reconstruction is to train from scratch an end-to-end network, which transforms the multiplexed measurements into a 3D volume. But it leads to unsatisfactory results (Fig. 13), due to the huge input/output domain gap. To tackle this issue, we incorporate a differentiable 3D-FBP network as a backend; doing so harnesses the excellent existing domain knowledge to prevent overfitting the training data and generalize better to novel cases [Wang and Liu 2020]. We also carefully design an efficient network architecture to exploit the coherence in measured data.

6.2 Architecture

The input to our network is a multi-view CT image $\mathbf{I}$ of 1920×1920 (which is never directly acquired). The output is a 3D density volume of 32×128×128.

The network consists of three parts. First, an encoder network probes the physical multi-view CT image with optimized lighting patterns, resulting in multiplexed CT images. Next, these raw measurements are converted to a sinogram via a decoder network. Finally, we perform volumetric reconstruction from the sinogram. Please refer to Fig. 10 for an illustration. The three parts will be described in details below (Sec. 6.3-6.5).

6.3 Encoder

The first part of our network contains a linear fc layer, whose weights correspond to the lighting patterns used during acquisition. The patterns help probe the physical multi-view CT images into multiplexed CT images, according to Eq. 5.

6.4 Decoder

This part of the network transforms the multiplexed CT images to a sinogram. Before introducing the details, we first observe that the detector measurements are similar with respect to different sources on a *single* acquisition module, due to the spatial proximity (see Fig. 8 for an example). This motivates our network design, which divides the entire job to 24 tasks on a per-module basis as follows.

Specifically, we partition the sinogram $\mathbf{D}$ into 24 sub-sinograms, each of which corresponds to the pairs between a source on a *particular* acquisition module and one of all detectors in the device. Next, we pretrain 24 autoencoders for each of the modules. Here an autoencoder takes the multi-view CT images as input, and outputs a sub-sinogram corresponding to a particular module. After pretraining, we take the decoder out and denote it as a DecodeNet. Finally, we stack a network on top of 24 DecodeNets, which correspond to 24 acquisition modules in our prototype. This network is denoted as TransferNet, whose job is to convert the input multiplexed CT images into the appropriate latent codes for each DecodeNet. To obtain a complete sinogram, we simply concatenate all 24 output sub-sinograms from all DecodeNets. A graphical illustration is shown in Fig. 10.

6.5 Volumetric Reconstruction

Finally, we pass the decoded sinogram to a differentiable 3D-FBPNet, similar to Wang and Liu [2020], to reconstruct a 3D volume. The 3D-FBPNet contains a 3D FBP module following a 6-layer 3D U-Net network. The filtering weights and back projection weights in 3D FBP module are trainable. Note that other networks that compute a 3D volume out of a sinogram can also be plugged in here.

6.6 Loss Function

Our loss function is defined as $\mathcal{L}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{P}) = \mathcal{L}_{\text{vol}}(\mathbf{x}, \tilde{\mathbf{x}}) + \mathcal{L}_{\text{pat}}(\mathbf{P})$. Its two terms are defined as:

$$\mathcal{L}_{\text{vol}}(\mathbf{x}, \tilde{\mathbf{x}}) = \|\mathbf{x} - \tilde{\mathbf{x}}\|_2^2 + \lambda_{\text{vgg}} \mathcal{L}_{\text{vgg}}(\mathbf{x}, \tilde{\mathbf{x}}), \tag{7}$$

$$\mathcal{L}_{\text{pat}}(\mathbf{P}) = \lambda_{\text{bar}} \mathcal{L}_{\text{bar}}(\mathbf{P}) + \lambda_{\text{bin}} \mathcal{L}_{\text{bin}}(\mathbf{P}). \tag{8}$$

Here $\mathcal{L}_{\text{vol}}$ measures the reconstruction error of predicted density volume $\tilde{\mathbf{x}}$. Its first part measures the difference between network output $\tilde{\mathbf{x}}$ and the label $\mathbf{x}$. The second part $\mathcal{L}_{\text{vgg}}$ computes the perceptual reconstruction quality [Johnson et al. 2016] as:

$$\mathcal{L}_{\text{vgg}}(\mathbf{x}, \tilde{\mathbf{x}}) = \sum_{i=1}^{4} \|\text{VGG}_i(\Phi(\mathbf{x})) - \text{VGG}_i(\Phi(\tilde{x}))\|_1, \tag{9}$$

where VGG_i computes the output from the i-th layer of VGG16 Net, and $\Phi(\mathbf{x}) = \dfrac{\log(1 + \mu\mathbf{x})}{\log(1 + \mu)}$, which is a transform to ensure better data alignment with VGG16 [Santos et al. 2020]. In all experiments, we set $\lambda_{\text{vgg}} = 10^{-5}$ and $\mu = 5000$.

Next, $\mathcal{L}_{\text{pat}}$ is a loss to constrain the optimization of lighting patterns. Here $\mathcal{L}_{\text{bar}}(\mathbf{P}) = \sum[\tanh(\epsilon(\mathbf{P}_{ij} - 1)) - \tanh(\epsilon(\mathbf{P}_{ij} + 1))]$. It acts as a barrier function to ensure the weights in the encoder, which correspond to the lighting patterns, to be in the range of $[-1, 1]$ for physical realization. The final term $\mathcal{L}_{\text{bin}}(\mathbf{P}) = -\sum |\mathbf{P_{ij}}|$, which encourages each LED intensity to be as close to 1 or -1 as possible, since our device can project binary patterns more rapidly than, e.g., 8-bit patterns. In all experiments, we use $\lambda_{\text{bar}} = 1/\#_w$, λ_{bin}=0.1 and ϵ=50, where $\#_w$ is the number of weights in the first fc layer of our encoder.

Note that since the intensities of lighting patterns are constrained to $[-1, 1]$, we can convert each such pattern into two for physical realization: one with all positive values, and the other with all negative values with signs flipped. Throughout this paper, we report the number of physical lighting patterns for consistency.

6.7 Training

According to Sec. 6.4, we first pretrain 24 autoencoders for each of the acquisition modules. Next, we pretrain TransferNet using the latent outputs of the pretrained encoders in the previous step as labels. Finally, we jointly train TransferNet, DecodeNet and 3D-FBPNet in an end-to-end fashion.

Our training dataset is synthetic and contains two types of data. We first follow existing work [Ding et al. 2019; Häggström et al. 2019; Hauptmann et al. 2019; Xu et al. 2018] to randomly generate combinations of predefined primitives with different locations, sizes and orientations (ellipsoids, cubes, spheres, ellipsoidal shells, cubic shells and spherical shells). The density of each primitive is in the range of $(0, 0.5]$. To further increase the diversity, we add simulated sequences of fluid with a resolution of $256 \times 256 \times 256$ by Mantaflow [Thuerey and Pfaff 2018]. Random, non-empty sub-volumes of $32 \times 128 \times 128$ are cropped to fit to our valid volume (refer to Fig. 9 for examples). Next, Pytorch3D [Ravi et al. 2020] is employed to accumulate density along rays of each source-detector pair to synthesize a sinogram (Eq. 2). Finally, we compute multi-view CT images with the help of calibrated $\tilde{\mathbf{I}}$ according to Eq. 3. Note that our approach is fully data-driven, and thus can switch to other training data depending on applications.

To increase robustness in physical experiments, we multiply each of our simulated measurement with a relative Gaussian noise (μ = 1, σ = 10%) during training, to model noise/effects not accounted in our pipeline. Note that σ is determined from an experiment that compares physical measurements with simulated ones under different lighting patterns.

7 RESULTS & DISCUSSIONS

All computation experiments are conducted on a server with dual AMD EPYC 7763 CPUs, 768GB DDR4 memory and 8 NVIDIA GeForce RTX 4090 GPUs. We implement our network with PyTorch and use Adam optimizer for training. PyTorch3D [Ravi et al. 2020] is employed to synthesize multi-view CT images. The learning rate is 10^{-4}, and the batch size is 1. The total training time for the network is 43 hours.

During acquisition, the exposure time for taking a photograph under one lighting pattern is 2.5 ms. Currently, our unoptimized pipeline requires an additional 0.9 ms per pattern to drive both the camera and LED array from PC, resulting in 36.8/8.2 volumes per second using 8/36 lighting patterns, respectively. This extra time can be avoided by driving the entire acquisition directly from our circuit board, which will be equivalent to 400 images/50 volumes per second using 8 lighting patterns. Note that this reaches the maximum speed of our camera. The reconstruction time for a volume is 0.026 s for the number of light patterns ranging from 8 to 36.

We demonstrate the sequences of reconstructed 3D volumes of 5 real dynamic scenes. Selected frames are shown in Fig. 12. Please refer to the accompanying video for animations. All results are rendered with path tracing in Blender. For the erupting vapor scene that is quickly changing, we use #=8 lighting patterns for acquisition. For the remaining scenes that change more slowly, we set # to 36 for a better reconstruction quality. In Fig. 8, we show the multiplexed CT images measured under different lighting patterns. It is clear that multiplexed patterns make better use of the sampling capability of our device, compared with a single light source.

Note that our scanning geometry prioritizes a dense and complete sampling of the projection space over spatial resolution. This differs from previous work [Atcheson et al. 2008; Zang et al. 2020], which typically has a higher spatial resolution and a lower angular one. In reconstruction, our approach relies less on prior information and more on measurements carefully sampled in the projection space, towards the goal of general CT acquisition.

7.1 Comparison

In Fig. 7, we compare our approach with related technique on acquiring and reconstructing rapidly varying vapor. First, OLAT cannot produce a plausible result, due to the extremely long time to finish one scan, during which the scene has already changed. Traditional multiplexing [Zhang et al. 2008] cannot capture fast enough either, as 3840 photographs are needed. For the state-of-the-art high-speed capture technique of Zang et al. [2021], sampling completeness is sacrificed for acquisition speed. The limited angle case employs 8 sources on a single acquisition module, corresponding to an angular coverage of approximately 15°; the sparse view case uses 8 sources evenly distributed over all 24 acquisition modules. Due to the substantially longer exposure time (caused by the limited power of a single LED) and the limited input information, their reconstructions are not as good as ours using an equal number of input photographs. Our system is the only one that can acquire this dynamic phenomenon and reconstruct a 3D volume that resembles the photograph using as few as 8 lighting patterns per volume, which corresponds to a capture time of 0.027 s only.

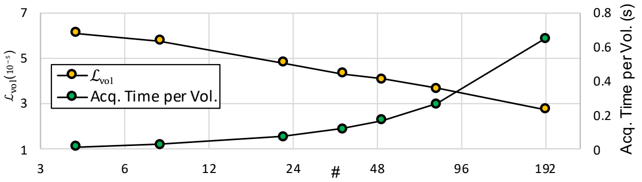


Figure 6: Volumetric reconstruction error $\mathcal{L}_{\text{vol}}$ and acquisition time per volume as a function of the number of lighting patterns #. The horizontal axis is spaced on a log scale.

7.2 Evaluations

Here we mainly focus on analyzing the effectiveness of our current system, as both the hardware and software are jointly optimized towards reconstruction quality. It will be interesting future work to conduct experiments using alternative components, e.g. SART [Andersen and Kak 1984], from existing literature.

We first evaluate the impact of the number of lighting patterns over the reconstruction quality. In Fig. 6, we plot the volumetric reconstruction error $\mathcal{L}_{\text{vol}}$ over a test dataset generated with the method in Sec. 6.7, as a function of #, the lighting pattern number. We also plot the time for capturing a single volume as a function of #. As expected, while the reconstruction error decreases with # (i.e., more captured information), the acquisition time increases. Our framework offers the flexibility to trade between quality and speed, to cater the demands in different applications. In Fig. 11, we test our networks trained with different # on reconstructing a synthetic and a real static scene. Please refer to the figure for qualitative and quantitative evaluations.

Next, we evaluate the impact of different lighting patterns in Fig. 13. We compare with patterns generated with Gaussian noise and randomly selected rows of a Hadamard matrix. In experiments, we fix these alternative patterns and train corresponding networks to adapt to them. At the same number of lighting patterns, we achieve a higher reconstruction quality compared with hand-crafted ones, as our patterns are optimized in conjunction with the reconstruction network towards optimal quality, which fully exploits the capabilities of our prototype.

In the same figure, our network is compared with a naïvely designed end-to-end network, a 3D version of Häggström [2019], which directly maps the measurements to the 3D volume. Their result is not satisfactory. We believe there are two main reasons: (1) the network does not explicitly make use of the coherence with respect to source in a sub-sinogram (Sec. 6.4); (2) existing domain knowledge is not exploited (Sec.6.2).

We evaluate the effect of our trainable 3D-FBPNet in Fig. 13. There are two benefits [Wang and Liu 2020]. First, compared with a pure neural network, 3D-FBPNet does not overfit to the training data. Second, the trainable filter and back projection weights make it possible to further fine-tune the performance. As a result, we achieve a higher reconstruction quality over the vanilla 3D-FBP.

Moreover, we evaluate the impact of the loss term $\mathcal{L}_{\text{bin}}$ (Fig. 13). We train a network without this term to solely focus on volumetric reconstruction error. This results in a more varied distribution of learned source intensities. The reconstruction is improved, despite a longer exposure time due to the longer time it takes for our device to project non-binary patterns. End users of our system can choose to go with binary or non-binary patterns, depending on the changing speed of the scene of interest and the desired reconstruction quality.

Finally, we test if a NeRF-like algorithm (IntraTomo [Zang et al. 2021]) can work well directly on our measurements in the same figure. We employ a 3D version of IntraTomo to minimize the differences between the measurements computed from the current estimation of the volume and its ground-truth, under our lighting patterns. The result is less satisfactory, due to the considerably low number of constraints (i.e., measurements). In comparison, our network implicitly learns a data prior for better reconstructions.

8 LIMITATIONS & FUTURE WORK

Unlike X-ray CT in which attenuation is the main phenomenon, our visible-light system is more influenced by reflections and refractions. It will be interesting to explicitly model these optical effects, in order to acquire interesting fluid motions. In addition, the physical light transport deviates from our single-channel assumption to some degree: our LEDs have a broad spectrum, and the optical fibers transfer the light at different wavelengths differently. We expect that the deviation will be reduced when replacing with a narrow-band LEDs. Moreover, we independently reconstruct each 3D volume in this paper. The quality might be improved, by combining existing work that exploits temporal coherence. It will be also intriguing to apply latest work on neural representation [Rückert et al. 2022] to fine-tune the result against the measurements for a higher quality. And our reconstruction on the erupting vapor suggests that our prototype might be useful to validate fluid simulation techniques. Last but not least, we are excited to push the idea towards the first general dynamic X-ray CT.

ACKNOWLEDGMENTS

We would like to thank anonymous reviewers for their comments. This work is partially supported by NSF China (62022072 & 62227806), Zhejiang Provincial Key R&D Program (2022C01057), the Fundamental Research Funds for the Central Universities, the XPLORER PRIZE and Information Technology Center and State Key Lab of CAD&CG, Zhejiang University.

REFERENCES

Anders H Andersen and Avinash C Kak. 1984. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrasonic imaging* 6, 1 (1984), 81–94.

Rushil Anirudh, Hyojin Kim, Jayaraman J. Thiagarajan, K. Aditya Mohan, Kyle Champley, and Timo Bremer. 2018. Lose the Views: Limited Angle CT Reconstruction via Implicit Sinogram Completion. In *CVPR*.

Y Arai, E Tammisalo, K Iwai, K Hashimoto, and K Shinoda. 1999. Development of a compact computed tomographic apparatus for dental use. *Dentomaxillofacial Radiology* 28, 4 (1999), 245–248.

Bradley Atcheson, Ivo Ihrke, Wolfgang Heidrich, Art Tevs, Derek Bradley, Marcus Magnor, and Hans-Peter Seidel. 2008. Time-resolved 3d capture of non-stationary gas flows. *TOG* 27, 5 (2008), 1–9.

Mojtaba Bemana, Karol Myszkowski, Jeppe Revall Frisvad, Hans-Peter Seidel, and Tobias Ritschel. 2022. Eikonal fields for refractive novel-view synthesis. In *ACM SIGGRAPH 2022 Conference Proceedings*. 1–9.

Guang-Hong Chen, Jie Tang, and Shuai Leng. 2008b. Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Medical physics* 35, 2 (2008), 660–663.

Guang-Hong Chen, Pascal Thériault-Lauzier, Jie Tang, Brian Nett, Shuai Leng, Joseph Zambelli, Zhihua Qi, Nicholas Bevins, Amish Raval, Scott Reeder, et al. 2011. Time-resolved interventional cardiac C-arm cone-beam CT: An application of the PICCS algorithm. *IEEE transactions on medical imaging* 31, 4 (2011), 907–923.

Lingyun Chen, Chris C Shaw, Mustafa C Altunbas, Chao-Jen Lai, and Xinming Liu. 2008a. Spatial resolution properties in cone beam CT: a simulation study. *Medical physics* 35, 2 (2008), 724–734.

Yi Chen, Yan Xi, and Jun Zhao. 2014. A stationary computed tomography system with cylindrically distributed sources and detectors. *Journal of X-ray science and technology* 22 6 (2014), 707–25.

Bruno De Man, Jorge Uribe, Jongduk Baek, Dan Harrison, Zhye Yin, Randy Longtin, Jaydeep Roy, Bill Waters, Colin Wilson, Jonathan Short, et al. 2016. Multisource inverse-geometry CT. Part I. System concept and development. *Medical physics* 43, 8Part1 (2016), 4607–4616.

Guanglei Ding, Yitong Liu, Rui Zhang, and Huolin L Xin. 2019. A joint deep learning model to recover information and reduce artifacts in missing-wedge sinograms for electron tomography and beyond. *Scientific reports* 9, 1 (2019), 1–13.

M-L Eckert, Wolfgang Heidrich, and Nils Thuerey. 2018. Coupled fluid density and motion from single views. In *CGF*, Vol. 37. 47–58.

L. A. Feldkamp, L. C. Davis, and J. W. Kress. 1984. Practical cone-beam algorithm. *J. Opt. Soc. Am. A* 1, 6 (Jun 1984), 612–619. https://doi.org/10.1364/JOSAA.1.000612

Richard Gordon, Robert Bender, and Gabor T Herman. 1970. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of theoretical Biology* 29, 3 (1970), 471–481.

Ida Häggström, C Ross Schmidtlein, Gabriele Campanella, and Thomas J Fuchs. 2019. DeepPET: A deep encoder–decoder network for directly solving the PET image reconstruction inverse problem. *Medical image analysis* 54 (2019), 253–262.

Samuel W Hasinoff and Kiriakos N Kutulakos. 2007. Photo-consistent reconstruction of semitransparent scenes by density-sheet decomposition. *TPAMI* 29, 5 (2007), 870–885.

Andreas Hauptmann, Simon Arridge, Felix Lucka, Vivek Muthurangu, and Jennifer A Steeden. 2019. Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning–proof of concept in congenital heart disease. *Magnetic resonance in medicine* 81, 2 (2019), 1143–1156.

Godfrey Hounsfield. 1973. Computerized transverse axial scanning (tomography): Part 1. Description of system. *The British journal of radiology* 46, 552 (1973), 1016–1022.

Jiang Hsieh. 2003. Computed tomography: principles, design, artifacts, and recent advances. PM344 (2003).

Jing Huang, Yunwan Zhang, Jianhua Ma, Dong Zeng, Zhaoying Bian, Shanzhou Niu, Qianjin Feng, Zhengrong Liang, and Wufan Chen. 2013. Iterative image reconstruction for sparse-view CT using normal-dose image induced total variation prior. *PloS one* 8, 11 (2013), e79709.

Yixing Huang, Xiaolin Huang, Oliver Taubmann, Yan Xia, Viktor Haase, Joachim Hornegger, Guenter Lauritsch, and Andreas Maier. 2017. Restoration of missing data in limited angle tomography based on Helgason–Ludwig consistency conditions. *Biomedical Physics & Engineering Express* 3, 3 (2017), 035015.

Matthias B Hullin, Martin Fuchs, Ivo Ihrke, Hans-Peter Seidel, and Hendrik PA Lensch. 2008. Fluorescent immersion range scanning. *TOG* 27, 3 (2008), 87.

Ivo Ihrke, B Goidluecke, and Marcus Magnor. 2005. Reconstructing the geometry of flowing water. In *ICCV*, Vol. 2. IEEE, 1055–1060.

Ivo Ihrke, Kiriakos N Kutulakos, Hendrik PA Lensch, Marcus Magnor, and Wolfgang Heidrich. 2010. Transparent and specular object reconstruction. In *CGF*, Vol. 29. 2400–2426.

Ivo Ihrke and Marcus Magnor. 2004. Image-based tomographic reconstruction of flames. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*. 365–373.

Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. 2017. Deep convolutional neural network for inverse problems in imaging. *TIP* 26, 9 (2017), 4509–4522.

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*. Springer, 694–711.

Kaizhang Kang, Zimin Chen, Jiaping Wang, Kun Zhou, and Hongzhi Wu. 2018. Efficient Reflectance Capture Using an Autoencoder. *TOG* 37, 4, Article 127 (2018), 10 pages.

Kyeong-Hyeon Kim, Dong-Seok Shin, Sang-Won Kang, Seong-Hee Kang, Tae-Ho Kim, Jin-Beom Chung, Tae Suk Suh, and Dong-Su Kim. 2021. Four-dimensional inverse-geometry computed tomography: a preliminary study. *Physics in Medicine & Biology* 66, 6 (2021), 065028.

Yueting Luo, Derrek Spronk, Yueh Z Lee, Otto Zhou, and Jianping Lu. 2021. Simulation on system configuration for stationary head CT using linear carbon nanotube x-ray source arrays. *Journal of Medical Imaging* 8, 5 (2021), 052114–052114.

K Aditya Mohan, SV Venkatakrishnan, John W Gibbs, Emine Begum Gulsoy, Xianghui Xiao, Marc De Graef, Peter W Voorhees, and Charles A Bouman. 2015. TIMBIR: A method for time-space reconstruction from interlaced views. *IEEE Transactions on Computational Imaging* 1, 2 (2015), 96–111.

P Mozzo, C Procacci, A Tacconi, P Tinazzi Martini, and IA Bergamo Andreis. 1998. A new volumetric CT machine for dental imaging based on the cone-beam technique: preliminary results. *European radiology* 8, 9 (1998), 1558–1564.

Thiago Pereira, Szymon Rusinkiewicz, and Wojciech Matusik. 2014. Computational light routing: 3d printed optical fibers for sensing and display. *TOG* 33, 3 (2014), 1–13.

Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. 2020. Accelerating 3D Deep Learning with PyTorch3D. *arXiv:2007.08501* (2020).

Darius Rückert, Yuanhao Wang, Rui Li, Ramzi Idoughi, and Wolfgang Heidrich. 2022. NeAT: Neural Adaptive Tomography. *TOG* 41, 4 (2022).

Marcel Santana Santos, Tsang Ing Ren, and Nima Khademi Kalantari. 2020. Single image HDR reconstruction using a CNN with masked features and perceptual loss. *TOG* 39, 4 (2020), 80–1.

Taly Gilat Schmidt, Rebecca Fahrig, Norbert J Pelc, and Edward G Solomon. 2004. An inverse-geometry volumetric CT system with a large-area scanned source: A feasibility study. *Medical physics* 31, 9 (2004), 2623–2627.

PR Schwoebel, John M Boone, and Joe Shao. 2014. Studies of a prototype linear stationary x-ray source for tomosynthesis imaging. *Physics in Medicine & Biology* 59, 10 (2014), 2393.

Jan-Jakob Sonke, Lambert Zijp, Peter Remeijer, and Marcel Van Herk. 2005. Respiratory correlated cone beam CT. *Medical physics* 32, 4 (2005), 1176–1186.

Derrek W Spronk. 2021. *Development and Evaluation of a Stationary Head Computed Tomography Scanner*. Ph. D. Dissertation. UNC Chapel Hill.

Nils Thuerey and Tobias Pfaff. 2018. MantaFlow. *http://mantaflow.com*.

Borislav Trifonov, Derek Bradley, and Wolfgang Heidrich. 2006. Tomographic Reconstruction of Transparent Objects. In *EGSR*. 51–60.

Bo Wang and Huafeng Liu. 2020. FBP-Net for direct reconstruction of dynamic PET images. *Physics in Medicine & Biology* 65, 23 (2020), 235008.

Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi. 2018. Deep image-based relighting from optimal sparse samples. *TOG* 37, 4 (2018), 1–13.

Yidi Yao, Liang Li, and Zhiqiang Chen. 2021. A Novel Static CT System: The Design of Triple Planes CT and its Multi-Energy Simulation Results. *Frontiers in Physics* (2021), 213.

Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. 2021. IntraTomo: self-supervised learning-based tomography via sinogram synthesis and prediction. In *ICCV*. 1960–1970.

Guangming Zang, Ramzi Idoughi, Ran Tao, Gilles Lubineau, Peter Wonka, and Wolfgang Heidrich. 2018. Space-Time Tomography for Continuously Deforming Objects. *TOG* 37, 4, Article 100 (Jul 2018), 14 pages.

Guangming Zang, Ramzi Idoughi, Ran Tao, Gilles Lubineau, Peter Wonka, and Wolfgang Heidrich. 2019. Warp-and-project tomography for rapidly deforming objects. *TOG* 38, 4 (2019), 1–13.

Guangming Zang, Ramzi Idoughi, Congli Wang, Anthony Bennett, Jianguo Du, Scott Skeen, William L Roberts, Peter Wonka, and Wolfgang Heidrich. 2020. TomoFluid: reconstructing dynamic fluid from sparse view videos. In *CVPR*. 1870–1879.

J Zhang, R Peng, S Chang, JP Lu, and O Zhou. 2010. Imaging quality assessment of multiplexing x-ray radiography based on multi-beam x-ray source technology. In *Medical Imaging 2010: Physics of Medical Imaging*, Vol. 7622. SPIE, 1450–1457.

J Zhang, G Yang, S Chang, JP Lu, and O Zhou. 2008. Hadamard multiplexing radiography based on carbon nanotube field emission multi-pixel x-ray technology. In *Medical Imaging 2008: Physics of Medical Imaging*, Vol. 6913. SPIE, 628–635.

Tao Zhang, Hewei Gao, Li Zhang, Yuxiang Xing, and Zhiqiang Chen. 2021. 3D image reconstruction for symmetric-geometry CT with linearly distributed source and detector in a stationary configuration. In *Medical Imaging 2021: Physics of Medical Imaging*, Vol. 11595. SPIE, 977–980.

Tao Zhang, Yuxiang Xing, Li Zhang, Xin Jin, Hewei Gao, and Zhiqiang Chen. 2020. Stationary computed tomography with source and detector in linear symmetric geometry: Direct filtered backprojection reconstruction. *Medical physics* 47, 5 (2020), 2222–2236.

Shuyao Zhou, Tianqian Zhu, Kanle Shi, Yazi Li, Wen Zheng, and Junhai Yong. 2021. Review of light field technologies. *Visual Computing for Industry, Biomedicine, and Art* 4, 1 (2021), 29.

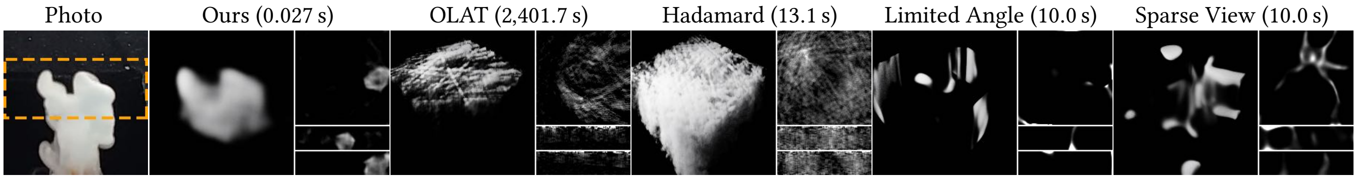


Figure 7: Comparison of different techniques on a dynamic scene. For each approach, we show the rendering result of the reconstructed volume, visualize 3 axis-aligned slices of the volume and report the time for capturing one 3D volume on top. From the left to right: photograph (with the valid volume roughly marked in yellow), our network (# = 8), OLAT (# = 1920), Hadamard multiplexing [Zhang et al. 2008] (# = 3840), and the limited angle/sparse view approach [Zang et al. 2021] (# = 8).

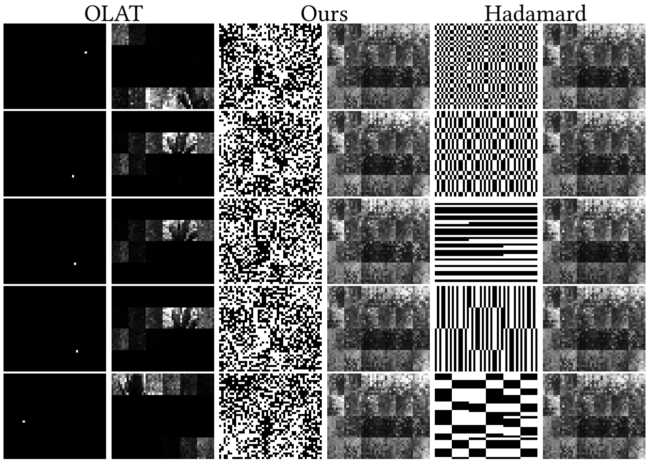


Figure 8: Visualization of different lighting patterns and corresponding measurements of the same static scene. For each pair of columns, the left one visualizes lighting patterns and the right is corresponding measurements. From the left to right: OLAT patterns, ours and Hadamard patterns. Only a subset of all lighting patterns are shown due to the space limit. For OLAT patterns, the sources in the center three rows are spatially adjacent. The intensities of measurements are scaled for a better visualization.

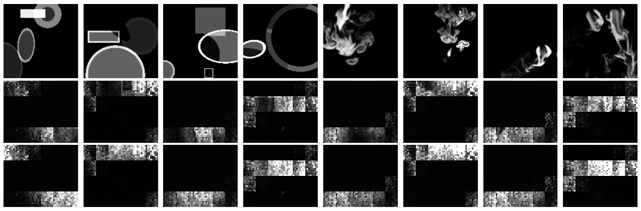


Figure 9: Examples of synthetic training data. Each column shows a randomly generated training sample. The first row is an XY slice of the volume. The second row shows the CT image with respect to a source *j*, simulated with calibrated parameters of our device. The corresponding $\tilde{I}_j$ is in the third row. The first 4 columns are generated from random combinations of volumetric primitives, and the remaining columns are from fluid simulations.

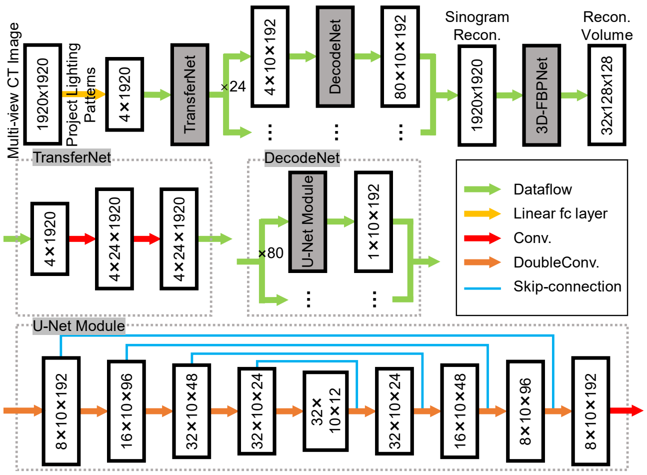


Figure 10: Network architecture. Our autoencoder consists of three parts. The encoder is a simple linear fc layer, which maps the lighting patterns during acquisition. Then a neural network transforms the multiplexed CT image to a corresponding sinogram. Finally, a differentiable 3D-FBP network reconstructs a 3D volume from the sinogram. Please refer to Sec. 6.2 for details.

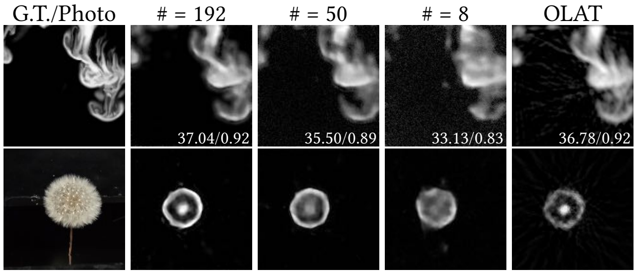


Figure 11: Impact of the number of lighting patterns over reconstruction quality. From the left to right: a slice of simulated flow data/a photograph, reconstruction results from our networks with different numbers of lighting patterns #, and the reconstructions from OLAT measurements with vanilla 3D FBP. Quantitative errors in PSNR/SSIM are reported at the bottom-right corner of corresponding images.

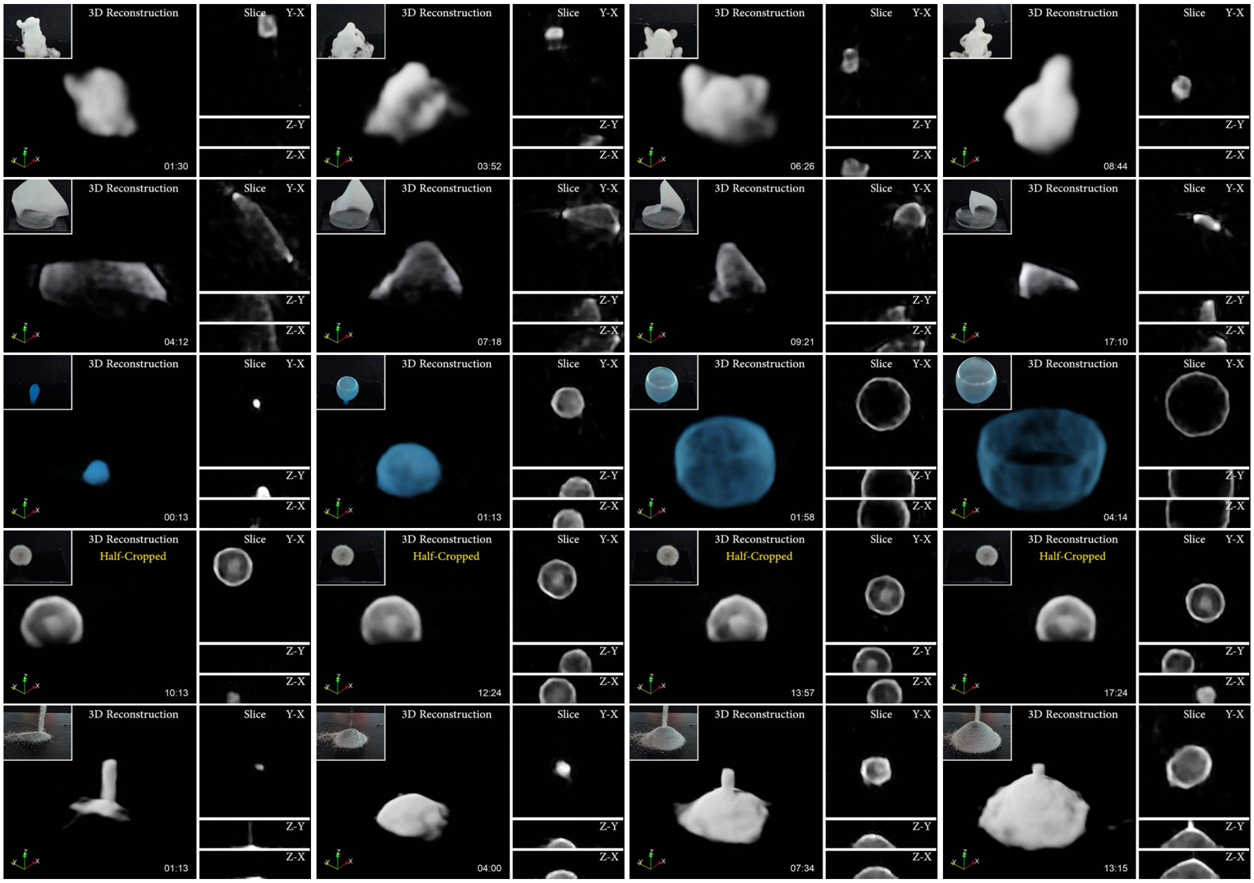


Figure 12: Reconstructions of different dynamic scenes. We visualize a subset of the reconstructed sequences of 5 dynamic scenes: erupting vapor (after putting dry ice in the water), soaking napkin, in-/de-flating balloon, moving dandelion and falling powder. Note that in the 4th row, we cut open the reconstructed volume and render only half of it to better reveal the inner structures of the dandelion. In addition to the visualization of the reconstructed volume, we show the photograph of the scene at the same time on the top-left corner, 3 slices of the reconstructed volume on the right, and the time stamp on the bottom-right corner.

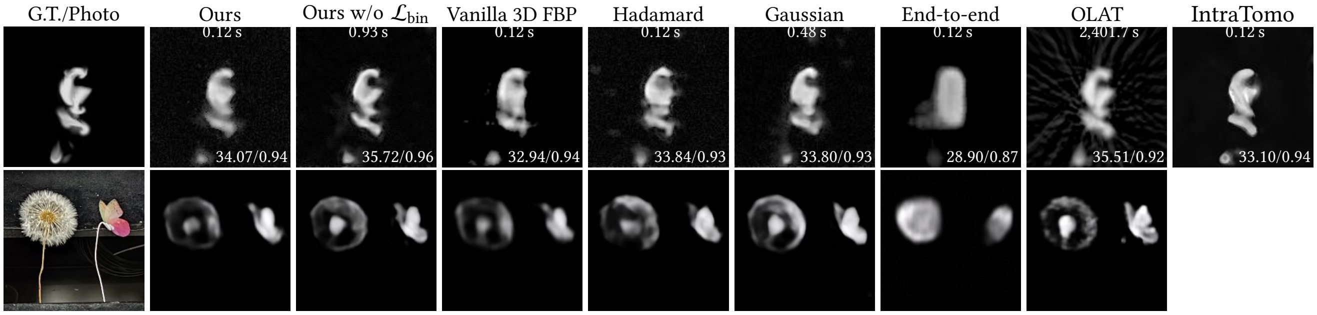


Figure 13: Impact of various factors over reconstruction quality. Here all neural networks use 36 lighting patterns. The first row is a synthetic example, while the second a physical scene. From the left to right: a slice of the ground-truth volume/a photograph, the results using our network, our network trained without $\mathcal{L}_{bin}$, our network with vanilla 3D FBP, our network with fixed, randomly selected Hadamard/Gaussian noise patterns, an end-to-end network [Häggström et al. 2019], direct 3D FBP from OLAT measurments and a 3D version of IntraTomo [Zang et al. 2021]. The time to acquire a volume/quantitative errors (PSNR/SSIM) are reported at the top and bottom-right of corresponding images, respectively.