

# Space-time Tomography for Continuously Deforming Objects

GUANGMING ZANG, RAMZI IDOUGHI, RAN TAO, GILLES LUBINEAU,  
PETER WONKA, and WOLFGANG HEIDRICH,

King Abdullah University of Science And Technology, Saudi Arabia

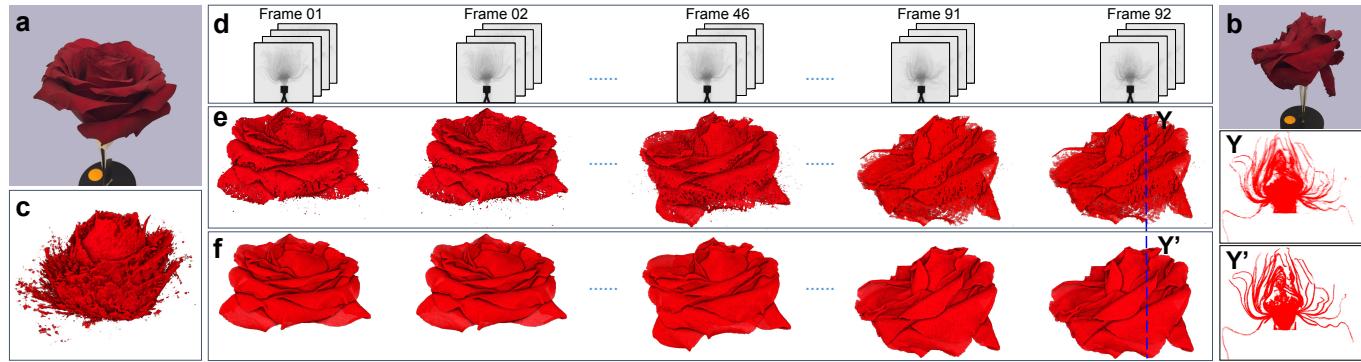


Fig. 1. X-ray tomographic reconstruction of a rose undergoing significant wilting during the scan due to loss of water. Images (a) and (b) show photographs of the rose directly before and directly after the scanning process. Traditional CT reconstruction (c) from all 5520 projections in the scan sequence show significant distortions due to misalignment of features. When grouping the projections into 92 frames of 60 projections each (d), the deformation over each frame becomes negligible, but now the number of projections per frame is insufficient for high-quality reconstruction of the corresponding volumes (e, Y). By comparison, our full space-time reconstruction algorithm yields a time sequence of highly detailed volumes for different time steps (f, Y').

X-ray computed tomography (CT) is a valuable tool for analyzing objects with interesting internal structure or complex geometries that are not accessible with optical means. Unfortunately, tomographic reconstruction of complex shapes requires a multitude (often hundreds or thousands) of projections from different viewpoints. Such a large number of projections can only be acquired in a time-sequential fashion. This significantly limits the ability to use x-ray tomography for either objects that undergo uncontrolled shape change at the time scale of a scan, or else for analyzing dynamic phenomena, where the motion itself is under investigation.

In this work, we present a non-parametric space-time tomographic method for tackling such dynamic settings. Through a combination of a new CT image acquisition strategy, a space-time tomographic image formation model, and an alternating, multi-scale solver, we achieve a general approach that can be used to analyze a wide range of dynamic phenomena. We demonstrate our method with extensive experiments on both real and simulated data.

CCS Concepts: • Computing methodologies → Computer graphics; 3D imaging; Motion capture;

Additional Key Words and Phrases: X-ray computed tomography, 4D reconstruction, Optimization

Authors' address: Guangming Zang; Ramzi Idoughi; Ran Tao; Gilles Lubineau; Peter Wonka; Wolfgang Heidrich, King Abdullah University of Science And Technology, Thuwal, 23955-6900, Saudi Arabia, Email:<firstname>.<lastname>@kaust.edu.sa.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2018 Copyright held by the owner/author(s).

XXXX-XXXX/2018/4-ART

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## ACM Reference Format:

Guangming Zang, Ramzi Idoughi, Ran Tao, Gilles Lubineau, Peter Wonka, and Wolfgang Heidrich. 2018. Space-time Tomography for Continuously Deforming Objects. 1, 1 (April 2018), 14 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 INTRODUCTION

Capture of 3D geometry and dynamic phenomena has long been a topic of interest in computer graphics. While much of the work in this area has been based on traditional computer vision methods with either passive or active illumination [Dou et al. 2016; Józsa et al. 2013], tomographic methods have also played an important role in computer graphics research [Gregson et al. 2014; Ijiri et al. 2014; Reche-Martinez et al. 2004; Stuppy et al. 2003; Zhao et al. 2011].

In engineering and medicine, x-ray computed tomography has become a valuable tool for its ability to reveal internal structures in objects that cannot be sensed by optical means. Tomographic reconstruction in graphics has been primarily focused on visible light tomography [Reche-Martinez et al. 2004; Trifonov et al. 2006; Zheng et al. 2017], although there are a few examples of using x-ray CTs as well [Stuppy et al. 2003; Zhao et al. 2011]. While visible light tomography is inherently limited to transparent materials, it does have the advantage of using regular cameras and light sources, which makes it feasible to construct multi-camera setups that can be used to capture dynamic phenomena such as fluids (e.g. [Atcheson et al. 2008; Elsinga et al. 2006; Gregson et al. 2014, 2012; Hasinoff and Kutulakos 2007; Ihrke and Magnor 2004]). By comparison, x-ray CT is typically limited to a single source and sensor array or a very small number of such sources and arrays, and the required diversity of view points is created through mechanical motion of either the

sample (typical of industrial CTs) or the source-sensor assembly (as in medical CTs). Unfortunately, the need for mechanical motion has so far hindered the ability to use x-ray CTs for imaging dynamic phenomena or continuously changing geometries, except in some very specific setting such as objects that undergo periodic motion (e.g. beating hearts [Mory et al. 2014]).

In this work, we aim to generalize x-ray CT reconstruction to the 4-dimensional case of space-time reconstruction of continuously deforming objects and phenomena. We target the general case of non-periodic motion, although we do assume relatively slow, smooth motion fields. Specifically, we assume that the deformation is negligible when considering a small number of projections (e.g 10-60) acquired in sequence, but that the motion is significant over the time it takes to acquire the hundreds or even thousands of projections needed to scan the whole object at high spatial resolution.

Figure 1 shows an example of such a setting. The scan object is a rose that undergoes significant deformation due to loss of water during the scanning process. This deformation makes it impossible to reconstruct a detailed volume using traditional 3D tomography methods: reconstructions from short sequences of projections are severely under-determined and thus lack detail, while reconstructions from large sequences of frames fail due to misalignment of the geometry. On the other hand, our 4D space-time tomographic reconstruction is able to reconstruct not just a single volume but a full time sequence of highly detailed volumes. It also estimates the deformation fields between these frames which allows our method to be used for a full analysis of the wilting process.

The specific technical contributions of this paper are

- a new CT image acquisition strategy based on low-discrepancy sampling, which improves the distribution of projection angles over time.
- a new image formation model for non-parametric 4D (space-time) tomographic reconstruction, together with the appropriate regularization strategies for tackling the ill-posed inverse problem.
- an alternating, matrix-free, multi-scale optimization algorithm for solving this inverse problem
- an experimental evaluation and demonstration of our approach on a wide range of real data, as well as detailed comparisons on simulated data with ground truth.

With these contributions, we demonstrate a versatile, non-parametric space-time tomographic method that can be used to scan and analyze a wide range of dynamic phenomena, including de-hydration and re-hydration of organic objects, raising dough, fluid flows, etc. We believe this is a major step towards making x-ray CT useful for analyzing internal structures of objects whose shape change during scanning cannot be avoided, as well as objects that undergo deliberate deformations that are to be investigated.

## 2 RELATED WORK

**Scanning of non-rigid and deforming geometries** has a long history in computer graphics and computer vision. Common dynamic surface reconstruction methods use template models [Bradley et al. 2008; De Aguiar et al. 2008], or assume slow deformations that ensure large overlapping regions between successive scans [Li

et al. 2012; Mitra et al. 2007; Wand et al. 2009]. More recently, new pipelines allow for real-time 4D reconstruction of non-rigidly deforming scenes using a single [Innmann et al. 2016] or multiple [Dou et al. 2016] RGB-D cameras. However, such work is based on visible light imaging of surfaces, and is therefore incapable of recovering internal structures of objects, or complicated shapes with occluded visibility.

**Computed tomography in graphics** has been successfully employed for 3D reconstruction and rendering. In order to obtain a volumetric rendering of trees, Reche-Martinez et al. [2004] introduced an approach that combines a volumetric opacity estimation with a view-dependent texturing from different photographs. For fluid imaging, computed tomography allows the 3D reconstruction of flames [Hasinoff and Kutulakos 2007; Ihrke and Magnor 2004], the capture of non-stationary gas flows [Atcheson et al. 2008], the 3D imaging of turbulent fluids [Gregson et al. 2014, 2012], as well as the 3D velocity estimation inside a fluid [Elsinga et al. 2006]. After suspending a transparent object inside a fluid with the same refractive index, Trifonov et al. [2006] realized a tomographic reconstruction of that object. Tomography was also introduced for other applications like 3D displays [Lanman et al. 2011; Wetzstein et al. 2011], where multi-layered systems are employed. X-ray computed tomography has also been used for some applications in graphics, including the utilization of micro CT images to build volumetric appearance models for fabrics [Zhao et al. 2011]. In addition, X-ray tomography has proven to be an excellent tool for flower modeling [Ijiri et al. 2014] or flower structure analysis [Stuppy et al. 2003].

**Numerical methods for tomographic reconstruction** can be grouped into two families: direct methods and iterative methods [Herman 2009; Kak and Slaney 2001]. Direct methods, like the filtered back-projection (FDK) algorithm [Feldkamp et al. 1984], rely on the use of the Radon transform and its inverse. On the other hand, iterative methods propose to solve the discrete formulation of the reconstruction problem [Herman 2009]. This approach involves very large matrices, which makes the memory consumption a key concern. The Simultaneous Algebraic Reconstruction Technique (SART), proposed by Andersen et al. [1984], is one of the most efficient iterative reconstruction methods. It is based on the projection method of Kaczmarz, where the reconstructed volume is updated for each scan view. This approach has modest memory requirements, and yields better reconstruction results than the FDK algorithm when the number of used projections is limited [Chlewicki et al. 2001]. Another advantage of iterative methods is their flexibility in incorporating prior information into the reconstruction process. For example, a total variation (TV) prior has been used for tomography reconstruction [Mory et al. 2012; Sidky and Pan 2008].

**Discrete deformations** refer to the situation where the target can be scanned before and after it undergoes a deformation, but where either no deformation is taking place during the scan, or the time scale of the deformation is very long compared to the scan time, such as in recent work by Kato et al. [2017] on imaging the growing process of plants. This scenario often occurs in mechanical engineering and material science, where the primary goal is to measure

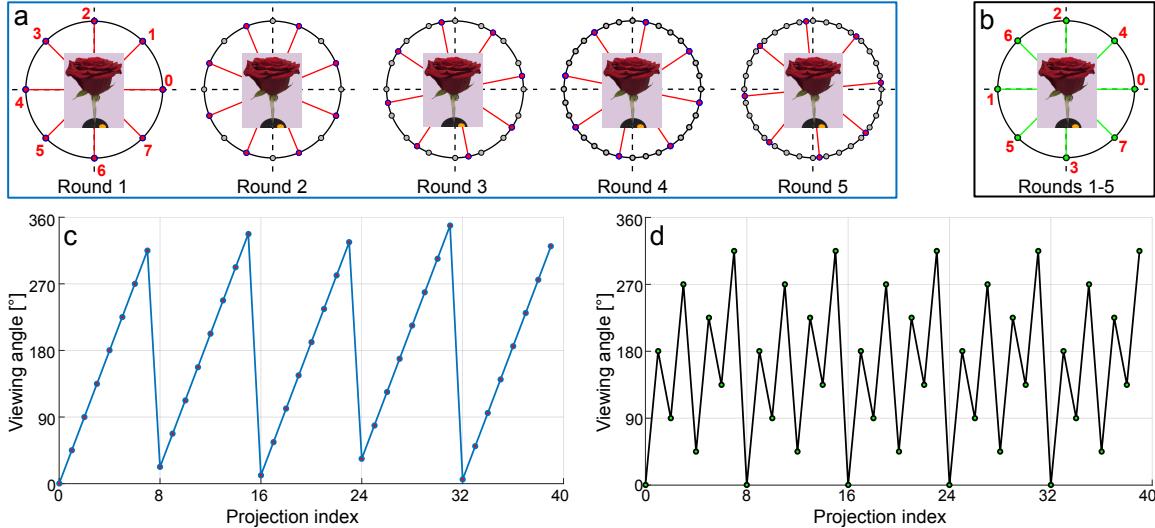


Fig. 2. Comparison between our acquisition strategy (a and c) and interlaced view sampling method [Mohan et al. 2015] (b and d). The parameters taken for this comparison are:  $N_s = N_\theta = 8$  and  $N_{sf} = 4$ . The acquired viewing angles for each round are represented by red circles (a) or green circles (b). The grey circles correspond to the previously acquired angles. The plots (c and d) represent the viewing angles vs. projection index for the first five rounds.

the displacement through of a target structure before and after mechanical loading, in order to retrieve the strain field. To estimate this strain field, Digital Volume Correlation (DVC), essentially a 3D variant of optical flow [Schreier et al. 2009], is widely employed [Bay et al. 1999; Hild et al. 2014; Lachambre et al. 2015; Morgeneyer et al. 2013; Verhulp et al. 2004].

**Periodic deformations.** For medical CT imaging, the deformation time scale is smaller than the scan time, however, the motion is often cyclic, with a period corresponding to a heart beat or respiratory cycle. Based on this observation, several approaches proposed a 4D reconstruction of the heart or thorax [Chen et al. 2012; Sonke et al. 2005], by regrouping the scans into different phases of the cycle and then reconstructing each phase independently, using only the relevant x-ray projections. In order to get a good reconstruction for all phases, the total number of scans should be sufficiently large, which mandates a longer irradiation time and a higher radiation dose for the patient. To address this issue, Mory et al. [2014] proposed a 4D iterative reconstruction with spatial and temporal total variation regularization. This technique is applied on all projections together. The quality of the yielded reconstruction still suffers from the angular undersampling for each cardiac phase. Another recent approach consists of estimating and iteratively compensating the motion that occurs during a cycle, in order to obtain a 4D reconstruction [Taubmann et al. 2015]. The motion estimation is realized using a pair-wise 3D/3D image registration. A major limitation of this approach lies in the low accuracy of the generated 3D volumes representative of each cardiac phase. Finally, one common limitation of all presented 4D CT reconstruction approaches is their reduced applicability to specific deformation phenomenon.

**General deformations.** There has been some limited work on handling general deformations during a tomographic scan, and obtaining a true 4D tomographic reconstruction. A first approach

consists of compensating for known time-dependent transformations of the scanned object [Desbat et al. 2007]. Another approach is to project the volume densities into a finite element basis, and then track deformations of the finite elements over time [Neggers et al. 2015; Roux et al. 2008]. This can be used in digital volume correlation to speed up the distortion estimation between two static frames, but such concepts can also be used as regularizers in continuously deforming objects. Unfortunately, finite elements are very cumbersome to fit to geometries with complex topologies, such as the rose in Figure 1. Recent work like e.g [Leclerc et al. 2015; Taillandier-Thomas et al. 2016] first scans a high-quality template geometry, and then requires only a few projections per deformation state to track the motion. This approach will fail if the object is not stationary long enough to obtain the template. Our method, by comparison, is non-parametric and does not require a template.

Finally, Mohan et al. [2015] combine a new acquisition strategy with a 4D model-based iterative reconstruction (MBIR). This method is perhaps most closely related to ours, as it involves a maximum a posteriori optimization of a penalized likelihood objective function with temporal and spatial smoothness penalties. However, their method requires careful tuning of the acquisition strategy to the range and speed of deformations in a given dataset, making it difficult to scan an arbitrary deformation in a single setting.

### 3 LOW DISCREPANCY VIEW SAMPLING

Computed tomography requires obtaining x-ray projection images of the target under different observation angles. In cone-beam tomography, the most common sampling strategy for obtaining these projections is that the target undergoes a full circle of rotation relative to the x-ray source and sensor, and images are taken in regular intervals. That is, the angles  $\theta_k$  of the  $k^{th}$  projection are given as

$$\theta_k = \theta_0 + k \cdot 2\pi/N, \quad k = 0 \dots N - 1, \quad (1)$$

where  $N$  is the total number of projections, and  $\theta_0$  is some starting angle. In the following, we will call such a scan sequence a *linear scan round*, or just *round* for short.

While this simple strategy works well for static targets, we seek a more uniform distribution of angles vs. time in the case of dynamic targets. Several acquisition strategies were proposed in the literature with this goal in mind. The closest to our approach is the interlaced view sampling approach by Mohan et al. [2015], who proposed to divide the total number of projections  $N$  into  $N_{sf}$  groups consisting of  $N_\theta$  projections each (Figure 2, right). Each group of projections is used to reconstruct a different volume, corresponding to a time frame in an animation of the deforming object.

Within each frame, the projection angles are distributed according to a low discrepancy sequence, which considerably improves the sampling of angle vs. time compared to the simple linear strategy. However, this approach still has several downsides that we would like to address in our work:

- In Mohan et al.’s approach [2015] the tradeoff between  $N_{sf}$  and  $N_\theta$  is a parameter that needs to be set according to the speed of the deformation (many frames with a small number of projections for fast motions vs. few frames with a large number of projections for slow motion). This might mean having to repeat scans until the best tradeoff is found. In comparison, we seek a strategy that is independent of the speed of motion, and only depends on the capabilities of the hardware.
- In interlaced view sampling, the sequence of angles repeats from frame to frame, so that static or slow moving parts of the geometry do not benefit from additional information as the number of frames increases. Since the number of projections per frame can be small, reconstruction quality will suffer as a result. Instead, we seek an approach where all of the  $N$  projection angles are unique in order to extract the maximal amount of information from each additional projection.
- The interlaced view sampling strategy cannot be implemented on all CT hardware without changes to control software. For example, the scanner used in our experiments only supports linear scan rounds as discussed above, with a minimum number of  $N_{\min} = 10$  projections per round. We aim for a sampling strategy that takes such restrictions into account by decomposing naturally into linear scan rounds.

For our space-time tomography method, we propose to split the total sequence of  $N$  projections into  $N_r$  rounds, where each round consists of  $N_{\min}$  projections that are spaced  $\Delta\theta = 2\pi/N_{\min}$  apart. Note that in our approach, the rounds should be as short as possible on a given CT hardware, and the number of rounds should not be confused with the number of frames in the space-time reconstruction (see Section 4.1). We choose the starting angle  $\theta_{0,i}$  for each round  $i$  according to a low discrepancy distribution over the range  $[0, \Delta\theta[$  to uniformly cover all directions over time (see Figure 2, left). Specifically we choose the base-2 Van der Corput sequence [Van der Corput 1935], which for a given index  $i \geq 0$  is defined as

$$h_2(i) = \sum_{j=0}^{\infty} \frac{a_j}{2^{j+1}} \quad \text{with} \quad i = \sum_{j=0}^{\infty} a_j \cdot 2^j, \quad (2)$$

where  $(a_j)$  is the binary representation of  $i$ .  $h_2(i)$  is always in the interval  $[0, 1[$ . The starting angle for the  $i^{\text{th}}$  round is then given as  $\theta_{0,i} = h_2(i) \cdot 2\pi/N_{\min}$ , and the angle of the  $k^{\text{th}}$  projection in the  $i^{\text{th}}$  round is given as

$$\theta_{k,i} = \theta_{0,i} + k \cdot 2\pi/N_{\min}. \quad (3)$$

We note that this strategy simplifies to just the Van der Corput sequence over the full circle if the CT scanner supports “rounds” with just a single image. However, even if the scanner imposes a minimum length for rounds, with our scheme it can still be used to image deformable phenomena, so long as the motion is small over  $N_{\min}$  successive frames. All results in this paper are obtained for  $N_{\min} = 10$ . We also note that the sampling strategy is completely decoupled from the magnitude of the deformation, and that in particular the number of reconstructed frames in the space-time reconstruction (Section 4.1) is independent of the number of rounds and can be chosen post-capture. Moreover, since all projection angles are unique, each additional image provides extra information for a high-quality reconstruction, even in the case of stationary or partially stationary geometry.

## 4 SPACE-TIME TOMOGRAPHY

Given the sampling scheme from the previous section, any sequence of successively acquired projections with at least  $N_{\min}$  images provides a full coverage of the angular space. Furthermore, if that sequence is short enough that the deformation is negligible, then all projections in the sequence are consistent with a specific deformation state of the target and can be thought of as representing a frame of the deforming volume. Short sequences however usually do not contain sufficient information for high-resolution volume reconstructions as the resulting linear system is under determined by several orders of magnitude, depending on the volume resolution and the number of projections in the sequence. We therefore propose a full 4D, space-time image formation model that aggregates information from all  $N$  projections, properly compensated for the deformation to represent each frame in the animation.

In the following, we present the 4D image formation model, before discussing how to solve the resulting optimization problem in Section 4.2.

### 4.1 4D Image Formation Model

**Image formation model.** Consider a 3D object undergoing continuous deformation over time. We represent this object as a continuous 4D density function  $f(x, y, z, t)$ . This density function can be discretized into a sequence of  $N_t$  frames, where each frame is represented by a voxel grid with  $N_v$  voxels. We choose  $N_t \ll N$  and assume that each frame represents an approximately static geometry that can be reconstructed at least at low spatial resolution and low quality by a short sequence of  $N_\theta$  successively captured projections. In our implementation, we choose  $N_\theta = N/N_t$ , however one could also choose larger values of  $N_\theta$ , so that projections contribute to multiple frames.

After regrouping all voxels in each frame into a column vector, the 4D density volume is described by a vector in  $\mathbb{R}^{N_v N_t}$ :  $\mathbf{f} = \{\mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_{N_t-1}\}$ , where  $\mathbf{f}_t \in \mathbb{R}^{N_v}$  represents the 3D volume at

the  $t^{th}$  time step. With these assumptions, the basic tomography problem on the sequence of frames can be expressed as

$$\underbrace{\begin{pmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_t & \\ & & & \ddots & \\ & & & & \mathbf{A}_{N_t} \end{pmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{pmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_t \\ \vdots \\ \mathbf{f}_{N_t} \end{pmatrix}}_{\mathbf{f}} = \underbrace{\begin{pmatrix} \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_t \\ \vdots \\ \mathbf{p}_{N_t} \end{pmatrix}}_{\mathbf{p}}, \quad (4)$$

where  $\mathbf{A}_t \in \mathbb{R}^{N_\theta M \times N_v}$  is the matrix that models the Radon transform operator for the  $N_\theta$  projections contributing to the  $t^{th}$  frame.  $M$  is the number of pixels for a projection image.  $\mathbf{p}_t \in \mathbb{R}^{N_\theta M}$  is the vector of the measured data (sinogram).

**Joint optimization framework.** In general, this computed tomography reconstruction is an ill-posed inverse problem, especially since only few projections are used for the reconstruction at each frame (for the results in this paper,  $N_\theta$  ranges from 10 to 60). This problem can be alleviated by aggregating information across frames, using the regularized optimization framework shown in Equation 5. This framework jointly reconstructs the 4D volume  $\mathbf{f}$  as well as the deformation fields  $\mathbf{u} = \{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{N_t-2}\}$  between successive time frames. Recently, this joint motion estimation and image reconstruction approach has successfully been used in other applications, like in dynamic image reconstruction [Burger et al. 2018] and fluid flow estimation [Gregson et al. 2014; Xiong et al. 2017].

$$\begin{aligned} (\mathbf{f}^*, \mathbf{u}^*) = & \underset{\mathbf{f}, \mathbf{u}}{\operatorname{argmin}} \sum_{t=1}^{N_t} \|\mathbf{A}_t \mathbf{f}_t - \mathbf{p}_t\|_2^2 \\ & + \kappa_1 \sum_{t=1}^{N_t-1} \|\nabla_T \mathbf{f}_t + \nabla_S \mathbf{f}_t \cdot \mathbf{u}_t\|_1 \\ & + \kappa_2 \sum_{t=1}^{N_t} \|\nabla_S \mathbf{f}_t\|_{H_\epsilon} + \kappa_3 \sum_{t=1}^{N_t} \|\nabla_T \mathbf{f}_t\|_2^2 \\ & + \kappa_4 \sum_{t=1}^{N_t-1} \sum_{i=x,y,z} \|\nabla_S \mathbf{u}_{t,i}\|_{H_\tau} \end{aligned} \quad (5)$$

Here, the operators  $\nabla_S$  and  $\nabla_T$  represent respectively the spatial and temporal discrete gradient,  $\mathbf{u}_t$  represents the deformation vector field at time  $t$ .

The first term in Equation 5 corresponds to the least-square data fitting term, derived from Equation 4. The second term is an  $L_1$  volume correlation term, similar to a 3D version of the Horn-Schunck style brightness constancy term [Horn and Schunck 1981] in optical flow. As discussed in Section 4.2 below, we use a multi-scale version of optical flow [Meinhardt-Llopis et al. 2013] which does not have a closed form representation, but is capable of dealing with larger deformations than basic Horn-Schunck.

The remaining terms are regularizers for the volume and the deformation field, respectively. On the third line of Equation 5 are two terms for regularizing the spatial and the time dimensions of the reconstructed volumes. For the spatial regularization we use the Huber penalty [Huber 2011] on the spatial gradient over

the reconstructed volumes, which is parameterized by a positive parameter  $\epsilon$ . The idea of this term is similar to a total variation (TV) prior, except that the Huber penalty prevents the usual staircase artifacts introduced by a TV prior by not penalizing small gradients. Finally, on the last line we introduce a spatial regularization of the deformation field, also utilizing the Huber penalty.

## 4.2 Solver

Solving the joint optimization problem in Equation 5 is challenging, especially because of the optical flow prior. Indeed, this term is neither linear nor convex. Moreover, it involves both variables  $\mathbf{f}$  and  $\mathbf{u}$ . We address these issues by solving the optimization problem in an alternating way as shown in Algorithm 1.

We first initialize the 4D volume ( $\mathbf{f}$ ) by applying a variant of the SART algorithm [Andersen and Kak 1984] independently for each frame, where the individual projections are processed in a random order to improve convergence [Trifonov et al. 2006]. Then, our algorithm alternates between estimating the deformation field ( $\mathbf{u}$ ) for each pair of successive frames, and refining the reconstruction of  $\mathbf{f}$ .

---

### Algorithm 1 Space-Time tomography

---

```

1: procedure ST-TOMOGRAPHY( $F_f, G_f, F_u, G_u, \rho$ )
2:   Initialize  $\mathbf{u} = \mathbf{0}$ , and  $\mathbf{f}$  by using the SART algorithm
3:   for  $i$  from 1 to maximum outer iteration do
4:     // generate multi-scale data
5:      $\mathbf{f}^1 \leftarrow \mathbf{f}$ ,  $\mathbf{u}^1 \leftarrow \mathbf{u}$ 
6:     for  $s$  from 1 to  $N_{scales} - 1$  do
7:        $\mathbf{f}^{s+1} \leftarrow \downarrow \mathbf{f}^s$ 
8:        $\mathbf{u}^{s+1} \leftarrow \rho \downarrow \mathbf{u}^s$ 
9:     end for
10:    // update deformation field variables
11:    // from the coarsest scale to the finest
12:    for  $s$  from  $N_{scales} - 1$  to 1 do
13:       $\mathbf{u}^s \leftarrow \text{ESTIMATEDEFORMATIONS}(F_{f^s}(\mathbf{u}^s), G_{f^s}(\mathbf{u}^s))$ 
14:       $\mathbf{u}^{s-1} \leftarrow \frac{1}{\rho} \uparrow \mathbf{u}^s$ 
15:    end for
16:     $\mathbf{u} \leftarrow \mathbf{u}^1$ 
17:    // update 4D volume variables
18:     $\mathbf{f} \leftarrow \text{RECONSTRUCTVOLMEDENSITY}(F_u, G_u)$ 
19:  end for
20:  return  $\mathbf{f}$  and  $\mathbf{u}$ 
23: end procedure

```

---

In addition, our alternating joint optimization framework, given in Algorithm 1, follows a multi-scale coarse-to-fine scheme [Meinhardt-Llopis et al. 2013]. The deformation field is estimated first for the coarsest scale, then the estimation is propagated step-by-step to the finest levels. For each scale ( $s$ ), the deformation field is initialized by up-sampling the previous result obtained with a coarser scale. This multi-scale strategy aims to take into account the large deformations in the optical flow prior. In Algorithm 1, the finest and the coarsest scales are respectively obtained for  $s = 1$  and  $s = N_{scales}$ . The

operators  $\uparrow$  and  $\downarrow$  respectively perform cubic up- and down-sampling by a factor of  $\rho$ , and the down-sampling operator also includes a Gaussian smoothing.

Detailed explanations of the deformation field estimation and the volume reconstruction and the associated operators  $F_f$ ,  $G_f$ ,  $F_u$ , and  $G_u$  are provided in the following.

**4.2.1 Deformation field estimation.** For estimating the deformation field between two volumes, we adapt the multi-scale optical flow approach introduced by Meinhardt-Llopis et al. [2013] to include additional regularizers and to work on 3D volumes rather than 2D images.

At each scale  $s$ , we estimate the residual deformation between frame  $f_t$  and a backwards warped version of frame  $f_{t+1}$ , where the warping is performed using the previously estimated low resolution deformation field  $u_t^{s+1}$ . We denote this backward-warped frame as  $\text{warp}(f_{t+1}, \uparrow u_t^{s+1})$ , and define the following function

$$\begin{aligned} G_f(u^s) = & \kappa_1 \sum_{t=1}^{N_t-1} \|(\text{warp}(f_{t+1}, \uparrow u_t^{s+1}) - f_t) \\ & + \nabla_S \text{warp}(f_{t+1}, \uparrow u_t^{s+1}) \cdot (u_t^s - \uparrow u_t^{s+1})\|_1 \quad (6) \\ & + \kappa_4 \sum_{t=1}^{N_t-1} \sum_{i=x,y,z} \|\nabla_S u_{t,i}^s\|_{H_\tau} \end{aligned}$$

Based on Equation 5, the optimization problem for each scale is then given as

$$u^{s,*} = \underset{u^s}{\operatorname{argmin}} \quad G_f(u^s). \quad (7)$$

This approach corresponds to a first-order Taylor approximation of the non-linear warping function, and was first proposed by Meinhardt-Llopis [2013]; we refer to their work for a detailed derivation and discussion. In order to ensure high accuracy results, a BFEC method [Kim et al. 2005] was selected for all our warping operations.

---

**Algorithm 2** CP-based method for deformation field estimation

---

```

1: procedure ESTIMATEDEFORMATIONS( $F_f, G_f$ )
2:   while not converged do
3:     // update slack variable
4:      $w^{j+1} \leftarrow \text{prox}_{\lambda_1 G_f^*}(w^j + \lambda_1 K_f \bar{u}^j)$ 
5:     // update deformation field
6:      $u^{s,j+1} \leftarrow \text{prox}_{\mu_1 F_f}(u^{s,j} - \mu_1 K_f^T w^{j+1})$ 
7:     // update dual variable
8:      $\bar{u}^{j+1} \leftarrow 2 \cdot u^{s,j+1} - u^{s,j}$ 
9:   end while
10:  return  $u^s$ 
11: end procedure

```

---

Due to the presence of the L1-norm and the Huber penalty in Equation 7 we reformulate this optimization in the first-order primal-dual framework (CP algorithm), introduced by Chambolle and Pock [2011]. The strategy proposed by this algorithm, consists in splitting the optimization problem into different sub-problems that are solved independently in the form of proximal operators. The pseudo-code

shown in Algorithm 2, summarizes the CP algorithm used to solve the Equation 7, where  $w$  and  $\bar{u}$  are respectively the slack and the dual variables.  $\text{prox}_{\lambda_1 G_f^*}$  and  $\text{prox}_{\mu_1 F_f}$  are the proximal operators, based on the functions  $F_f(u) = 0$  and  $G_f$  as defined above. Since  $F_f(u)$  is a constant function, its proximal operator is simply the identity. The derivation of  $\text{prox}_{\lambda_1 G_f^*}$  is given in the Supplemental Material. The operator  $K_f$  is defined as follows:

$$K_f = \begin{pmatrix} \nabla_S^T & 0 & 0 & \nabla_x \text{warp}(f_{t+1}, \uparrow u_t^{s+1}) \\ 0 & \nabla_S^T & 0 & \nabla_y \text{warp}(f_{t+1}, \uparrow u_t^{s+1}) \\ 0 & 0 & \nabla_S^T & \nabla_z \text{warp}(f_{t+1}, \uparrow u_t^{s+1}) \end{pmatrix}^T \quad (8)$$

**4.2.2 Volumetric reconstruction.** After the deformation field estimation, the 4D volume density ( $f$ ) is reconstructed using the optimization framework given in the Equation 9, which follows directly from Equation 5.

$$\begin{aligned} f^* = & \underset{f}{\operatorname{argmin}} \quad \sum_{t=1}^{N_t} \|A_t f_t - p_t\|_2^2 \\ & + \kappa_1 \sum_{t=1}^{N_t-1} \|\nabla_T f_t + \nabla_S f_t \cdot u_t\|_1 \\ & + \kappa_2 \sum_{t=1}^{N_t} \|\nabla_S f_t\|_{H_\epsilon} + \kappa_3 \sum_{t=1}^{N_t} \|\nabla_T f_t\|_2^2 \end{aligned} \quad (9)$$

In order to deal with large-scale deformations, the implementation of the volume update in practice uses volume warping instead of the Horn-Schunck-style energy term, just as in the deformation estimation (Section 4.2.1). That is, we approximate

$$\nabla_T f_t + \nabla_S f_t \cdot u_t \approx \text{warp}(f_{t+1}, u_t) - f_t. \quad (10)$$

Including this term in the volume reconstruction is what allows for aggregation of information from projections across frames, and is the reason for improved volume reconstruction quality.

---

**Algorithm 3** CP-based method for tomographic reconstruction

---

```

1: procedure RECONSTRUCTVOLUME(DENSITY( $F_u, G_u$ )
2:   while not converged do
3:     // update slack variable
4:      $g^{j+1} \leftarrow \text{prox}_{\lambda_2 G_u^*}(g^j + \lambda_2 K_u \bar{f}^j)$ 
5:     // update volume variable
6:      $f^{j+1} \leftarrow \text{prox}_{\mu_2 F_u}(f^j - \mu_2 K_u^T g^{j+1})$ 
7:     // update dual variable
8:      $\bar{f}^{j+1} \leftarrow 2 \cdot f^{j+1} - f^j$ 
9:   end while
10:  return  $f$ 
11: end procedure

```

---

As before, a first-order primal-dual algorithm is employed, since Equation 9 also contains the Huber penalty and the L1-norm. The used scheme for solving the Equation 9 is provided in Algorithm 3, where  $g$  and  $\bar{f}$  are respectively the slack and the dual variables.  $\text{prox}_{\lambda_2 G_u^*}$  and  $\text{prox}_{\mu_2 F_u}$  are the proximal operators for the volume reconstruction, their derivations are provided in the Supplemental Material. The function  $F_u$  is defined as the first term of the objective

function in Equation 9, while  $G_u$  consists of the remaining terms of the same objective function. The operator  $K_u$  is given by

$$K_u = (\nabla_S, \nabla_T, W)^T. \quad (11)$$

where  $W$  is a warping operator representing the volume warping term  $\text{warp}(\cdot, \cdot)$ . Note that for the sake of notational simplicity we are abusing the linear operator notation to also describe the non-linear  $W$ . Specifically, we denote the adjoint of the image warping operator as  $W^T$ . If  $W$  is the backward warp from frame  $i+1$  to frame  $i$ , the  $W^T$  is the forward warp from frame  $i-1$  to frame  $i$ .

The SART algorithm is applied as a solver for the proximal operator of  $F_u(f)$  instead of the traditional conjugated gradient (CG) method, since the CG solver usually requires good preconditioners to be stable, and is generally more sensitive to measurement noise [Aly et al. 2016].

## 5 EXPERIMENTS AND RESULTS

In this section, we first quantitatively assess our reconstruction method using a simulated deforming object. Then, we demonstrate its applicability to six selected deformation phenomena: flow of a high-viscosity fluid, wilting of a rose, re-hydration of a dried mushroom, rising of a dough, magnification of lentil/lupin seeds soaked in water, and dissolution of crystal sugar in water.

### 5.1 Quantitative evaluation with synthetic data

We perform experiments on a synthetic dataset to obtain quantitative results for comparisons with alternative acquisition schemes and reconstruction methods. The volume is based on a high-quality CT scan of a static copper foam, depicted in Figure 3, on which we impose a synthetic deformation. The volume resolution in these experiments was  $125 \times 250 \times 125$ .

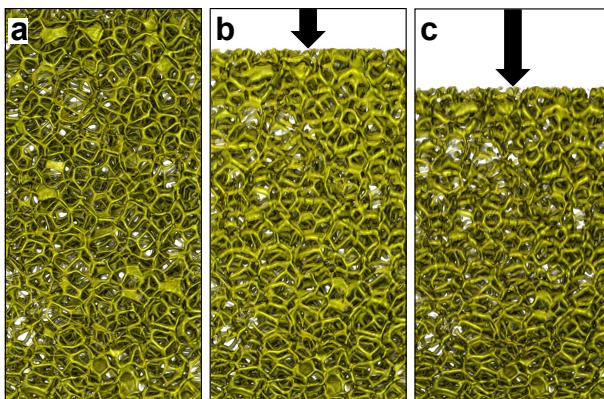


Fig. 3. Synthetic deformation for a copper foam volume with an initial size of  $125 \times 250 \times 125$ . (a) Initial volume obtained with real CT acquisition. The deformation is a uniform compression, where the top edge moves downwards at a constant speed, and the bottom edge is stationary, resulting in a linear velocity gradient throughout the volume. Frames (b) and (c) frames 150 and 300 of this sequence.

The deformation in these experiments is a uniform compression in the vertical direction (Figure 3), where the top edge moves downwards by 0.2 voxels between successive projection images. While

this is a very simple, non-physical motion, it does allow us to observe a range of different velocities in a single experiment, with  $v = 0.2$  voxels/ $\Delta t$  at the top edge,  $v = 0$  at the bottom edge, and a linear ramp of velocities inbetween. This allows us to collect quality statistics for different horizontal slabs of the volume to analyze the impact of deformation velocity on the quality of reconstructions for different algorithmic choices and parameter settings. In the following comparisons, we consider the volume generated at the mid-time frame as the ground truth for that time frame.

**Impact of the acquisition strategy.** First, we study the impact of the view sampling strategy on our space-time reconstruction method. From a total number of  $N = 160$  projections we reconstruct  $N_t = 5$  frames (i.e. we use  $N_\theta = 32$  projections per frame). We test three different view sampling strategies: (a) a single **linear** round with  $N = 160$  projections, (b) the strategy of Mohan et al. [2015] with  $N_\theta = 32$ , and (c) **our** low discrepancy sampling, where we use  $N_r = 5$  rounds of  $N_{\min} = 32$  projections each, since a power of two value for  $N_{\min}$  makes our method more directly comparable to Mohan et al.'s method. On all datasets, we attempt our full space-time tomographic reconstruction method.

Table 1 shows the numerical results of this experiment separately for five different regions of the volume, corresponding to different velocities of the deformation. Two numerical metrics are collected on the data, namely: Peak SNR (PSNR), and SSIM index. For SSIM comparisons, a binary mask is added to allow us to focus on the geometric comparison. As expected, all the metrics indicate improved reconstruction with decreasing speed of deformation for all sampling strategies. Our low discrepancy strategy offers the best performance at all speeds, with the difference being larger on the fastest moving parts, where the advantages of a good sample distribution in both in space *and* time are the most evident.

Region	Metric	Linear	Mohan	Our
Top (fastest)	PSNR	18.34	23.83	28.08
	SSIM	0.53	0.75	0.84
Center (medium)	PSNR	21.54	25.27	29.15
	SSIM	0.68	0.80	0.88
Bottom (slowest)	PSNR	25.43	28.17	31.45
	SSIM	0.78	0.87	0.92
	PSNR	28.32	31.44	33.57
	SSIM	0.84	0.93	0.96
	PSNR	30.28	33.23	34.52
	SSIM	0.89	0.96	0.97

Table 1. Interaction of the sampling strategy with our reconstruction method. PSNR [dB], and SSIM are applied.

**Comparison of different reconstruction algorithms.** While the previous experiment used our reconstruction method with different sampling strategies, we now compare different reconstruction methods and study in detail the impact of each prior on the reconstructed quality. The methods we compared are:

- plain SART,

- Rudin-Osher-Fatemi total variation denoising model [Gentreuer 2012], implemented by solving it in a primal dual scheme [Chambolle and Pock 2011], and using SART as the solver for data term (SART-ROF),
- SART-Huber (SART-H), where the TV term is replaced by the corresponding Huber variant,
- SART-Huber&Temporal smoothing prior (SART-H&T), but no optical flow warping, and
- our full method (Ours).

Considering the structural complexity and relative motion of this simulated data, we choose  $N_\theta$  as 30, thus we have 10 frames of volumes in total, the numeric results for each metric shown in the Table 2 are averaged over all 10 frames. As shown in Table 2, the three methods that rely solely on spatial reconstruction (SART, SART-ROF, and SART-H) perform significantly worse than the two methods with using some form of temporal regularization (SART-H&T and Ours). However, even within that latter group, the advantage of our full model including optical flow warping is evident. Analyzing the results for the individual horizontal slices, we note that the errors get worse from bottom to top, but our method degrades much slower than the comparison approaches.

Region	Metric	SART	ROF	H	H&T	Ours
Top (fastest)	PSNR	19.53	21.83	21.88	23.58	28.15
	SSIM	0.65	0.71	0.71	0.76	0.83
Center (medium)	PSNR	21.79	23.13	23.17	24.17	29.03
	SSIM	0.72	0.77	0.77	0.82	0.88
Bottom (slowest)	PSNR	26.67	27.53	27.58	28.69	31.96
	SSIM	0.83	0.86	0.86	0.89	0.93
	PSNR	29.23	30.09	30.07	32.32	33.92
	SSIM	0.89	0.91	0.92	0.93	0.96
	PSNR	30.65	31.07	31.07	32.81	34.35
	SSIM	0.91	0.93	0.93	0.94	0.96

Table 2. Calculated PSNR [dB], and SSIM for different reconstruction methods: SART, Rudin-Osher-Fatemi (SART+TV), SART+H(uber), SART+H&T(emporal) smoothing, and our full method.

**Failure case determination.** In order to explore the failure cases of our method, we conduct a compression experiment with larger velocities than before (0.9 voxels/ $\Delta t$  at the top edge of the volume).

Metric	(0.0-0.1)	(0.1-0.3)	(0.3-0.5)	(0.5-0.7)	(0.7-0.9)
PSNR	33.77	27.15	22.52	18.39	15.14
SSIM	0.94	0.81	0.72	0.59	0.48

Table 3. Calculated PSNR [dB], and SSIM for different compression velocities [voxels/ $\Delta t$ ].

The Table 3 shows how the quality of our reconstruction degrades with faster velocities. For example, if the deformation is larger than 0.5 voxels between successive projections, the PSNR drops below 20. This corresponds to approximately 15 voxels for one time frame

of our reconstruction, since we are using 30 projections to reconstruct each time frame. Another experiment was done by rotating the volume between successive projections with a fixed angle  $\phi$ . Different values of  $\phi$  were tested (see Supplemental Material), and demonstrate that the method starts breaking down around values of  $\phi > 0.3^\circ$ . It must be emphasized that these results cannot necessarily be generalized to arbitrary data, since the performance of our method also depends on the amount of local volume structure. However, overall our synthetic experiments demonstrate that the performance of our method deteriorates gracefully as the speed of motion increases.

## 5.2 Qualitative evaluation with real scans

To demonstrate the versatility of our proposed approach in studying a variety of dynamic phenomena, we apply it to six real experiments comprising very different types of objects and deformations: (a) a **rose** undergoing significant wilting during the 9.5 hour scan process. The wilting was further accelerated by adding salt to the bottom of the stem. (b) Flow of a high viscosity transparent **fluid** over a 3D-printed mold. This fluid, with 20 million times the viscosity of water, has several included air bubbles that move and pop as the fluid fills the mold. (c) Re-hydration of dried black **mushrooms** from a melting ice cube. (d) Rising **dough**, made from flour and yeast. A hazelnut is placed inside this dough, in order to have additional internal structures. (e) Hydration of a mixture of lentil and lupin **seeds** soaked in water. (f) Dissolution of crystal **sugar** inside water.

*Scanning and Optimization Parameters.* All of these phenomena were captured on the same CT device, a Nikon XT H 225, where the acceleration voltage of the x-ray tube was in the range of [93 kV, 175 kV], depending on the dataset. The detector of this device has a resolution of  $1910 \times 1524$  pixels, with a pixel size of  $0.127 \times 0.127$  mm. Given the different nature of these experiments, their acquisition parameters (capture time, number of frames) are very different. Table 4 provides the value of these parameters for all datasets.

In our optimization framework some parameters are common for the six datasets. The weights for the Huber penalty priors are set respectively to:  $\kappa_2 = 0.05$  (In practice,  $\kappa_2$  ranging from 0.01 to 0.1 yields accurate results) and  $\kappa_4 = 1.2$ . For the multi-scale scheme we choose  $N_{scales} = 3$ ,  $\rho = 0.5$ ,  $\sigma = 0.65$  (Algorithm 1). Otherwise, the scales used in the proximal operators are set to:  $\lambda_1 = \mu_1 = 0.3$  for the deformation estimation problem (Algorithm 2), and  $\lambda_2 = 1.0$ ,  $\mu_2 = 0.1$  for the volume density reconstruction (Algorithm 3). In the inner loops we used two iterations for the SART algorithm and one iteration for the deformation field reconstruction. Other parameters are specific for each dataset (see Table 4 for these parameters).  $\kappa_3$  controls the temporal coherence of each dataset, and usually to get good results  $\kappa_3$  should be in the range 0.1 to 0.8. Our algorithms were implemented in C++, and were run on a computer with a dual-core 3.00GHz Intel Xeon E5-2687W processor and 512 GB of RAM. The deformation estimation ( $\mathbf{u}$ -problem) and the volume reconstruction ( $\mathbf{f}$ -problem) run times for one complete iteration are also given in the Table 4.

The deformation scale and the complexity of the structures are very different from one dataset to another. Consequently, we choose

Dataset	Cap. time [h:mm]	$N_t$	$N_\theta$	Volume size	Vox. pitch [mm]	$\kappa_1$	$\kappa_3$	# Outer iterations	u-problem [h:mm]	f-problem [h:mm]
rose	9:32	92	60	510×384×456	0.20	0.2	0.3	6	4:48	3:19
fluid	2:12	152	10	384×304×384	0.24	0.2	0.5	6	4:34	1:21
mushrooms	6:05	200	20	240×190×240	0.34	0.1	0.1	6	5:50	1:07
dough	4:27	132	20	276×221×271	0.20	0.1	0.1	8	4:50	0:51
seeds	5:12	60	50	508×332×506	0.17	0.1	0.1	8	2:38	2:02
sugar	0:43	50	10	475×380×475	0.17	0.1	0.1	8	2:09	0:18

Table 4. Parameters used in the acquisition and for the optimization, and total run times per sub-problem (summed over all outer iterations) for each of the six datasets. The parameters  $\kappa_2 = 0.05$  and  $\kappa_4 = 1.2$  were fixed for all datasets.

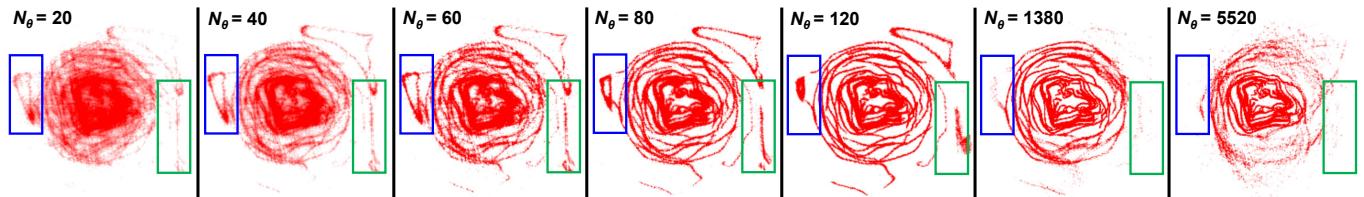


Fig. 4. Comparison of rose slices obtained after SART reconstruction using a different number of projections ( $N_\theta$ ). The blue and green rectangles highlight the regions where the differences are the most prominent.

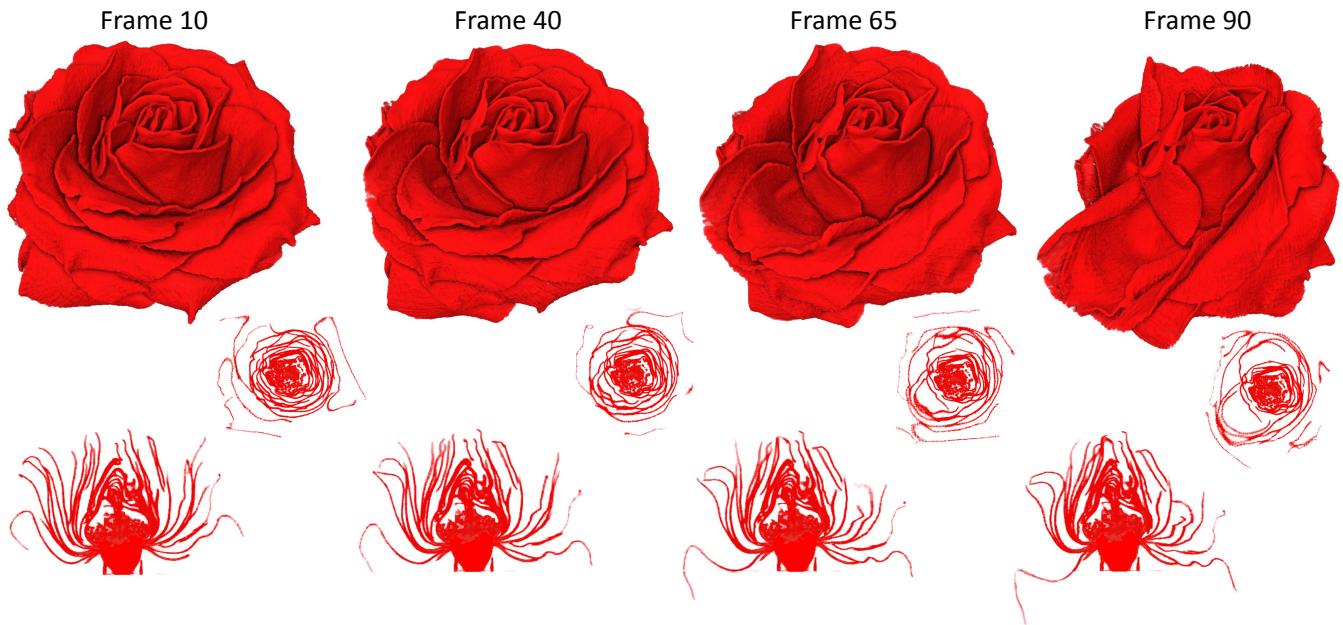


Fig. 5. Reconstruction obtained for the wilting rose dataset. A 3D rendering of the rose is given at four different time frames (first row). The second and the third rows show respectively a top and a side slices for the same time frames.

different number of projections by time frame ( $N_\theta$ ) in our reconstructions (see Table 4). This parameter is of great importance, and has to be set carefully in order to obtain accurate reconstructions. Intuitively, it should be as high as possible to have a better reconstruction of complex structures in each individual frame, which can then be used for optical flow tracking. On the other hand,  $N_\theta$  should

be small enough to avoid the "motion blur" caused by the deformations. In other words, there is a trade-off between spatial and temporal resolution that requires parameter selection, but unlike in the work of Mohan et al. [Mohan et al. 2015], this parameter choice is strictly post-capture and can be informed by visually analyzing the amount of motion in the raw projection images.

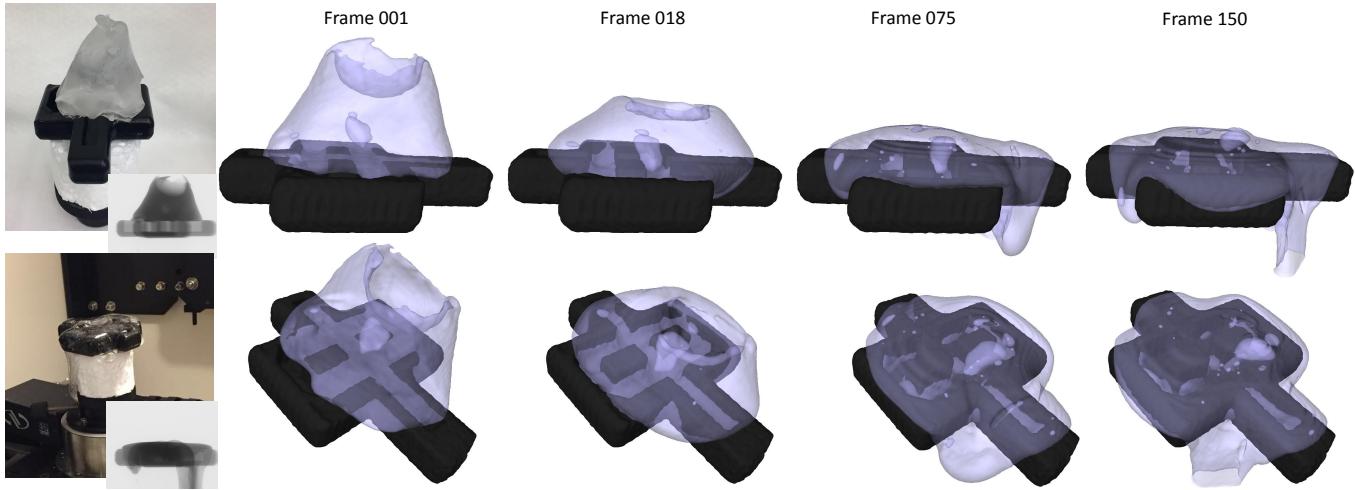


Fig. 6. Reconstruction obtained for the high viscosity fluid dataset. Two photographs are given for the before and after scanning process (left). These photos are accompanied by thumbnail images of the first and last projections. A 3D rendering is given for four different time frames from two different views (right).

**Rose dataset.** In Figure 4 we illustrate the impact of  $N_\theta$  on the quality of a standard (3D) SART reconstruction of the wilting rose. This figure shows the same rose slice obtained using reconstructions with different values of  $N_\theta$ . The central petals of this rose are almost static during the scanning process. Thus, the reconstruction quality of these features improves with an increasing number of projections. This can also be seen in Figure 1-(c). However, when the structures are dynamic, using a large number of projections yields blurred reconstructions, causing these features to disappear as their density gets distributed over many voxels. Note that for Figure 1-(c), we use a threshold for the rendering, that is why these dynamic structures look less blurred in the 3D rendering, but certain features are missing entirely. The blue and green rectangles in Figure 4 illustrate two dynamic features of the rose. One can notice that for  $N_\theta$  higher than 60, some details of these features are blurred. Therefore, for this dataset we choose to use 60 projections to reconstruct each time step using our space-time method. Moreover, Figure 1-(f) shows that the reconstructed rose is quite similar for successive frames (Frames 01-02 and Frames 91-92), which reinforces our choice of  $N_\theta$ . For other datasets we also set this parameter empirically, by looking at the changes in projections of similar angles. The reconstructions shown in Figure 5 demonstrate the accuracy of our approach. Note that the petals curl in on themselves, creating a quite intricate deformation field. Our non-parametric method can handle this situation well, while fitting an appropriate FEM model to this data would be extremely challenging. Despite the high geometric complexity of this dataset, as well as the low contrast in the projections (see Figure 1-(d) and our supplemental video), our reconstruction provides sharp details of the rose during the whole wilting process. This makes our approach very suitable for studying botanical processes that happen at time scales of the scan process.

**Fluid dataset.** The second dataset (see Figure 6) is composed of a static part (mold) and a dynamic part (high viscosity fluid, (PSF-20,000,000 cSt Pure Silicone Fluid) that contains internal structures

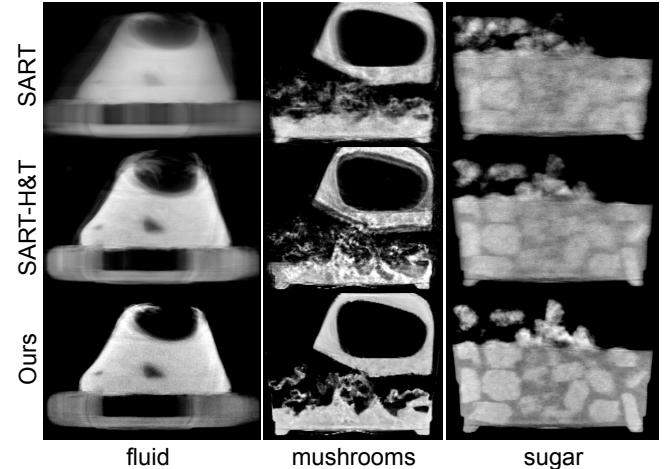


Fig. 7. Comparison of three reconstruction approaches for different datasets (from left to right: fluid, mushrooms and sugar datasets). The first row corresponds to the SART result. The second row are reconstructions using the SART algorithm with a Huber penalty and a temporal smoothing prior. The last row shows the reconstruction obtained with our method.

(bubbles). Given the relatively fast motion of the fluid, only ten projections are used to reconstruct each time frame. Thus, when a SART algorithm is applied on this dataset, the volumes are poorly reconstructed (see the first column of Figure 7). Adding just the spatial priors – the Nonlinear TV prior (ROF) or the Huber penalty prior – is not sufficient to improve the reconstruction quality, as shown in the supplementary material. Moreover, on the provided video we observe a temporal incoherence between the retrieved volumes at successive time steps. With our reconstruction, this is solved by adding temporal smoothness and optical flow priors. The temporal prior ensures a temporal coherence for the reconstructed

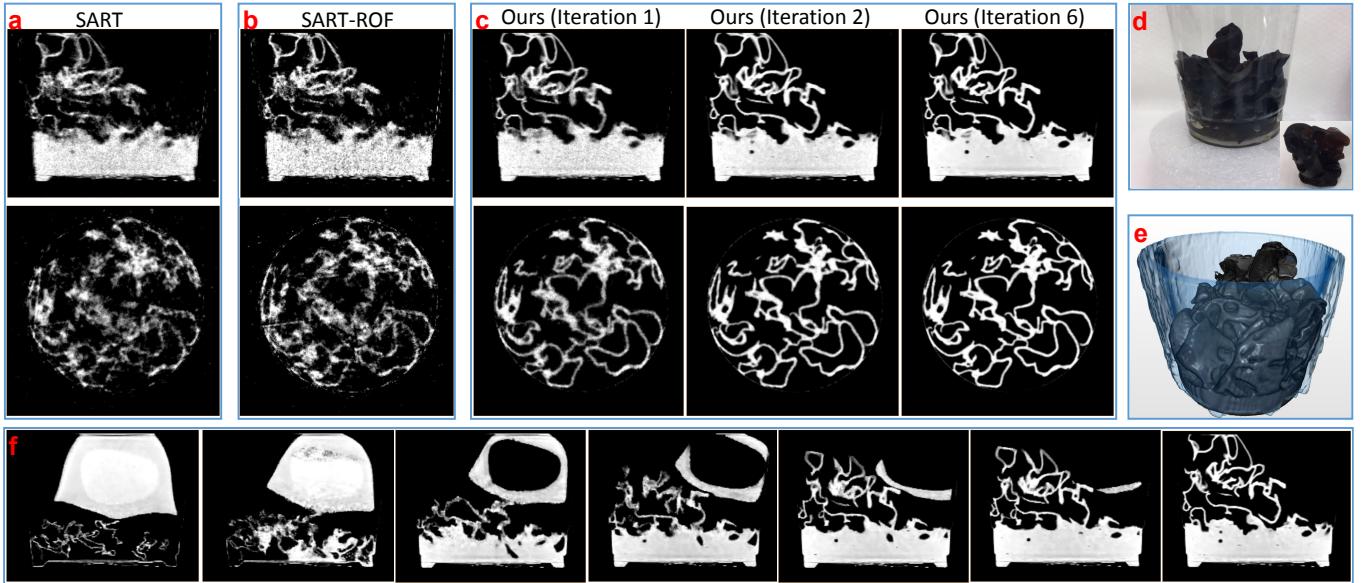


Fig. 8. Reconstruction obtained for the dried black mushrooms dataset. A comparison of different reconstruction methods at one time frame is shown for 2 slices in (a) SART, (b) SART-ROF and (c) Our method with 1 iteration (left), 2 iterations (middle) and 6 iterations (right). Images (d) and (e) represent respectively a photograph and a 3D rendering of the last time frame. Finally, a time sequence of the same slice illustrates the steps of the melting ice and the re-hydration of the mushrooms in (f).

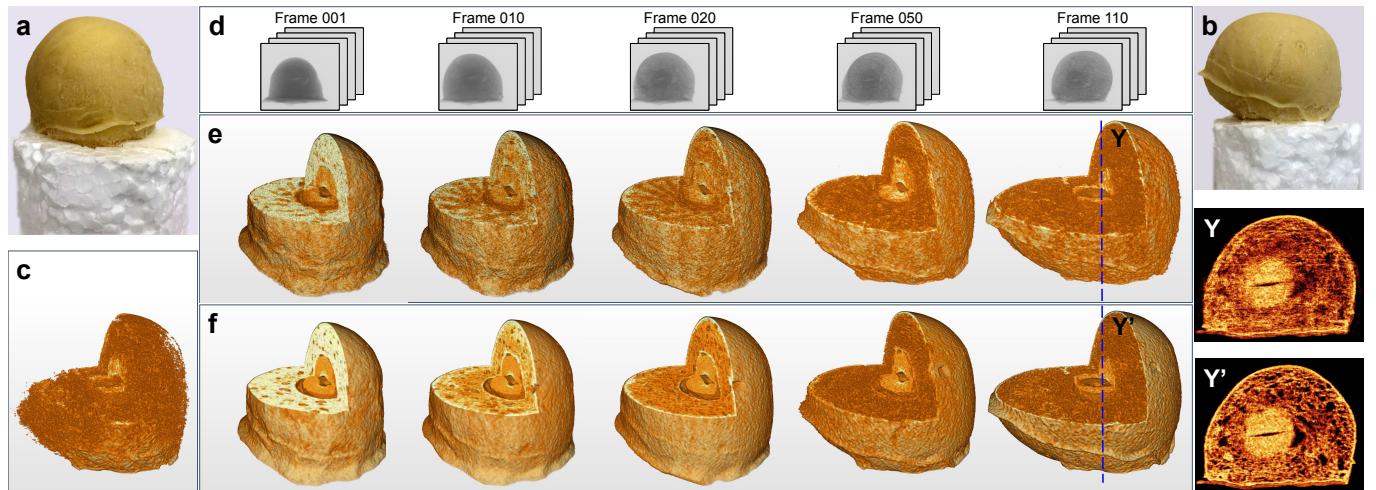


Fig. 9. Reconstruction obtained for the rising dough dataset. Images (a) and (b) show photographs of the dough directly before and directly after the scanning process. Traditional CT reconstructions from all 132 frames (d) in the scan sequence show significant distortions due to misalignment of features in the x-ray projections (c). While the deformation is gradual enough to be negligible over shorter sequences of 20 successive images, this number of projections is too small for reconstructing accurately the internal structures of the dough (e, Y). By comparison, our full space-time reconstructions algorithm yields a time sequence of highly detailed volumes for different time steps (f, Y').

volumes (see the supplemental video). In addition, it allows enriching the angular information, for both the static parts (mold) and the dynamic parts (fluid), since projections used in successive time frames are different. The accuracy of the reconstruction is then improved, especially for static and quasi-static features. In Figure 7, we see clearly that the SART-H&T algorithm (SART algorithm +

Huber penalty + temporal prior) has a better reconstruction for the mold. But since the top of the fluid has large displacements, we observe some blurring and smearing effects. This is solved in our approach by using the optical flow prior and the joint optimization framework. Figure 7 shows the improvement of our method in the

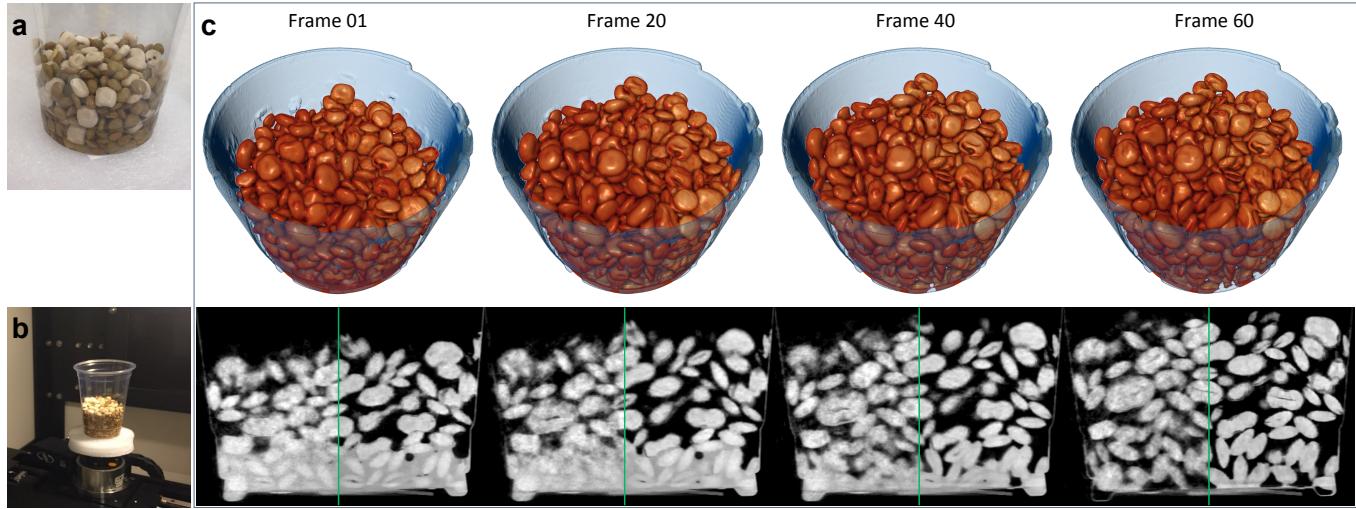


Fig. 10. Reconstruction obtained for the lentils/lupin seeds dataset. (a) and (b) Captured photographs of the seeds sample before and at the end of the scan. (c) 3D rendering (top) and slices visualization (bottom) of the obtained result for 4 different time frames. For these slices the left half corresponds to the reconstruction using a classical method (SART-ROF); while the right half corresponds to our reconstruction result.

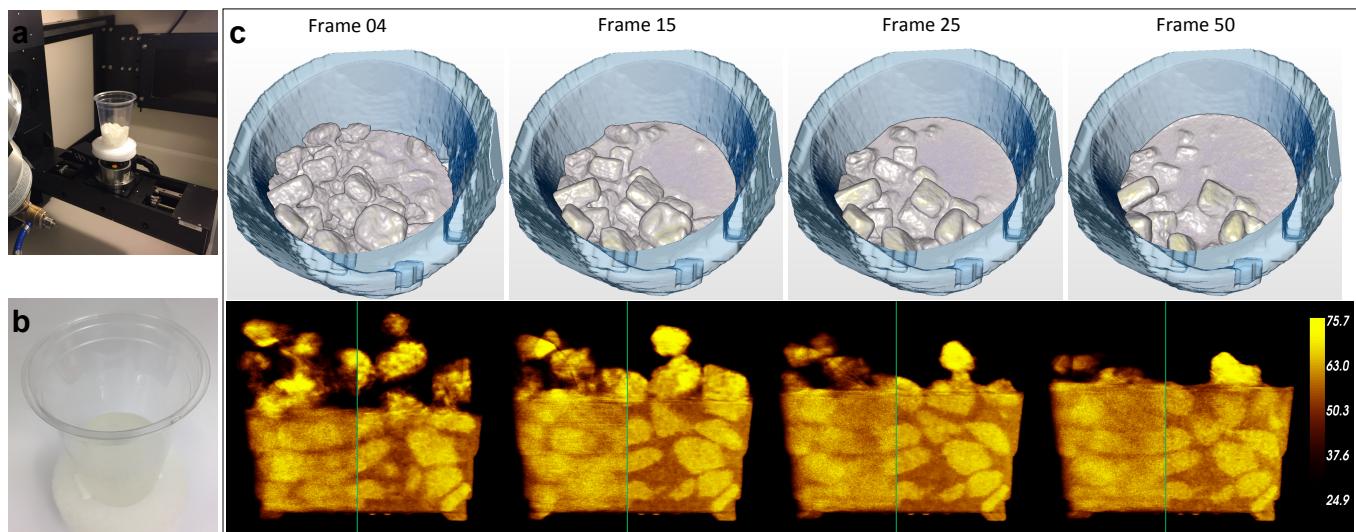


Fig. 11. Reconstruction obtained for the crystal sugar dataset. On the left side, photographs of the sample before (a) and after (b) the scanning process are given. On the right side (c), a 3D rendering for four time frames is given (top), as well as a slices visualization for the same time frames (bottom). For these slices the left hand part corresponds to the reconstruction using a classical method (SART-ROF); while the right side corresponds to our reconstruction result.

reconstruction of the dynamic features of the fluid. In the supplemental video improvements on the reconstruction of the mold are also clearly visible.

**Mushroom dataset.** Similar observations can be made for the slices representing the mushrooms dataset shown in the middle column of Figure 7. In this data, a partially frozen ice cube with a liquid water core is placed on a bed of dried mushrooms (*Auricularia auricula-judae*). The slice sequence (Figure 8-(f)) makes it possible to follow the complete process of re-hydration of the mushrooms.

The first image illustrates the presence of a water cavity inside the ice, showing that the cube was not frozen completely. The water has higher density than ice in the x-ray images, as is to be expected. The second image shows the beginning of the melting of the ice. The thickness of the ice is reduced, and liquid water is now present at the bottom of the container. On the top side of the ice the density is darker than before, while it is brighter on the bottom side. This shows how the liquid water drains from the cavity in the ice cube and is replaced by air. In the third image, all the liquid water leaves the cavity and ends up on the bottom of the container. The remaining

four images show the continuous re-hydration of the mushrooms from the melting ice. The final state of this process is illustrated by the photograph in Figure 8-(d) and the 3D rendering in Figure 8-(e).

As can be seen in Figure 7, the SART and the SART-H&T approaches fail in reconstructing sharp features, especially for the cavity inside the melting ice. Figure 8-(a),(b) and (c) shows a comparison between the SART, the SART-ROF algorithms and three different iterations of our approach for the last time frame. At this stage all the ice has melted, which accelerates the re-hydration of the mushrooms. The increased rate of deformations in the mushrooms causes a reduction in reconstruction quality for the SART and SART-ROF algorithms. After few iterations of our algorithm the result is accurate.

*Dough dataset.* The fourth dataset corresponds to a rising dough shown in the photographs in Figure 9-(a) and (b). In comparison to the rose, the projections of the dough have a good contrast (see Figure 9-(d)). In addition, the dough initially has a relatively simple overall geometric shape comparing to the rose. Thus, the classical method reconstructs the external shape with good accuracy (see Figure 9-(e) and (Y)). However as the dough rises, the yeast creates air bubbles of different sizes. Many air bubbles are too small to be geometrically resolved by the CT scanner. Their impact is visible through a change in the absorption coefficient of the dough. Other bubbles, however, are developing at a macroscopic scale, and show up as internal, sponge-like structures in the geometry. In the classical reconstruction method, these structures are blurred out due to the low projection count. Our space-time reconstruction (see Figure 9-(f) and (Y')), however, is able to recover these bubbles quite well.

*Seeds dataset.* Figure 10 illustrates the results obtained for the reconstruction of the seeds dataset. Before the scan, the seeds were soaked in water for several minutes, but the water was then mostly drained for the actual scan. The supplemental video shows the reconstructed motion of these seeds over the time. By absorbing water these seeds increase in size, and push the other seeds upward. The slices in Figure 10 point up that the magnification of seeds is more important for those in the bottom of the container. We can notice also that at the last frame, all the water is absorbed.

For this dataset, the impact of the "motion blur" is clearly seen on the SART-ROF reconstruction (left side of the slices). This artifact is well corrected using our approach.

*Sugar dataset.* The last experiment consists of imaging the dissolution of sugar crystals in water. A comparison of different reconstruction methods (SART, SART-ROF, SART-H&T and ours) is also given for this dataset in the right column in Figure 7 and in Figure 11. Our reconstruction outperforms the other methods for this dataset as well, by yielding sharper features. When comparing the slices corresponding to Frame 04 and 50 in Figure 11, we notice the density of the water increases slightly over time. This can be explained by the presence of dissolved sugar molecules in the water at the Frame 50. Finally, in the provided video we can see some temporal incoherence in the beginning of the sequence of this dataset. For this case, we are out of the scope of our method since the deformation time scale is smaller than the scanning time for ten projections (here  $N_\theta = 10$ ).

## 6 CONCLUSIONS AND FUTURE WORK

We have presented a method for space-time tomographic reconstruction of objects that undergo significant deformations during the scanning process. We demonstrate our method on a wide variety of input data, ranging from deforming surfaces (e.g. rose, mushrooms) to volumes with changing internal structure (e.g. rising dough).

The success of our method relies on two novel contributions. First, we devise a new sampling strategy for selecting a sequence of viewing angles from which to obtain the x-ray projections. This strategy provides a dense, approximately uniform coverage of angles vs. time, and can be implemented on commercial x-ray CTs without modifications. Furthermore, the sampling strategy does not require any data-dependent parameter selection and can naturally handle both static objects as well as deformations of various magnitudes and speeds.

The second component of our method is a joint image formation model and optimization method for simultaneously recovering a sequence of shapes over time, as well as the deformation fields between them. By jointly solving for both variables, we successfully transfer information between time steps of different deformation states, and are able to overcome quality and resolution issues that would result from independent reconstructions of each frame. Our method is non-parametric; we do not require projections of either the volumes or the motions into basis functions. This makes it easy to apply our method to a wide range of geometries with varying topologies without manual tweaking.

The limitation of the approach is that we still have to assume a relatively small motion at the scale of a small number of projections (10-60 in our examples). Our method does not work if there already is significant deformation between two successively captured projections. In the future we believe we can address this situation by moving to a continuous time scale, in which we respect the exact capture time for each individual projection image in the optimization method.

Despite this limitation on deformation speed, our method already provides an efficient and effective means of dealing with deformations in x-ray computed tomography. This allows to analyze time-varying phenomena with interesting changes in internal structure at time scales that could not previously be handled.

## ACKNOWLEDGMENTS

This work was supported by King Abdullah University of Science and Technology as part of VCC Center Competitive Funding. The authors would like to thank Thomas Theussl for helping with the volume renderings, and Vinicius Lube for helping with the CT calibration. We also thank Jing Ren, Qilin Sun, Qiang Fu and Abdulrahman Aljedaani for insightful discussions on the datasets.

## REFERENCES

- Mohamed Aly, Guangming Zang, Wolfgang Heidrich, and Peter Wonka. 2016. TRex: A Tomography Reconstruction Proximal Framework for Robust Sparse View X-Ray Applications. *arXiv preprint arXiv:1606.03601* (2016).
- 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.
- 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. *ACM Trans. Graph.* 27, 5 (2008), 132.

- Brian K Bay, Tait S Smith, David P Fyhré, and Malik Saad. 1999. Digital volume correlation: three-dimensional strain mapping using X-ray tomography. *Exp. Mech.* 39, 3 (1999), 217–226.
- Derek Bradley, Tiberiu Popa, Alla Sheffer, Wolfgang Heidrich, and Tammy Boubekeur. 2008. Markerless garment capture. 27, 3 (2008), 99.
- Martin Burger, Hendrik Dirks, and Carola-Bibiane Schönlieb. 2018. A variational model for joint motion estimation and image reconstruction. *SIAM Journal on Imaging Sciences* 11, 1 (2018), 94–128.
- Antonin Chambolle and Thomas Pock. 2011. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging and Vision* 40, 1 (2011), 120–145.
- Guang-Hong Chen, Pascal Thériault-Lauzier, Jie Tang, Brian Nett, Shuai Leng, Joseph Zambelli, Zhihua Qi, Nicholas Bevins, Amish Raval, Scott Reeder, et al. 2012. Time-resolved interventional cardiac C-arm cone-beam CT: An application of the PICCS algorithm. *IEEE Trans. Med. Img.* 31, 4 (2012), 907–923.
- W Chlewicki, C Badea, and N Pallikarakis. 2001. Cone based 3D reconstruction: a FDK-SART comparison for limited number of projections. *Proceedings of MEDICON 2001* (2001).
- Edilson De Aguiar, Carsten Stoll, Christian Theobalt, Naveed Ahmed, Hans-Peter Seidel, and Sebastian Thrun. 2008. Performance capture from sparse multi-view video. 27, 3 (2008), 98.
- Laurent Desbat, Sébastien Roux, and Pierre Grangeat. 2007. Compensation of some time dependent deformations in tomography. *IEEE Trans. Med. Img.* 26, 2 (2007), 261–269.
- Mingsong Dou, Sameh Khamis, Yury Degtyarev, Philip Davidson, Sean Ryan Fanello, Adarsh Kowdle, Sergio Orts Escalano, Christoph Rhemann, David Kim, Jonathan Taylor, et al. 2016. Fusion4D: Real-time performance capture of challenging scenes. *ACM Trans. Graph.* 35, 4 (2016), 114.
- Gerrit E Elsinga, Fulvio Scarano, Bernhard Wieneke, and Bas W van Oudheusden. 2006. Tomographic particle image velocimetry. *Exp. Fluids* 41, 6 (2006), 933–947.
- LA Feldkamp, LC Davis, and JW Kress. 1984. Practical cone-beam algorithm. *JOSA A* 1, 6 (1984), 612–619.
- Pascal Getreuer. 2012. Rudin-Osher-Fatemi total variation denoising using split Bregman. *Image Processing On Line* 2 (2012), 74–95.
- James Gregson, Ivo Ihrke, Nils Thuerey, and Wolfgang Heidrich. 2014. From capture to simulation: connecting forward and inverse problems in fluids. *ACM Trans. Graph.* 33, 4 (2014), 139.
- James Gregson, Michael Krímerman, Matthias B Hullin, and Wolfgang Heidrich. 2012. Stochastic tomography and its applications in 3D imaging of mixing fluids. *ACM Trans. Graph.* 31, 4 (2012), 52–1.
- Samuel W Hasinoff and Kiriakos N Kutulakos. 2007. Photo-consistent reconstruction of semitransparent scenes by density-sheet decomposition. *IEEE Trans. PAMI* 29, 5 (2007), 870–885.
- Gabor T Herman. 2009. *Fundamentals of computerized tomography: image reconstruction from projections*.
- François Hild, Hugo Leclerc, and Stéphane Roux. 2014. Performing DVC at the Voxel Scale. In *Advancement of Optical Methods in Experimental Mechanics, Volume 3*, 209–215.
- Berthold KP Horn and Brian G Schunck. 1981. Determining optical flow. *Artificial Intelligence* 17, 1-3 (1981), 185–203.
- Peter J Huber. 2011. Robust statistics. In *Int. Encyclopedia of Statistical Science*. 1248–1251.
- Ivo Ihrke and Marcus Magnor. 2004. Image-based tomographic reconstruction of flames. In *Proc. SCA*. 365–373.
- Takashi Ijiri, Shin Yoshizawa, Hideo Yokota, and Takeo Igarashi. 2014. Flower modeling via X-ray computed tomography. *ACM Trans. Graph.* 33, 4 (2014), 48.
- Matthias Innmann, Michael Zollhöfer, Matthias Nießner, Christian Theobalt, and Marc Stamminger. 2016. VolumeDeform: Real-time volumetric non-rigid reconstruction. In *Proc. ECCV*. Springer, 362–379.
- Oszkár Józsa, Attila Börcs, and Csaba Benedek. 2013. Towards 4D virtual city reconstruction from Lidar point cloud sequences. *ISPRS*.
- Avinash C Kak and Malcolm Slaney. 2001. *Principles of computerized tomographic imaging*. SIAM.
- Sakiho Kato, Tomofumi Narita, Chika Tomiyama, Takashi Ijiri, and Hiroya Tanaka. 2017. 4D computed tomography measurement for growing plant animation. In *SIGGRAPH ASIA 2017 Posters*.
- ByungMoon Kim, Yingjie Liu, Ignacio Llamas, and Jaroslaw R Rossignac. 2005. *Flowfixer: Using bfec for fluid simulation*. Technical Report. Georgia Institute of Technology.
- Joël Lachambre, Julien Réthoré, Arnaud Weck, and Jean-Yves Buffiere. 2015. Extraction of stress intensity factors for 3D small fatigue cracks using digital volume correlation and X-ray tomography. *Int. J. Fatigue* 71 (2015), 3–10.
- Douglas Lanman, Gordon Wetzstein, Matthew Hirsch, Wolfgang Heidrich, and Ramesh Raskar. 2011. Polarization fields: dynamic light field display using multi-layer LCDs. *ACM Trans. Graph.* 30, 6 (2011), 186.
- Hugo Leclerc, Stéphane Roux, and François Hild. 2015. Projection savings in CT-based digital volume correlation. *Exp. Mech.* 55, 1 (2015), 275–287.
- Hao Li, Linjie Luo, Daniel Vlasic, Pieter Peers, Jovan Popović, Mark Pauly, and Szymon Rusinkiewicz. 2012. Temporally coherent completion of dynamic shapes. *ACM Trans. Graph.* 31, 1 (2012), 2.
- Enric Meinhardt-Llopis, Javier Sánchez Pérez, and Daniel Kondermann. 2013. Horn-Schunck Optical Flow with a Multi-Scale Strategy. *Image Processing On Line*, 2013: 151–172, 2013. (2013).
- Niloy J Mitra, Simon Flöry, Maks Ovsjanikov, Natasha Gelfand, Leonidas J Guibas, and Helmut Pottmann. 2007. Dynamic geometry registration. In *Symposium on geometry processing*. 173–182.
- K Aditya Mohan, SV Venkatakrishnan, John W Gibbs, Emine Begum Gulsoy, Xianghui Xiao, Marc De Graef, Peter W Voorhees, and Charles A Bouman. 2015. TIMBR: A method for time-space reconstruction from interlaced views. 1, 2 (2015), 96–111.
- Thilo F Morgeneyer, Lukas Helfen, Hazem Mubarak, and François Hild. 2013. 3D digital volume correlation of synchrotron radiation laminography images of ductile crack initiation: an initial feasibility study. *Exp. Mech.* 53, 4 (2013), 543–556.
- Cyril Mory, Vincent Auvray, Bo Zhang, Michael Grass, Dirk Schäfer, S James Chen, John D Carroll, Simon Rit, Françoise Peyrin, Philippe Douek, et al. 2014. Cardiac C-arm computed tomography using a 3D + time ROI reconstruction method with spatial and temporal regularization. *Med. Phys.* 41, 2 (2014).
- Cyril Mory, Bo Zhang, Vincent Auvray, Michael Grass, Dirk Schäfer, Françoise Peyrin, Simon Rit, Philippe Douek, and Loïc Boussel. 2012. ECG-gated C-arm computed tomography using L1 regularization. In *Proc. EUSIPCO*. 2728–2732.
- J Neggers, JPM Hoefnagels, MGD Geers, F Hild, and S Roux. 2015. Time-resolved integrated digital image correlation. *Internat. J. Numer. Methods Engrg.* 103, 3 (2015), 157–182.
- Alex Reche-Martinez, Ignacio Martin, and George Drettakis. 2004. Volumetric reconstruction and interactive rendering of trees from photographs. *ACM Trans. Graph.* 23, 3 (2004), 720–727.
- Stéphane Roux, François Hild, Philippe Viot, and Dominique Bernard. 2008. Three-dimensional image correlation from X-ray computed tomography of solid foam. *Composites Part A: Applied science and manufacturing* 39, 8 (2008), 1253–1265.
- Hubert Schreier, Jean-José Orteu, and Michael A Sutton. 2009. *Image correlation for shape, motion and deformation measurements*. Springer US.
- Emil Y Sidky and Xiaochuan Pan. 2008. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine and Biology* 53, 17 (2008), 4777.
- Jan-Jakob Sonke, Lambert Zijp, Peter Remeijer, and Marcel van Herk. 2005. Respiratory correlated cone beam CT. *Med. Phys.* 32, 4 (2005), 1176–1186.
- Wolfgang H Stupp, Jessica A Maisano, Matthew W Colbert, Paula J Rudall, and Timothy B Rowe. 2003. Three-dimensional analysis of plant structure using high-resolution X-ray computed tomography. *Trends in Plant Science* 8, 1 (2003), 2–6.
- Thibault Taillandier-Thomas, Stéphane Roux, and François Hild. 2016. Soft route to 4D tomography. *Phys. Rev. Letters* 117, 2 (2016), 025501.
- Oliver Taubmann, Günter Lauritsch, Andreas Maier, Rebecca Fahrig, and Joachim Horngger. 2015. Estimate, compensate, iterate: joint motion estimation and compensation in 4-D cardiac C-arm computed tomography. In *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*. 579–586.
- Borislav Trifonov, Derek Bradley, and Wolfgang Heidrich. 2006. Tomographic reconstruction of transparent objects. In *Proc. EGSR*.
- JG Van der Corput. 1935. *Verteilungsfunktionen I, II*. *Nederl. Akad. Wetensch. Proc.* 38 (1935).
- Elco Verhulp, Bert van Rietbergen, and Rik Huiskes. 2004. A three-dimensional digital image correlation technique for strain measurements in microstructures. *J. Biomechanics* 37, 9 (2004), 1313–1320.
- Michael Wand, Bart Adams, Maksim Ovsjanikov, Alexander Berner, Martin Bokeloh, Philipp Jenke, Leonidas Guibas, Hans-Peter Seidel, and Andreas Schilling. 2009. Efficient reconstruction of nonrigid shape and motion from real-time 3D scanner data. *ACM Trans. Graph.* 28, 2 (2009), 15.
- Gordon Wetzstein, Douglas Lanman, Wolfgang Heidrich, and Ramesh Raskar. 2011. Layered 3D: tomographic image synthesis for attenuation-based light field and high dynamic range displays. *ACM Trans. Graph.* 30, 4 (2011), 95.
- Jinhui Xiong, Ramzi Idoughi, Andres A Aguirre-Pablo, Abdulrahman B AlJedaani, Xiong Dun, Qiang Fu, Sigurdur T Thoroddsen, and Wolfgang Heidrich. 2017. Rainbow particle imaging velocimetry for dense 3D fluid velocity imaging. *ACM Trans. Graph.* 36, 4 (2017), 36.
- Shuang Zhao, Wenzel Jakob, Steve Marschner, and Kavita Bala. 2011. Building volumetric appearance models of fabric using micro CT imaging. *ACM Trans. Graph.* 30, 4 (2011), 44.
- Qian Zheng, Xiaochen Fan, Minglun Gong, Andrei Sharf, Oliver Deussen, and Hui Huang. 2017. 4D Reconstruction of Blooming Flowers. *CGF* 36, 6 (2017), 405–417.