

Spatiotemporal 3D Models of Aging Fruit from Multi-view Time-Lapse Videos

Lintao Guo, Hunter Quant, Nikolas Lamb, Benjamin Lowit, Sean Banerjee,
and Natasha Kholgade Banerjee^(✉)

Clarkson University, Potsdam, NY 13699, USA
{linguo, quanthd, lambne, lowitbp, sbanerje, nbanerje}@clarkson.edu

Abstract. We provide an approach to reconstruct spatiotemporal 3D models of aging objects such as fruit containing time-varying shape and appearance using multi-view time-lapse videos captured by a microenvironment of Raspberry Pi cameras. Our approach represents the 3D structure of the object prior to aging using a static 3D mesh reconstructed from multiple photographs of the object captured using a rotating camera track. We manually align the 3D mesh to the images at the first time instant. Our approach automatically deforms the aligned 3D mesh to match the object across the multi-viewpoint time-lapse videos. We texture map the deformed 3D meshes with intensities from the frames at each time instant to create the spatiotemporal 3D model of the object. Our results reveal the time dependence of volume loss due to transpiration and color transformation due to enzymatic browning on banana peels and in exposed parts of bitten fruit.

Keywords: Spatiotemporal · 3D model · Multi view · Time-varying
Time lapse

1 Introduction

Public repositories of 3D models of objects are expanding rapidly, with ready-to-use 3D models available for a wide variety of applications such as animation [5], physics simulations [3, 4], robotic manipulation [8] and photo-editing [16]. However, the information contained in publicly available 3D models is still limited. Most 3D models only contain the geometry and textures for rigid objects. While rigged models exist for faces and bodies, current 3D model repositories lack the range of dynamic behaviors exhibited by real-world objects such as plush toys deforming under impact, metal corroding, chocolate melting, and bitten fruit such as the apple in Fig. 1 shrinking and browning when kept outside.

In this paper, we provide a data-driven approach to reconstruct spatiotemporal 3D models of objects such as fruit that undergo changes in shape and appearance due to aging when exposed to the environment. Unlike prior approaches on time-varying aging that focus on appearance transformations [11, 13, 19, 23], our approach models both appearance change due to enzymatic browning [25] and shape deformations due to water loss.

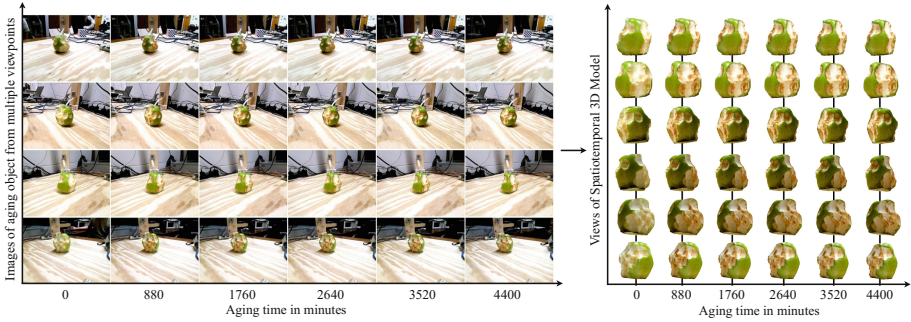


Fig. 1. Given time-lapse videos of an aging object from multiple viewpoints as shown on the left, our approach reconstructs a spatiotemporal 3D model as shown by the views on the right.

Our approach deforms a static 3D mesh of the object reconstructed prior to aging to fit synchronized time-lapse videos of the object captured from multiple viewpoints. As discussed in Sect. 3, we contribute a 3D printed camera track to automatically capture and stitch multiple photographs of the object into the static mesh. We contribute a microenvironment of multiple Raspberry Pi v2 cameras that automatically capture the synchronized time-lapse videos from various viewpoints as described in Sect. 4. We provide a manual interface to rigidly align and deform the static mesh to images at the first time instant as discussed in Sect. 5. Our approach automatically deforms the 3D mesh to match the object in the time-lapse videos and maps the deformed meshes with seam-free textures from the time-lapse images as discussed in Sects. 6 and 7.

Section 8 shows the spatiotemporal 3D models resulting from our approach. Our spatiotemporal 3D models enable data-driven analysis of the physical properties of objects aging under environmental effects as shown by our results on volume loss and appearance transformation in Sect. 8. The spatiotemporal 3D models provided by our work have the potential to enhance consumer quality of life by providing applications such as data-driven prediction of shelf-life of perishable food products from grocery store cameras, automated monitoring of the structural health of erosive materials used in building constructions, and intelligent updates to end users on the status of fruit, cooked items, and baked goods in their home environments.

2 Related Work

Many approaches simulate time-varying phenomena such as metal erosion, paint cracking, and plant aging using physics-based and biological models [10, 17, 26, 30]. These models often lack comprehensive representation of fine-scale appearance changes in real-world objects. Several data-driven approaches model real-world appearance by capturing material samples using a single camera [11], multiple cameras [13, 19], or scanner-camera setups [23]. None of these approaches

capture time-varying shape. In conjunction with appearance, our approach captures the shape deformation exhibited by fruit undergoing volume changes due to water loss. The approach of Li et al. [21] analyzes plant growth using 3D point clouds captured by rotating the plant and imaging it using a single structured light scanner. Unlike their approach, we use multiple cameras to capture the 3D structure of the object by keeping it still to avoid rotational forces from deforming the object shape. Their approach yields uncorresponded 3D point clouds, while our approach yields spatiotemporal models with mesh vertices corresponded in time.

Our work is more closely related to approaches that reconstruct spatiotemporal 3D shape and appearance models of facial and body motion from multiple viewpoint videos. These approaches need to address the challenge of estimating frame-to-frame correspondences where brightness constancy may not hold between adjacent frames due to aggressive face and body motions or due to occlusion from moving body parts. The approach of de Aguiar et al. [2] matches 3D body model points to interest point correspondences between multiple images at the same time instant, and uses silhouette rims to refine the alignment. Beeler et al. [6] cluster video frames as anchored and unanchored to a reference frame, propagate a 3D mesh to fit reference-to-anchored and anchored-to-unanchored correspondences obtained using normalized cross-correlation, and perform a global optimization to refine the propagated 3D meshes under constraints of image fidelity, mesh consistency, and geometric smoothness. Unlike such approaches, we use small time-lapses to capture slow aging transformations occurring over the span of hours to days, where brightness is nearly constant between adjacent frames, enabling the use of optic flow techniques such as [24].

3 Camera Track to Capture Static 3D Mesh

We provide a rotating camera track of diameter 620 mm shown in Fig. 2(a) to automatically capture photographs of the object from multiple viewpoints. The

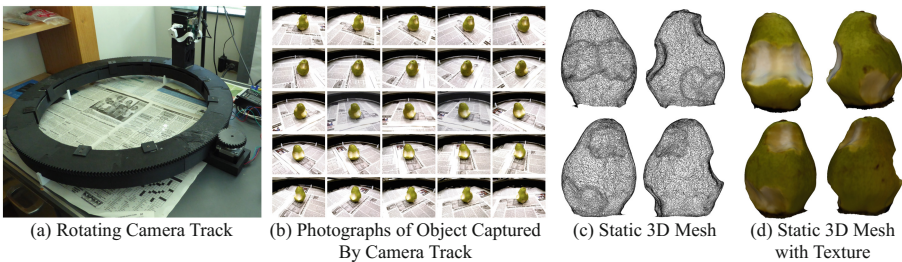


Fig. 2. (a) Rotating track with a Raspberry Pi camera attached to a tower to image an object from multiple viewpoints. (b) Photographs of a pear captured by the rotating camera. (c) Static 3D mesh reconstructed by stitching the photographs. (d) Static 3D mesh mapped with the texture from the photographs.

track is constructed by joining 12 programmatically modeled 3D printed arcs as discussed in [18], and is controlled using a Nema 23 stepper motor with a 1.8° step angle. The track rotates over ball bearings on a base assembled from slottable 3D printed segments. We use a Raspberry Pi v3 computer to program the movement of the stepper and the imaging of the object using a Raspberry Pi v2 camera. Figure 2(b) shows the images of a pear captured using the track. We stitch the photographs into a static 3D mesh using Autodesk ReMake [1] shown in Figs. 2(c) and (d).

4 Microenvironment of Multiple Raspberry Pi Cameras to Capture Time-Lapse Videos

As shown in Fig. 3(a), we use a microenvironment of four Raspberry Pi v2 cameras to capture multiple viewpoint time-lapse videos of the object aging. Details of the microenvironment construction can be found in [14]. We provide a Java graphical user interface (GUI) that allows a user to set a time-lapse duration and capture length, and run the time-lapse capture of an aging object. When a capture process is started, the server sends a parallelized capture signal to each Raspberry Pi computer at every time-lapse instant. On receiving the signal, each Raspberry Pi invokes an image capture. We manually set the exposure value and white balance to be constant. Each Raspberry Pi computer sends the captured image to the master computer to augment corresponding video sequences. Figure 3(b) shows multi-view images of a banana blackening captured using the Raspberry Pi microenvironment. We refer to each set of multi-view images at a single time instant in the videos as a frame F .

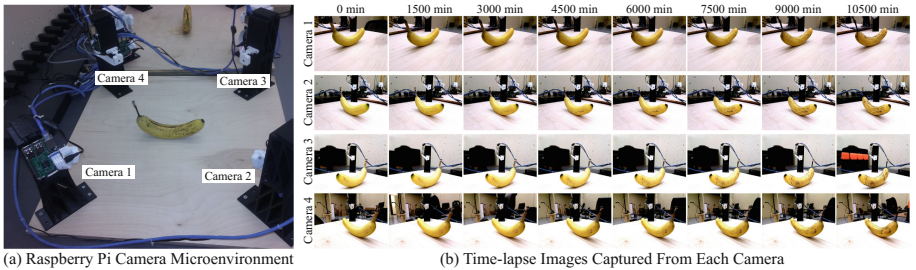


Fig. 3. (a) Microenvironment of four Raspberry Pi v2 cameras to capture time-varying phenomena such as aging. (b) Sample images shown from a time-lapse video of a blackening banana. While images are shown every 1000 min to show the range of the video, the time-lapse interval is finer at 5 min.

5 3D Mesh Alignment to Initial Frame

While the 3D mesh reconstructed using ReMake provides an accurate representation of the object, small-scale deviations due to ground contact, imperfect correspondences between photographs, or ambiguities in resolving illumination from

reflectance often render the 3D model imprecise. We provide a manual approach that adapts the geometry correction method of Kholgade et al. [16] to match the 3D mesh to the object in multiple viewpoint images at the initial frame of the time-lapse videos, i.e., at $F = 0$, using user-provided correspondences. While automated approaches to perform 3D model alignment exist [15, 22], they fail to provide unsupervised pixel-precise alignment for unstructured object collections such as the fruit in this work.

We first perform a rigid alignment of the 3D mesh to the 2D images from 2D-3D correspondences. As shown in Fig. 4(a), for the j^{th} image in the initial frame, we mark m_j 2D points \mathbf{x}_{k_j} in the image and corresponding 3D points \mathbf{X}_{k_j} in the 3D mesh, where $k_j \in \{1, 2, \dots, m_j\}$. We use the efficient Perspective n -Point algorithm (ePnP) [20] to estimate the rigid pose of each camera consisting of its rotation \mathbf{R}_j and translation \mathbf{t}_j with respect to the 3D mesh using $m_j \geq 4$. For each image, the ePnP algorithm requires $m_j \geq 4$ correspondences, necessitating $\sum_{j=1}^m m_j \geq 4m$ correspondences, where m is the number of cameras.

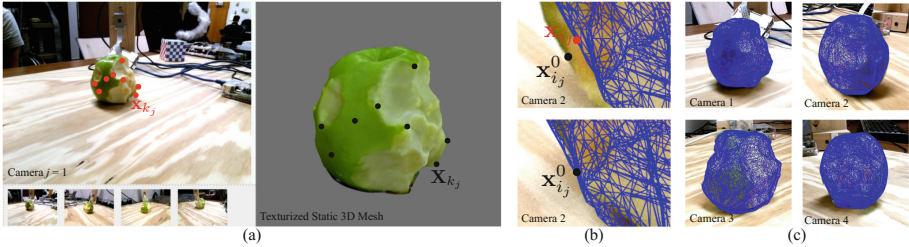


Fig. 4. Alignment of the 3D mesh to the frame at time instant 0 min, i.e., the initial frame. (a) Interface to manually perform rigid alignment of the 3D mesh on the right to the initial frame images from all four cameras on the left. The 3D mesh is texture-mapped with the stitched texture from the static mesh reconstruction for ease of point marking. (b) The user marks a single point correspondence in the second image (top) to deform the 3D mesh (bottom). (c) 3D mesh aligned to the object in all four images after several user-marked correspondences.

We use the pose $\{\mathbf{R}_j, \mathbf{t}_j\}$ of each camera to project the 3D mesh $\mathbf{X} \in \mathbb{R}^{3 \times N}$ into each camera image, where N is the number of vertices on the mesh. We manually deform \mathbf{X} to mesh \mathbf{X}^0 with pixel-precise alignment to the object contours at the initial frame $F = 0$ by marking n_j 2D correspondences $\mathbf{x}_{i_j} \in \mathbb{R}^2$ on the 3D mesh projection and $\mathbf{x}_{i_j}^0 \in \mathbb{R}^2$ on object contours, where $i_j \in \{1, 2, \dots, n_j\}$. The top image in Fig. 4(b) shows a single user-provided 2D-2D correspondence. We backproject rays from the camera center through \mathbf{x}_{i_j} to obtain 3D points $\mathbf{X}_{i_j} \in \mathbb{R}^3$ at the intersection of the rays and the faces of \mathbf{X} . Similar to Kholgade et al. [16], we estimate the deformed 3D mesh $\mathbf{X}^0 \in \mathbb{R}^{3 \times N}$ from \mathbf{X}_{i_j} and $\mathbf{x}_{i_j}^0$ by optimizing the objective function $E(\mathbf{X}^0)$ in \mathbf{X}^0 , where

$$E(\mathbf{X}^0) = \lambda_{\text{ray}} E_{\text{ray}}(\mathbf{X}^0) + \lambda_{\text{lap}} E_{\text{lap}}(\mathbf{X}^0) + \lambda_{\text{small}} E_{\text{small}}(\mathbf{X}^0). \quad (1)$$

As in Kholgade et al., the term E_{ray} in Eq. (1) constrains each 3D point $\mathbf{X}_{i_j}^0$ to lie as close as possible to the projection of $\mathbf{X}_{i_j}^0$ on the ray $\mathbf{v}_{i_j}^0 = \mathbf{K}_j^{-1} [\mathbf{x}_{i_j}^0; 1]$ ¹ through the 2D point $\mathbf{x}_{i_j}^0$. E_{ray} is described as

$$E_{\text{ray}}(\mathbf{X}^0) = \sum_{i_j=1}^{n_j} \left\| \left(\mathbf{I}_3 - \mathbf{v}_{i_j}^0 (\mathbf{v}_{i_j}^0)^T / \|\mathbf{v}_{i_j}^0\|^2 \right) \mathbf{R}_j \mathbf{X}_{i_j}^0 \right\|^2, \quad (2)$$

where \mathbf{I}_3 represents the 3×3 identity matrix. The term E_{lap} represents a simplified version of the Laplacian surface energy from the approach of Sorkine and Alexa [29]. Since our deformations are small, we eliminate co-tangent weights and local rotation estimation from their approach to yield E_{lap} as

$$E_{\text{lap}}(\mathbf{X}^0) = \sum_{i=1}^N \left\| (\mathbf{X}_i^0 - \mathbf{X}_i) - \sum_{l \in \mathcal{N}(i)} (\mathbf{X}_l^0 - \mathbf{X}_l) / \|\mathcal{N}(i)\| \right\|^2, \quad (3)$$

where $\mathcal{N}(i)$ is the 1-ring neighborhood of \mathbf{X}_i^0 . We drop the symmetry term from approach of Kholgade et al., as our objects show asymmetry due to slicing or bite marks. We introduce the term E_{small} in our approach, where

$$E_{\text{small}}(\mathbf{X}^0) = \sum_{i=1}^N \|\mathbf{X}_i^0 - \mathbf{X}_i\|^2 \quad (4)$$

to restrict the deformations to be as small as possible. The term prevents an arbitrary resizing and translation of the 3D mesh due to the scale ambiguity inherent to the relationship between the 3D mesh and the cameras. The bottom image of Fig. 4(b) shows the 3D mesh deformed to match the 2D-2D correspondence marked at the top. Figure 4(c) shows the deformed 3D mesh aligned to all four images of the first time-lapse frame for an apple. We use weights $\lambda_{\text{ray}} = 1$, $\lambda_{\text{lap}} = 5$, and $\lambda_{\text{small}} = .001$ in this work.

6 Automatic 3D Mesh Deformation to Time-Lapse Videos

Given the 3D mesh aligned to the initial frame, our approach automatically estimates a deformation of the 3D mesh to match point correspondences on the object over successive frames in the sequence of multi-view time-lapse videos. Our approach leverages the assumption that adjacent time-lapse images from a particular viewpoint show minimal change in shape and appearance, enabling the use of optic flow to estimate point correspondences between successive frames. We use the Kanade-Lucas-Tomasi (KLT) tracking approach [24] to estimate n_j^F 2D point correspondences $\mathbf{x}_{i_j}^F$ and $\mathbf{x}_{i_j}^{F+1}$ between the j^{th} images in the current frame F and the next frame $F+1$, where $i_j \in \{1, 2, \dots, n_j^F\}$. Figure 5(a) shows the 2D point correspondences estimated between two adjacent frames for the bitten apple. We use the projection of the 3D mesh in each image of the frame F to retain point correspondences within the object contour.

¹ In $[\mathbf{x}_{i_j}^0; 1]$, the semi-colon appends 1 at the end of column vector $\mathbf{x}_{i_j}^0$.

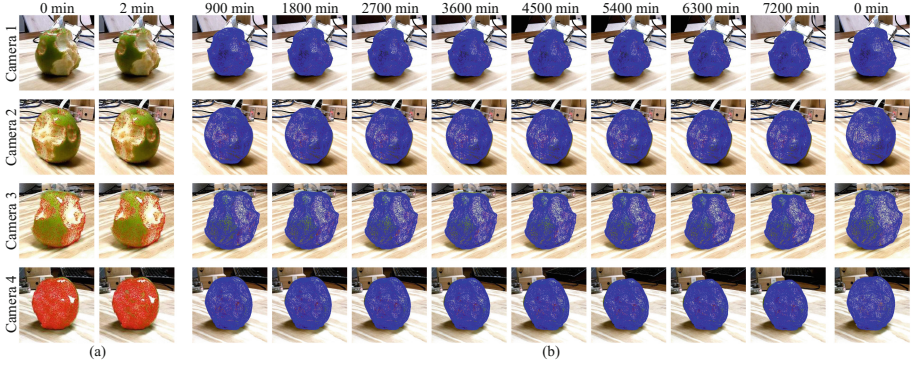


Fig. 5. Automatic deformation to multi-view time-lapse sequence of the bitten apple. (a) Point correspondences in red obtained using Kanade-Lucas-Tomasi (KLT) tracking between successive images from the first two frames. (b) Deformed 3D mesh for each camera image at various instances in the time-lapse sequence. The 3D mesh at time instant 0 min from Fig. 4(c) is shown at the right for size comparison to the 3D mesh at the last time instant of 7200 min. (Color figure online)

Given a deformed 3D mesh $\mathbf{X}^F \in \mathbb{R}^{3 \times N}$ at frame F , we determine each 3D point \mathbf{X}_{ij}^F on \mathbf{X}^F by back-projecting a ray from the camera center through \mathbf{x}_{ij}^F and determining barycentric coordinates α_{li_j} for vertices \mathbf{X}_l on the face $\mathcal{F}_k \in \mathcal{F}$ that intersects the ray, where $l \in \mathcal{F}_k$ and \mathcal{F} is the set of mesh faces. To reduce drift over the 3D model deformations, we deform the initial mesh \mathbf{X}^0 instead of \mathbf{X}^F to fit the corresponding points \mathbf{x}_{ij}^{F+1} in frame $F+1$. We compute 3D points on \mathbf{X}^0 corresponding to \mathbf{X}_{ij}^F as $\mathbf{X}_{ij}^0 = \sum_{l \in \mathcal{F}_k} \alpha_{li_j} \mathbf{X}_l^0$. We estimate the deformed 3D mesh \mathbf{X}^{F+1} in the frame $F+1$ by optimizing the objective function $E(\mathbf{X}^{F+1})$ in \mathbf{X}^{F+1} . We set up $E(\mathbf{X}^{F+1})$ by performing the substitutions $\mathbf{X} \leftarrow \mathbf{X}^0$, $\mathbf{X}^0 \leftarrow \mathbf{X}^{F+1}$, $n_j \leftarrow n_j^F$, and $\mathbf{v}_{ij}^0 \leftarrow \mathbf{v}_{ij}^{F+1}$ in Eqs. (1)–(4). The quantity $\mathbf{v}_{ij}^{F+1} = \mathbf{K}_j^{-1} [\mathbf{x}_{ij}^{F+1}; 1]$ represents the ray through the 2D point \mathbf{x}_{ij}^{F+1} in the frame $F+1$. Figure 5(b) shows the deformed mesh aligned to several frames in the time-lapse sequence of the bitten apple.

7 Texture Mapping the Deformed 3D Meshes

Given the 3D meshes deformed to match each time-lapse frame, we texture map each mesh with the intensities from the individual images. For the frame F of the time-lapse sequence, we use ray intersections to obtain the subset $\mathcal{F}_j^F \in \mathcal{F}$ of faces visible from the viewpoint of the j^{th} camera. We sample the intensities at the vertices indexed by \mathcal{F}_j^F in the j^{th} image. We use the sampled intensities to create a texture mapped sub-mesh consisting of the vertices indexed by \mathcal{F}_j^F . Figure 6(a) shows the sub-meshes for the bitten apple from the four cameras.

We blend the sub-meshes using TextureStitcher [7] to create a seamless texture-mapped 3D mesh at frame F , as shown in Fig. 6(b). The texture-mapped meshes for all time instants form a spatiotemporal 3D model for the object.

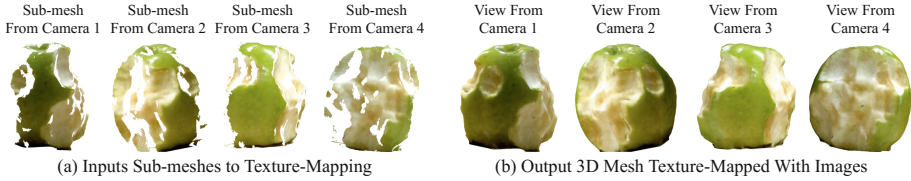


Fig. 6. Texture-mapping the 3D model. (a) As input to TextureStitcher [7], we provide four sub-meshes obtained by sampling the intensities from each image at the vertices of the faces visible in the corresponding camera. (b) TextureStitcher provides the complete 3D mesh as output containing seamlessly blended texture from the input sub-meshes.

8 Results

The top rows of Figs. 7 and 9(a)–(c) show spatiotemporal 3D models reconstructed for a pear, the bitten apple, a cut mango, and a unpeeled banana. Time-lapse intervals used in our captures were 10 min for the apple and banana, and 5 min for the mango and pear. We aged the apple and banana for 7200 min or 5 days, the mango for 3600 min or 2.5 days, and the pear for 2165 min or 1.503 days. The 3D models of the bitten objects reveal enzymatic browning over time when the exposed regions undergo oxidation catalyzed by polyphenol oxidase [25]. The spatiotemporal 3D model of the banana shows appearance change from yellow with black spots to nearly full black. The 3D model of the mango shows its leaning over under the influence of gravity. The mango model also shows slight browning due to decay at exposed edges.

Volume Loss Analysis. The middle rows of Figs. 7 and 9(a)–(c) show time-plots of $\log(V/V_0)$, the logarithmic rate of change of volume V expressed as a ratio of the initial volume V_0 . We compute the volume of each 3D mesh using the approach of [31]. We estimate linear fits shown in red dashed lines to analyze the match between the data plot and models that describe fruit senescence in terms of exponential decay [17]. We use random sample consensus (RANSAC) [12] to eliminate outliers in the linear fit. While our estimation shows a close match between the exponential decay model and the data for the banana and the pear, it under-estimates the rate of volume change for the mango and the apple. As part of future work, we are interested in understanding the occurrence of super-exponential volume decay in fruit by extending the time duration of capture for the banana to include the underripe green phase and rotten black phase, and by creating spatiotemporal 3D models for a greater diversity of fruit.

Appearance Change Analysis. The bottom rows of Figs. 7 and 9(a)–(c) show plots of the mean color in each textured 3D mesh across the time-lapse sequence.

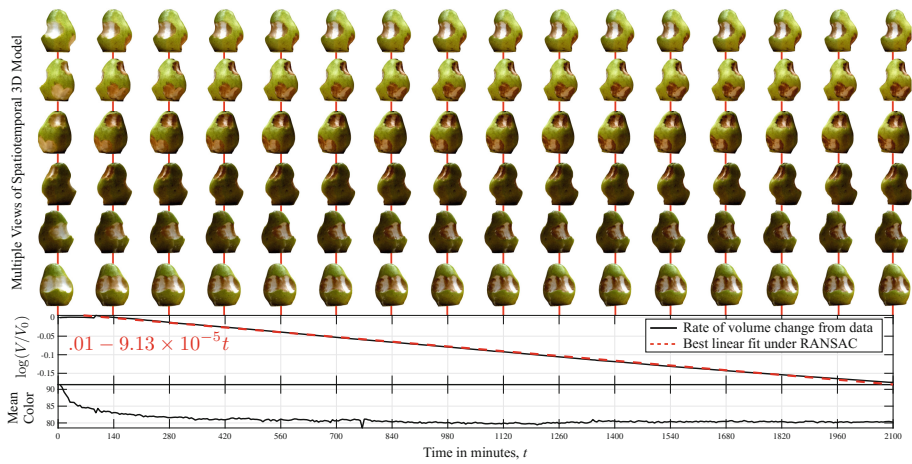


Fig. 7. Multiple views of the spatiotemporal 3D model and time plots of the log percentage of the initial volume and average color for bitten pear. (Color figure online)

For objects undergoing browning of exposed regions such as the apple and the pear, the rate of color change is high in the beginning and tapers toward the end, showing that browning occurs rapidly in the initial phases of exposure. For the banana where enzymatic browning occurs on the peel, we notice a gradual change in color in the beginning during the phase when the banana develops black spots, followed by an acceleration in color change as the spots saturate and the object becomes overripe. The mango shows a slower rate of color change over the time-lapse sequence, indicating lower levels of polyphenol oxidase compared to the apple, pear, and banana.

Brightness constancy analysis. Our 3D model deformation is based on the assumption that brightness is nearly constant over successive frames. Figure 8(a) shows a plot of root-mean-square (r.m.s.) difference in frame-to-frame appearance estimated using the textures obtained in Sect. 7. The mean value of the r.m.s. difference on a 0 to 255 intensity scale is $7.59 \pm .48$ for the apple, $7.05 \pm .60$ for the banana, $6.91 \pm .65$ for the pear, and $5.91 \pm .35$ for the mango, showing that per-frame differences in appearance are small. High values in the plot indicate



Fig. 8. (a) Plot of root-mean-square (r.m.s.) frame-to-frame appearance difference against time. (b) Frames showing interactions with a VR environment containing spatiotemporal 3D models. The user plays forward aging in the banana and pear.

lighting fluctuations, however, the slow rate of volume change introduces small point translations in the image, enabling the KLT tracker to accurately track points under slight fluctuations in illumination.

Interactions with Spatiotemporal 3D Models. The spatiotemporal 3D models created by our approach can be imported into standard 3D modeling and rendering software to create animations, games, and virtual reality (VR) environments. For instance as shown in Fig. 8(b), we use our spatiotemporal manipulation interface in VR discussed in [27] to perform spatial manipulations such as rotations, scaling, and translation and temporal manipulations such as playing, rewinding, and forwarding of changes in shape and appearance due to aging of objects.

9 Discussion

We provide an approach to reconstruct spatiotemporal 3D models of objects such as fruit that undergo aging on exposure to the environment. Our approach deforms a static 3D mesh reconstructed from photographs captured by a rotating camera track to time-lapse image sequences of the aging object captured from multiple viewpoints. The spatiotemporal 3D models provided by our approach reveal changes to shape changes due to water volume loss and appearance changes due to enzymatic browning characteristic of natural objects such as fruit.

The main limitation of our approach is that it requires manual interaction to perform precise alignment the 3D model to the initial frame. In future work, we will use matching of multi-viewpoint renders with interest point correspondences to perform automatic pixel-precise deformation of the mesh to the initial frame. Our approach retains the contribution of illumination in the texture as seen by the shadow on the surface of the pear and apple. As part of future work, we will evaluate BRDFs obtained using light probes [9] and using illumination estimation [16, 28]. Since our approach relies on approximate frame-to-frame brightness constancy to identify point correspondences using KLT tracking, it may not handle rapidly occurring changes such as popping corn kernels. We will investigate supervised learning approaches to predict point correspondences between frames with dissimilar brightness.

Our work is the first step in the creation of spatiotemporal 3D model repositories enriched with information on dynamic real-world object behaviors. As part of our future work in creating these repositories, we will expand the range of behaviors captured to include melting, weathering, and corrosion, perform data-driven estimation of physical properties, and perform controlled experiments to include the effect of variation in temperature and humidity on object aging. Our long term goal is to use spatiotemporal 3D models in performing data-driven understanding of the aging status of objects such as fruit ripeness and edibility and structural health of erosive objects from images taken by ubiquitous devices.

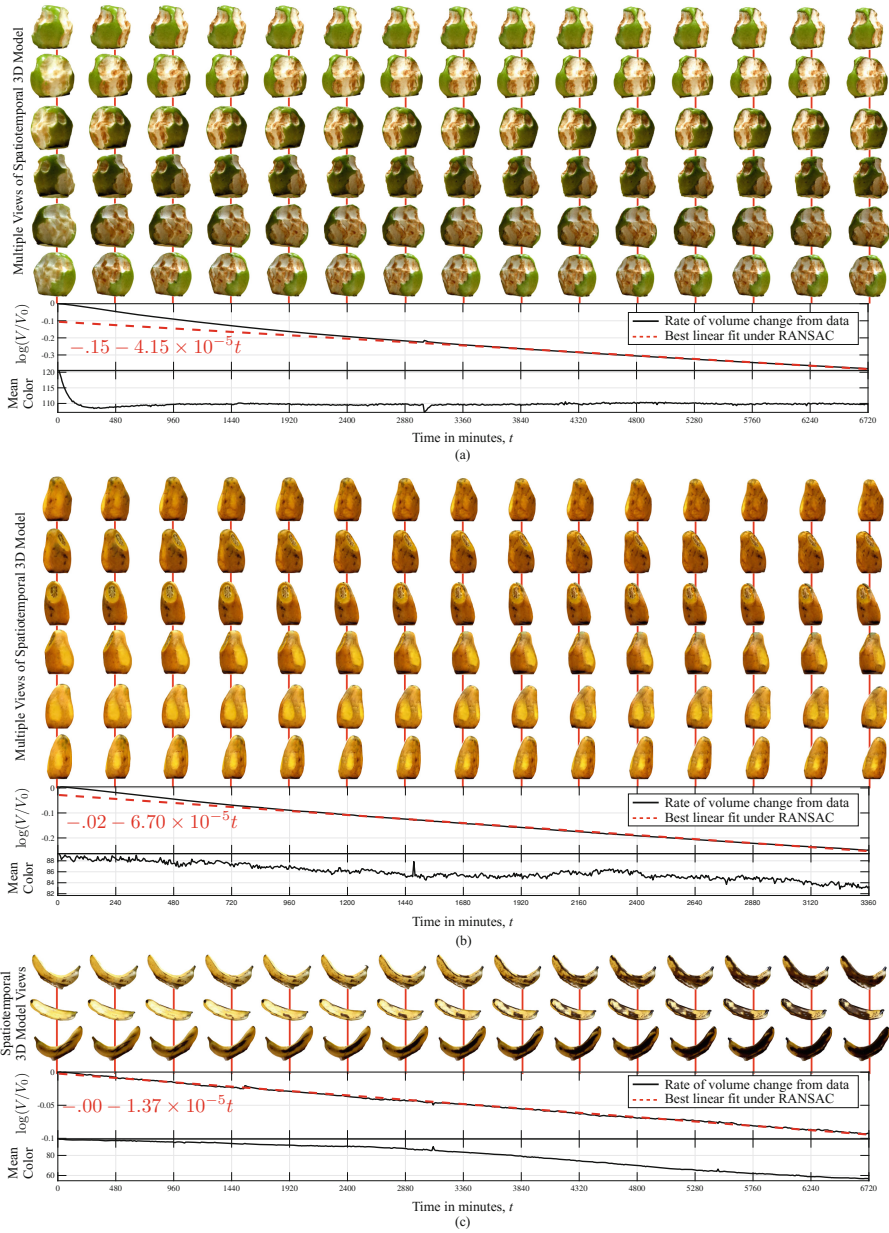


Fig. 9. Multiple views of the spatiotemporal 3D model and time plots of the log percentage of the initial volume and average color for (a) Apple (b) Mango, and (c) Banana. (Color figure online)

Acknowledgements. This work was partially supported by the National Science Foundation (NSF) grant #1730183.

References

1. Adobe: ReMake. <https://remake.autodesk.com/about/>
2. de Aguiar, E., Stoll, C., Theobalt, C., Ahmed, N., Seidel, H.P., Thrun, S.: Performance capture from sparse multi-view video. *ACM Trans. Graph.* **25**(3) (2008)
3. Baraff, D., Witkin, A.: Dynamic simulation of non-penetrating flexible bodies. In: *SIGGRAPH* (1992)
4. Baraff, D., Witkin, A.: Large steps in cloth simulation. In: *SIGGRAPH* (1998)
5. Baran, I., Popović, J.: Automatic rigging and animation of 3D characters. *ACM Trans. Graph.* **26**(3) (2007)
6. Beeler, T., Hahn, F., Bradley, D., Bickel, B., Beardsley, P., Gotsman, C., Sumner, R.W., Gross, M.: High-quality passive facial performance capture using anchor frames. In: *SIGGRAPH* (2011)
7. Chuang, M., Luo, L., Brown, B.J., Rusinkiewicz, S., Kazhdan, M.: Estimating the Laplace-Beltrami operator by restricting 3D functions. In: *CGF* (2009)
8. Collet, A., Berenson, D., Srinivasa, S.S., Ferguson, D.: Object recognition and full pose registration from a single image for robotic manipulation. In: *ICRA* (2009)
9. Debevec, P.: Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In: *SIGGRAPH Classes* (2008)
10. Dorsey, J., Pedersen, H.K., Hanrahan, P.: Flow and changes in appearance. In: *SIGGRAPH Courses*, p. 3 (2005)
11. Enrique, S., Koudelka, M., Belhumeur, P., Dorsey, J., Nayar, S., Ramamoorthi, R.: Time-varying textures: definition, acquisition, and synthesis. In: *SIGGRAPH Sketches* (2005)
12. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *CACM* **24**(6), 381–395 (1981)
13. Gu, J., Tu, C.I., Ramamoorthi, R., Belhumeur, P., Matusik, W., Nayar, S.: Time-varying surface appearance: acquisition, modeling and rendering. *ACM Trans. Graph.* **25**(3) (2006)
14. Guo, L., Quant, H., Lamb, N., Lowit, B., Banerjee, N.K., Banerjee, S.: Multi-camera microenvironment to capture multi-view time-lapse videos for 3D analysis of aging objects. In: *International Conference on Multimedia Modeling* (2018)
15. Jackson, A.S., Bulat, A., Argyriou, V., Tzimiropoulos, G.: Large pose 3D face reconstruction from a single image via direct volumetric CNN regression. *arXiv preprint arXiv:1703.07834* (2017)
16. Kholgade, N., Simon, T., Efros, A., Sheikh, Y.: 3D object manipulation in a single photograph using stock 3D models. In: *SIGGRAPH* (2014)
17. Kider, J.T., Raja, S., Badler, N.I.: Fruit senescence and decay simulation. In: *CG Forum* (2011)
18. Lamb, N., Banerjee, N.K., Banerjee, S.: Programmatic 3D printing of a revolving camera track to automatically capture dense images for 3D scanning of objects. In: *International Conference on Multimedia Modeling* (2018)
19. Langenbucher, T., Merzbach, S., Möller, D., Ochmann, S., Vock, R., Warnecke, W., Zschippig, M.: Time-varying BTFs. In: *CESCG* (2010)

20. Lepetit, V., Moreno-Noguer, F., Fua, P.: EPnP: an accurate $O(n)$ solution to the PnP problem. *IJCV* **81**(2), 155–166 (2009)
21. Li, Y., Fan, X., Mitra, N.J., Chamovitz, D., Cohen-Or, D., Chen, B.: Analyzing growing plants from 4D point cloud data. *ACM Trans. Graph.* **32**(6) (2013)
22. Lim, J.J., Khosla, A., Torralba, A.: FPM: fine pose parts-based model with 3D CAD models. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014*. LNCS, vol. 8694, pp. 478–493. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10599-4_31
23. Lu, J., Georgiades, A.S., Glaser, A., Wu, H., Wei, L.Y., Guo, B., Dorsey, J., Rushmeier, H.: Context-aware textures. *ACM Trans. Graph.* **26**(1) (2007)
24. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: *IJCAI* (1981)
25. Nicolas, J.J., Richard-Forget, F.C., Goupy, P.M., Amiot, M.J., Aubert, S.Y.: Enzymatic browning reactions in apple and apple products. *Critical Rev. Food Sci. Nutr.* **34**(2), 109–157 (1994)
26. Paquette, E., Poulin, P., Drettakis, G.: The simulation of paint cracking and peeling. In: *GI* (2002)
27. Quant, H., Banerjee, S., Banerjee, N.K.: A virtual reality interface for interactions with spatiotemporal 3D data. In: *International Conference on Multimedia Modeling* (2018)
28. Ramamoorthi, R., Hanrahan, P.: An efficient representation for irradiance environment maps. In: *SIGGRAPH* (2001)
29. Sorkine, O., Alexa, M.: As-rigid-as-possible surface modeling. In: *EUROGRAPHICS* (2007)
30. Yin, X., Fujimoto, T., Chiba, N., Tanaka, H.T.: Modeling of wood aging caused by biological deterioration. *JACIII* **22**(2), 125–131 (2008)
31. Zhang, C., Chen, T.: Efficient feature extraction for 2D/3D objects in mesh representation. In: *ICIP* (2001)