arXiv:2308.13678v1 [cs.CV] 25 Aug 2023

Textureless Deformable Surface Reconstruction with Invisible Markers

Xinyuan Li[1] Yu Ji[1] Yanchen Liu[2*] Xiaochen Hu[1]
Jinwei Ye[3] Changxi Zheng[1,2]
[1] Pixel Lab, Tencent America [2] Columbia University [3] George Mason University
https://fluorcap.github.io/

Abstract

Reconstructing and tracking deformable surface with little or no texture has posed long-standing challenges. Fundamentally, the challenges stem from textureless surfaces lacking features for establishing cross-image correspondences. In this work, we present a novel type of markers to proactively enrich the object's surface features, and thereby ease the 3D surface reconstruction and correspondence tracking. Our markers are made of fluorescent dyes, visible only under the ultraviolet (UV) light and invisible under regular lighting condition. Leveraging the markers, we design a multi-camera system that captures surface deformation under the UV light and the visible light in a time multiplexing fashion. Under the UV light, markers on the object emerge to enrich its surface texture, allowing high-quality 3D shape reconstruction and tracking. Under the visible light, markers become invisible, allowing us to capture the object's original untouched appearance. We perform experiments on various challenging scenes, including hand gestures, facial expressions, waving cloth, and hand-object interaction. In all these cases, we demonstrate that our system is able to produce robust, high-quality 3D reconstruction and tracking.

1. Introduction

In numerous applications, from animation synthesis and motion analysis to video recognition and robotic manipulation, a fundamental task is to reconstruct a 3D deformable object and track its surface correspondences over time. Typical use cases include 3D capture of hand gestures [58, 70], facial expressions [13, 21], and cloth deformation [26, 63].

To establish surface correspondences, a crucial clue is the deformable object's surface texture. A well textured surface provides rich features describing the surface's local appearance, allowing many existing methods to find surface point correspondences across multiple views of images at a time instant as well as across a sequence of images over time [4, 5, 6, 17, 22, 37, 40, 41, 42, 61]. Moreover, recent advances on deep neural networks also inspire the use of neural networks to 3D reconstruct deformable objects that are well textured [23, 43].

However, when an object has little texture or is *textureless*—such as human skin or a piece of clean paper—these existing methods all fall short. Several recent works aim to 3D reconstruct textureless deformable objects by leveraging neural networks trained on specific datasets (e.g., papers and cloths) [7, 18, 56]. Yet, they focus on surface reconstruction, unable to predict surface correspondences for tracking. They are trained on specific datasets and cross-category generalization remains challenging. Lastly, they all face a chicken-and-egg problem: to prepare a dataset to train these networks in the first place, one needs to 3D reconstruct and track deformable objects that are textureless.

Perhaps the most straightforward solution to work with textureless surfaces is by attaching optical makers to the surface, thereby explicitly introducing rich textures for 3D reconstruction and tracking. But this approach is intrusive to the object's surface, not able to preserve the object's authentic surface appearance, and thus has limited applications.

In this work, we overcome these limitations by introducing a novel type of optical markers—one that can be put on the object surface while fully preserving its appearance in normal lighting condition. Our markers are made of *fluorescent dyes*, unseen under the visible light but visible under the ultraviolet (UV) light.

To make use of these fluorescent markers, we propose a multi-camera imaging system that captures the deformable object under regular lights and UV lights in an interleaving fashion. Under UV lights, markers on the object surface are captured; they provide rich features that enable high-quality 3D reconstruction and tracking of the object's geometry. Under regular lights, markers disappear, and we are able to capture and reconstruct the object's original surface appearance.

Our imaging system can be used in a studio setting for volumetric acquisition and key point tracking, thus accel-

*Yanchen was an intern at Pixel Lab during this project

erating the process of digital asset production. It can also produce training data for neural networks that aim to infer from a conventional video the 3D shape and motion of a deformable object. One example is the neural network that tracks 3D hand gestures in a video [58]. Its training dataset must include sequences of 3D hand models and video images that capture the hand's untouched appearance. Yet, conventional optical markers, which would change the surface appearance, can not fulfill this requirement. As a consequence, while various such networks have been proposed, high-quality datasets for training them remain scarce.

We conduct experiments on three types of deformable objects, including hand, face, and cloth/paper. Testing on challenging scenes, such as hand-object interaction and gestures with crossed figures, we demonstrate that our system produces high-quality 3D reconstruction and feature tracking even in the presence of heavy occlusions. The dataset that we have captured in this work will be released to the public for research use.

In summary, we highlight the following contributions:

- We propose to apply fluorescent markers on textureless surfaces, and analyze the reflectance response of the marker on different types of surfaces.
- We develop a multi-camera imaging system with UV lights to capture the textureless surface simultaneously with and without the marker.
- We develop a template-based algorithm for 3D reconstruction and tracking correspondences of the textureless deformable surface.
- We validate our system on a variety of challenging scenes, and we will release a dataset for deformable object reconstruction and tracking.

2. Related Work

Invisible light imaging. In computational photography, the idea of invisible light imaging has been explored: use specialized light sources and sensors for emitting and capturing light outside of the visible spectrum (i.e., in infrared $> 750nm$ or ultraviolet $< 380nm$ region). This setup complements visible light imaging (between $380nm$ to $750nm$) in order to enhance imaging capacity. Some techniques combine color images with near-infrared (NIR) images for denoising [34, 59, 66], deblurring [52, 67], super-resolution [28, 54] and geometry estimation [15, 16, 65]. Notably, Wang *et al*. [60] use infrared illumination to relight faces, in order to reduce the effect of uneven face color in a video conference setting. Later, Krishnan and Fergus [34] propose the "dark flash" that uses near-infrared (NIR) and near-ultraviolet (NUV) flash light to replace the dazzling conventional flash. Their goal is to obtain high-quality images under weak ambient illumination without disturbing the human target. Inspired by their work, many methods have then been proposed to use NIR images (captured with NIR illumination) to denoise and deblur degraded color images captured under low-light conditions [29, 52, 59, 66, 67]. Other than that line of works, Blasinski and Farrell [9] propose using narrow-band multi-spectral flash for color balancing, and Choe *et al*. [15] derive an NIR reflectance model and use NIR images to recover fine-scale surface geometry.

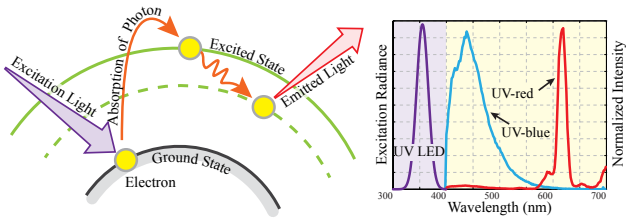
Unlike the widely studied NIR imaging techniques, ultraviolet (UV) imaging has received much less attention. The dark flash techniques [34, 59] analyze the object's spectral response under the UV light for denoising. In our work, we leverage fluorescent ink that is only visible under UV light to enrich features on object surface, without changing the object's appearance under visible light.

UV fluorescence imaging. UV light has been used to reveal substances or structures that are unseen under the visible light illumination. This is because some substances exhibit fluorescent reflectance, one that absorbs short wavelength light (*e.g*., UV light) and emit light at a longer wavelength (*e.g*., visible light) [35,47]. This phenomenon lays the foundation in a wide range of imaging applications, including forensics [44, 62], biomedical imaging [24, 33], and material analysis [57, 68].

In computer vision, the fluorescent reflection has been exploited for shape reconstruction [55], immersive range scanning [30], inter-reflection removal [19], multi-spectral reflectance estimation [10], and material classification [3], to name a few. Motivated by graphics rendering tasks, Wilkie *et al*. [64] derive a reflectance model for diffuse fluorescent surfaces. Inspired by this model are latter works on fluorescence-based reflectance analysis [31] and shape reconstruction [47]. There are also works on analyzing the spectral response of fluorescent reflection, leading to techniques for appearance separation [2,69], fluorescent relighting [20,36], and camera spectral sensitivity estimation [25].

We also leverage fluorescence. But in contrast to those existing work, we proactively introduce fluorescent markers to the object surface to enrich its surface texture. We also design an imaging system that simultaneously captures images under the UV light (to trigger the marker) and the visible light (to cancel the maker) via time multiplexing.

Deformable surface reconstruction & tracking. Our overarching goal is to recover the 3D shape and track correspondences on a dynamic deformable surface. In this direction, various sensor configurations have been explored, including the use of single camera [17, 37, 41], multiple cameras [8, 12, 48], and color cameras in tandem with depth cameras [11, 71] and with event cameras [39]. By far the most successful are methods that rely on local appearance



(a) (b)

Figure 1. **(a)** The physical principle of fluorescence. **(b)** The absorption and emission spectra of the fluorescent dyes used in our invisible markers. We use two types of dyes with the peak emission in blue (UV-blue) and red (UV-red) colors, respectively. The purple curve indicates the spectral profile of our UV LED light peaked at $365nm$.

to find feature point correspondences and then match 2D image features to a 3D shape template for surface reconstruction and tracking [4,5,14,40]. However, these methods are hampered when the surface is sparsely or repetitively textured, or has no texture at all, as in those cases, the number of feature point correspondences are insufficient and the matching quality is not reliable. When the deformable surface are weakly textured, dense pixel-level template matching may be helpful [37, 38, 41, 46]. Nevertheless, none of these methods can handle surfaces with no texture (*i.e.*, textureless surfaces)—for example, a piece of white paper.

In contrast, we are able to reconstruct high-fidelity 3D surface and track its deformation even for textureless surfaces, thanks to the markers that introduce local features while preserving the surface's original appearance.

3. Method

We now introduce the physical principle behind our fluorescent markers, their spectral response, and our imaging system that leverages fluorescent markers for robust feature tracking and surface reconstruction.

3.1. Invisible Marker

Fluorescent dyes. Fluorescent material exhibits the Stokes shift [51]—an optical phenomenon wherein the material absorbs short wavelength light, but re-emit light at a longer wavelength. This phenomenon is caused by the material molecules' quantum behavior: when electrons of a fluorescent material are irradiated by short wavelength light (*e.g.*, UV light), they enter into an excited state after absorbing the light energy, and then immediately de-excite and emit outgoing light at a longer wavelength (usually in the visible-light spectrum). This process is illustrated in Fig. 1-a.

Our invisible markers are made of fluorescent dyes, which are soluble in organic solvent (e.g., alcohol or acetone). The solution is transparent; and using it as the ink in a fountain pen, we draw dot- or line-shaped markers on the object surface. The resulting markers, due to their transparency, are invisible under the visible light, thereby preserving the object's authentic appearance. The markers emit visible light (and thus become visible) only under the UV light. Moreover, the emitted light fade immediately—often within $10^{-8}sec$—after the UV light is off. Therefore, by using the UV light to trigger the markers' emission and synchronizing the trigger with the camera shutters, we can capture images with and without the markers visible in a time multiplexing manner (see details in Section 3.2).

Figure 2. Objects covered with our fluorescent ink (UV-blue and UV-red). **(left)** Image under the visible light. **(right)** Image under the UV light, wherein the blue and red emission becomes apparent.

We look for fluorescent dyes that satisfy two more criteria: 1) the fluorescent emission under UV light has high contrast and strong visibility; and 2) the dyes are biologically safe and non-toxic to human skin. In this work, we find that the MaxMax UV dyes[1] meet our needs. In particular, we use two types of fluorescent dyes emitting blue and red lights, respectively. We refer to them as *UV-blue* and *UV-red* dyes. To prepare the fluorescent ink to draw markers, we dissolve the UV-blue dye in 70% alcohol and UV-red in acetone. Both types of markers can be triggered for emission under $365nm$ UV light (see Fig. 1-b). We choose this UV light, because it is in the range of UVA, the safest UV spectrum to human skin, abundant in natural sunlight. By using two types of dyes, we are able to use their emission colors to differentiate and segment multiple objects (see Fig. 2). This is useful in challenging scenes such as hand-object interactions.

Fluorescent emission has narrow-banded wavelengths (Fig. 1-b). Thus, the hue values it introduces on captured fluorescent images stay in a small range. Exploiting this fact, we can easily detect markers on the fluorescent images: we convert the images into the HSV color space, and label pixels whose hue values fall into the fluorescent material's emission range. This marker detection in HSV color space is robust even when the surface itself is fluorescent. This is because the narrow-banded fluorescent emission peaks at a certain spectral location, and it is very unlikely that the surface's emission, if any, peaks at the same location as our two types of markers, which by themselves have different emission peaks.

[1] https://maxmax.com/phosphorsdyesandinks

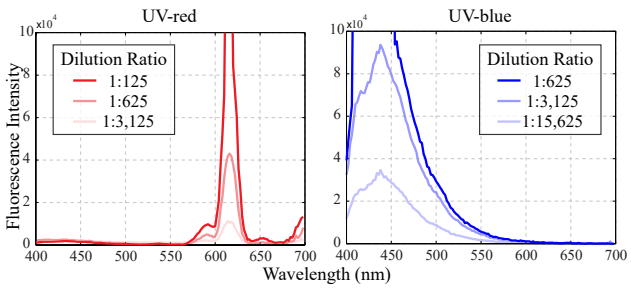


Figure 3. The spectral response of fluorescent solutions with different concentration levels of the UV-red (left) and UV-blue (right) dyes.

Spectral response. Our fluorescent ink is made by dissolving fluorescent dyes in the solvent (i.e., alcohol or acetone). Caution is needed when choosing the solution's concentration level. The higher the concentration is, the brighter the fluorescent emission under the UV light becomes. On the one hand, if the fluorescent emission is too bright, the markers will saturate image pixels, rendering the marker detection based on hue values much harder. On the other hand, if the concentration is too low, the markers appear too dim, and the detection also becomes less robust.

We choose the concentration level through systematic measurements. We test both UV-blue and UV-red dyes. For each type of dyes, we prepare the fluorescent solution of different concentrations spaced by $1/5$ dilution ratio (i.e., $1/5$, $1/25$, $1/125$, ...), covering concentration levels from $1/5$ to $1/15625$. We then measure the spectral responses of each sample using a modular multimode microplate reader (BioTek Synergy H1). Under $365nm$ UV light, the measurement records the response of fluorescent emission in the range of $400nm$ to $700nm$ (visible light spectrum) with a $2nm$ step. A subset of the measured response curves are shown in Fig. 3. Through the measurements, we find that for both types of dyes, the desirable concentration level is $1/625$, and we thereafter use this ratio throughout our experiments.

We also measure the spectral response of our markers painted on different types of surfaces, including the skin of a mouse (after hair removal), a piece of paper, and a cotton cloth. We include a mouse skin (kindly provided by a bio-medical lab) because its composition is very close to the human skin. We use a hyperspectral camera (Pika XC2) to measure the spectral response of fluorescent emission under $365nm$ UV light (see Fig. 4). The spectral response curves are shown in Fig. 5. The results indicate that each type of markers has similar spectral response on various types of surfaces, suggesting that our hue-based marker detection algorithm is applicable to a wide range of surfaces.

Biological safety. Since we will use our system to capture human hands and faces, we address the safety concerns carefully. First, the fluorescent dyes we use are non-toxic and stable under non-direct sunlight. When applying markers on human skin, we only use the UV-blue dye dissolved in 70% alcohol, because the alcohol solvent, commonly used for medical sterilization, is safe for skin contact. The UV-red dye is dissolved in acetone, which is commonly used as nail polish remover. But the acetone may irritate human skin. Therefore, we do not apply it on human skin and only use it on inanimate objects (such as paper and cloth).

UV light is present in natural sunlight—it constitutes about 10% of the total electromagnetic radiation from the sun, and about 95% of UV light is in the UVA spectrum (*i.e.*, from $315nm$ to $400nm$) [32]. But long-term exposure or short-term over-exposure to UV light can cause potential harm to human skin and eyes. To prevent from the harm, in our experiments, we strictly follow the Threshold Limit Values and Biological Exposure Indices (TLV/BEI) guidelines [1] to limit the UV illumination.

When we use our imaging system to acquire human hand gestures or faces, a typical acquisition session lasts no more than 10 minutes, during which the UV lights are turned on for only 36 seconds (see details on our illumination triggering in Section 3.2). We measure using a spectrometer the UV irradiance when our imaging system is in use, and the average UV irradiance is $24\mu W/cm^2$. As a comparison, the UV irradiance on a bright sunny day under direct sun light is about $250\mu W/cm^2$ [32], that is, around ten times of our UV light irradiance. Therefore, we believe our system will not cause any UV over-exposure risk. To be extra cautious, we always apply sunscreen cream on human hands and faces before painting the markers and exposing them to our UV lights. Also, our system operator always wears a UV protective glass with Optical Density (OD) > 6 during the acquisition process. In sum, our markers and UV lights cause no significant safety hazard, and we have taken additional safety measures for protection.

3.2. Acquisition System

System setup. We build a multi-camera system that captures surface deformation for 3D reconstruction and surface tracking. Our imaging system consists of 42 global-shutter color cameras each with a $16mm$ lens and 60 UV LED light units. Cameras and lights are uniformly mounted on a rhombicuboctahedral rig frame, facing inward to the frame center (see Fig. 6). Thanks to the rhombicuboctahedral frame being a discrete approximation of a sphere, their distances to the frame center are almost the same ($\approx 75cm$), thus avoiding uneven lighting attenuation.

All cameras capture videos with a resolution of 2448×2048 at $60fps$. They have a 30-degree field of view and a aperture of $f/5.6$ in order to capture all-focus images within

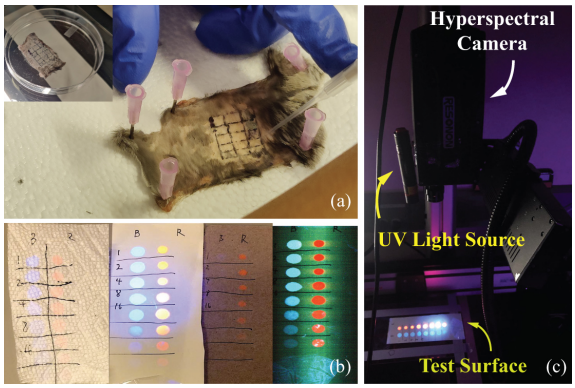


Figure 4. We measure the fluorescent spectral response on different types of surfaces. **(a)** Mouse skin sample. **(b)** Various paper and cloth samples. **(c)** Equipment for measuring the spectral response.

the range of the rig ($\approx 40cm$). In addition, we mount a $400nm$ long-pass filter in front of each cameras. The filter blocks the visible blue light component caused by the UV lights from contaminating the captured images.

Each UV LED unit has a power of $2.58W$, emitting light with a peak wavelength at $365nm$ and a beam angle of $155°$. The UV LEDs response rapidly to the triggering voltage: their rise and fall times are within $1ns$—an important trait for camera-light synchronization described next. In addition, we use 7 light panels on the top and bottom sides of the frame. These light panels are always on during a capture session in order to supply more light such that we can use a faster shutter speed and reduce captured motion blur and surface specularity.

Synchronization. We capture the deformable surface both with and without the markers visible in a time multiplexing fashion. To this end, we group the cameras into two sets: the first set (referred to as *UV cameras*) is triggered when the UV lights are turned on, and the second set (referred to as *reference cameras*) is triggered when the UV lights are off. In this way, the UV cameras capture the object with markers, while the reference cameras capture its original untouched appearance.

In practice, we use 33 UV cameras and 9 reference cameras in order to 3D reconstruct and track the markers more accurately.

Since the cameras capture at $60fps$, the time gap between two consecutive frames is $16ms$. If this time gap is much larger than the camera's exposure time, then we can shift the triggering time of the two sets of cameras within the $16ms$ time window so that their exposure time periods do not overlap with each other (see Fig. 6-a). In practice, we use an exposure time of $2ms$, although an exposure time up to $8ms$ works as well. As a result, the time difference between the frames captured by the two sets of cameras in the same time window is $2ms$. Although

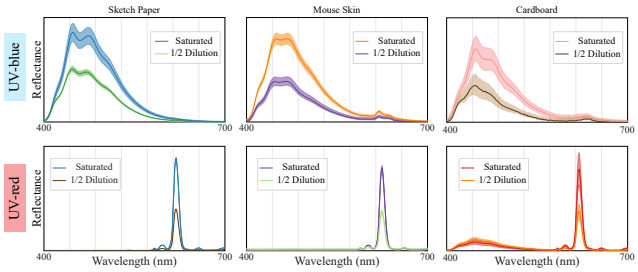


Figure 5. Spectral responses of two types of markers on sketch paper (left), mouse skin (middle) and cardboard (right).

the time difference is very short, we still use interpolation to reduce the amount of possible mis-alignment (see more details about the interpolation in Section 3.3). We custom-build an FPGA board to precisely trigger the on and off of the UV lights and synchronize the triggering events with the two sets of cameras.

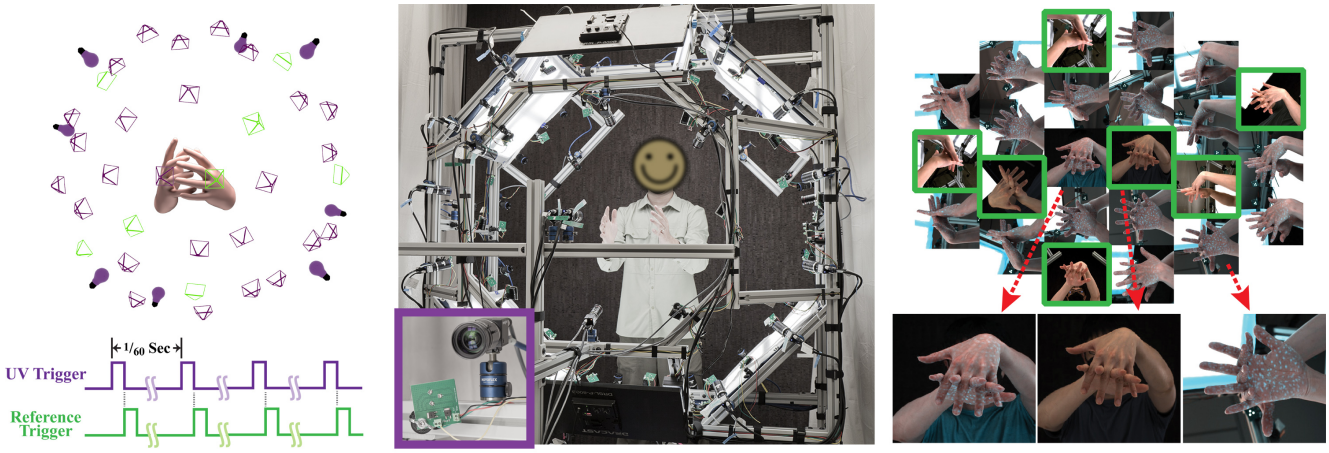
Calibration & data management. All cameras are geometrically calibrated using ChArUco patterns attached to each face of a rhombicuboctahedron. We also estimate the intrinsic and extrinsic parameters of all cameras, from which we compute the geometric transformation between each pair of cameras. This allows us to re-project images captured by one camera to another camera (with the re-projection error less than 0.6 pixels). Moreover, we chromatically calibrate all cameras through adjusting their white balance.

With 42 cameras capturing 2K images at $60fps$, we need an infrastructure to support high-speed data transfer and storage. To this end, we set up three high-performance servers, each with 16TB SSD hard drive and PCI-E to USB capture cards. On each server, we build an M2 SSD RAID system with read and sequential write speeds up to 28Gbps, and connect it to 14 cameras. In this way, we can directly save uncompressed raw images on the servers. The RAID systems can host up to 1 hour of acquired images. The data are later transferred to a NAS system for further processing.

3.3. Feature Tracking and Reconstruction

2D marker detection & 3D reconstruction. To detect markers on captured images, we convert the images into the HSV color space. The pixels corresponding to fluorescent markers typically exhibit high Saturation (S) and Value (V) values, and their Hue (H) values fall into a small range related to the dye's emission profile (i.e., $H \in [0, 15]$ for UV-red and $H \in [110, 125]$ for UV-blue). We label all the pixels that satisfy these criteria as marker pixels and detect the center position of each marker. The marker positions will be used in the subsequent step for template fitting.

Thanks to our markers, the deformable object captured by the multi-view fluorescent images has rich textures, even



(a) (b) (c)

Figure 6. **(a)** Conceptual illustration of our system with trigger scheme (green color refers to reference cameras, and purple color refers to UV cameras). **(b)** The real physical setup of our system with zoom-in view of the UV LED unit. **(c)** Samples images captured by our cameras (images with the green borders are captured by the reference cameras).

when the object itself has no texture at all. The rich textures allow us to follow the standard 3D photogrammetric reconstruction algorithm (*e.g.*, COLMAP [49,50]) to obtain a 3D point cloud for each frame in time. In practice, we compute dense disparity maps through semi-global matching [27], and use the commercial software Agisoft Metashape[2] for point cloud reconstruction.

Template fitting. Next, we fit each frame of 3D point cloud to an object-dependent predefined template. Depending on specific tasks (e.g., to capture hand gestures or facial expressions), different templates will be used, and their details will be provided in Section 4. Here we present our fitting algorithm that can be generally applied to different tasks.

Consider a 3D point cloud $\mathcal{S}$ consisting of M points $\mathbf{p} = \{\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_M\}$ and a 3D shape template described by N vertices $\mathbf{v} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_N\}$ and K faces. Our goal is to deform the shape template so that it aligns closely to the 3D point cloud. We adopt the embedded deformation graph [53] to deform the template: for every vertex i on the template shape, its deformed position is described by $\boldsymbol{v}_i + \boldsymbol{t}_i$. To ensure deformation smoothness, every vertex i also has a local region of influence. Its influence is described by a rotational matrix $\boldsymbol{R}_i \in SO(3)$, which maps any point $\boldsymbol{p}$ in its local region to the position $\boldsymbol{p}'$ according to

$$\boldsymbol{p}' = \boldsymbol{R}_i(\boldsymbol{p} - \boldsymbol{v}_i) + \boldsymbol{v}_i + \boldsymbol{t}_i. \tag{1}$$

We determine $\boldsymbol{t}_i$ and $\boldsymbol{R}_i$ (for $i = 1..N$) by solving the following optimization problem:

$$E_{\text{total}} = E_{\text{fit}} + \lambda_{\text{m}} E_{\text{marker}} + \lambda_{\text{s}} E_{\text{smooth}}. \tag{2}$$

[2]https://www.agisoft.com/

The first term E_{fit} measures the ℓ_2 distance between the 3D point cloud and the deformed template mesh in two ways: for each point j in the point cloud $\mathcal{S}$, its distance to the closest vertex and the closest face on the template. E_{fit} is thus a summation of two terms, namely,

$$\begin{aligned} E_{\text{fit}} = \sum_{j=1}^{M} \underbrace{\|\boldsymbol{v}_{c(i)} + \mathbf{t}_{c(j)} - \boldsymbol{p}_j\|_2^2}_{\text{point to vertex distance}} + \\ \beta \sum_{j=1}^{M} \underbrace{\|\boldsymbol{n}_{c(i)}^{\top}(\boldsymbol{v}_{c(j)} + \mathbf{t}_{c(i)} - \boldsymbol{p}_j)\|_2^2}_{\text{point to face distance}}, \end{aligned} \tag{3}$$

where $c(j)$ indicates the index of the deformed template vertex closet to the point $\boldsymbol{p}_j$, $\boldsymbol{n}_{c(j)}$ denotes the vertex normal, and β is a weight for balancing the two terms.

The second term E_{marker} measures the distance between the detected 2D markers and the projected marker positions from the template mesh. Consider N_f markers. Their 3D positions on the undeformed template mesh, denoted by $\boldsymbol{x}_j$ for $j = 1..N_f$, are initialized at the beginning of the capture session (see Section 4 for object-specific details of this initialization). $\boldsymbol{x}_j$ can be expressed using the barycentric coordinate $\boldsymbol{\alpha}_j = [\alpha_{j,1}\ \alpha_{j,2}\ \alpha_{j,3}]$ on the template triangle where it is located, that is, $\boldsymbol{x}_j = \sum_{k=1}^{3} \alpha_{j,k} \boldsymbol{v}_{j,k}$, where $\boldsymbol{v}_{j,k}$ $(k = 1, 2, 3)$ are the vertex positions of the template triangle. With these notations, E_{marker} is defined as

$$E_{\text{marker}} = \sum_{i=1}^{N_v} \sum_{j=1}^{N_f} w_{ij} \|\boldsymbol{p}_{ij} - \boldsymbol{\pi}_i \tilde{\boldsymbol{x}}_j\|_2^2, \tag{4}$$

where $\tilde{\boldsymbol{x}}_j$ is the j-th marker's 3D position on the deformed template mesh (i.e., $\boldsymbol{x}_j = \sum_{k=1}^{3} \alpha_{j,k}(\boldsymbol{v}_{j,k} + \boldsymbol{t}_{j,k})$), N_v

is the number of UV camera views; w_{ij} is the confidence weight for the marker j being viewed from the i-th camera. If the marker j is occluded from the i-th camera, w_{ij} vanishes. Moreover, $\boldsymbol{p}_{ij}$ is the 2D position of a marker j (if not occluded) on the i-th camera view, and $\boldsymbol{\pi}_i$ is the projection matrix of the i-th camera.

The last term E_{smooth} regulates the smoothness of the template mesh's deformation. This is where the local influence of each vertex i (and hence $\boldsymbol{R}_i$) is involved. Following a similar term defined in [53] (called E_{reg} therein), E_{smooth} encourages the mesh deformation to be locally rigid, defined as

$$E_{\text{smooth}} = \sum_{j=1}^{N} \sum_{k \in \mathcal{N}(j)} \gamma_{jk} \|\boldsymbol{R}_j \boldsymbol{v}_{kj} - \boldsymbol{v}_{kj} + \boldsymbol{t}_j - \boldsymbol{t}_k\|_2^2, \quad (5)$$

where $\boldsymbol{v}_{kj}$ is a shorthand for $\boldsymbol{v}_{kj} = \boldsymbol{v}_k - \boldsymbol{v}_j$; $\mathcal{N}(j)$ is the neighboring vertices of vertex i (here defined as the 10-nearest neighbors of i); and γ_{jk} is a weight parameter determined by distance between $\mathbf{v}_j$ and $\mathbf{v}_k$ (see more details in [53]).

When solving the optimization problem (2), we express each $\boldsymbol{R}_i$ in Lie algebra $so(3)$ and use the Levenberg-Marquardt optimizer. The optimization starts with large λ_{s} and λ_{m} values, and gradually reduces them in iterations until the optimization converges.

Feature warping. Apart from 3D reconstruction of the deformed shape using fluorescent images, we also need to align the 3D shape and features with reference-camera videos wherein markers are invisible and the object appearance remains untouched—for example, the reference video together with precisely aligned 3D shape deformation populate a training dataset useful for many deep neural networks.

But the videos captured by the UV cameras and reference cameras are not precisely synchronized (recall Section 3.2). There is a $2ms$ time difference. In many cases, this time difference is sufficiently small so that we can safely ignore it. However, when the object moves too quickly, this time difference causes a slight misalignment. We alleviate this misalignment by linearly interpolating the 3D shapes and features of two consecutive frames to the time instant at which the reference frame is captured. This strategy, albeit simple, is effective for reducing the misalignment (see experiment in Fig. 7).

4. Experiments

We perform experiments on various scenes to analyze the performance of our proposed imaging system and reconstruction algorithms.

Influence of delayed trigger. We first analyze how much mis-alignment could be caused by the trigger delay between the reference and UV cameras, and how effective our interpolation algorithm is at mitigating this artifact. This analysis is very important as our goal is to leverage the "invisible" makers in the UV view for reconstruction and tracking under the reference view.

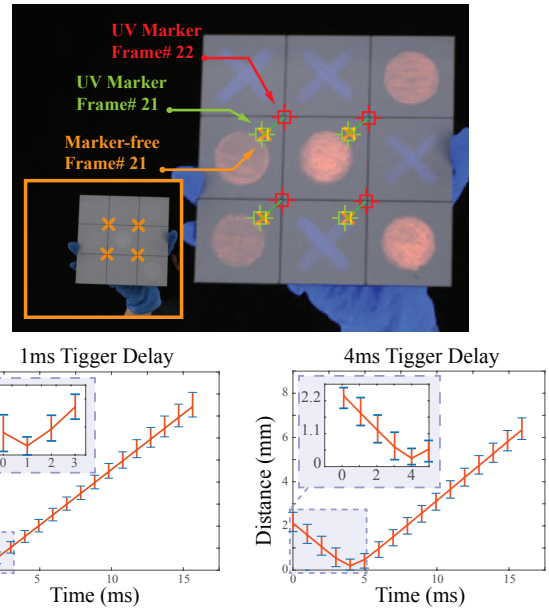


Figure 7. **(top)** Sample images of our grid target under UV and reference views. **(bottom)** Point-to-ray distance curves w.r.t. various σ values for $1ms$ and $4ms$ trigger delays.

Table 1. Average point-to-ray distances (in mm) w.r.t. σ and trigger delays

	σ	0	1	2	3	4	5	15
delay	1ms	0.49	0.20	0.56	1.06	1.56	2.06	8.12
	2ms	1.02	0.53	0.26	0.62	1.12	1.63	8.21
	4ms	2.23	1.69	1.16	0.64	0.28	0.56	7.85

We perform experiments using a board paper with grid. The grid is visible in both reference and UV views. We can therefore use the grid corners as features for measuring the pixel shift caused by the trigger delay. Sample images of the target are shown in Fig. 7. We illustrate feature points from corresponding reference and UV views (marked in orange and green, respectively), as well as those from the consecutive UV view (marked in red). We can see that features from the two consecutive UV images have apparent shift (whose time interval is $16ms$). The feature mis-alignment between the corresponding UV and reference views is slight in this case, but could vary with different motion speeds (*i.e.*, the faster the speed, the larger the shift).

We then quantitatively measure the amount of misalignment for different trigger delays, and evaluate the effectiveness of our linear interpolation algorithm described in Section 3.3. We first compute the 3D coordinates of the feature

Figure 8. **Facial expressions.** From top to bottom, we show the marker-free images, UV images with markers, tracked feature points, and recovered 3D face meshes warped on the marker-free views.

points via ray triangulation from all UV views. Assume the 3D points computed by the UV views at time t and time $t+1$ are $\mathbf{v}^t$ and $\mathbf{v}^{t+1}$, respectively. As the reference views are too sparse for accurate triangulation, we only trace out rays from features on the reference view. We use $\mathbf{r}^t$ to denote the rays traced out from reference view at time t. We then linearly interpolate $\mathbf{v}^t$ and $\mathbf{v}^{t+1}$ to obtain an intermediate point $\bar{\mathbf{v}}^\sigma$. We calculate the point-to-ray distance from $\bar{\mathbf{v}}^\sigma$ to $\mathbf{r}^t$ and use it to measure the accuracy of feature alignment in 3D.

We test on three different trigger delays: $1ms$, $2ms$, and $4ms$. We compute the point-to-ray distance with respect to various interpolation parameter σ, from 0 to 16. When $\sigma = 0$, it is equivalent as directly using $\mathbf{v}^t$ (*i.e.*, no interpolation). We test on motions of different types (*e.g.*, linear and circular), as well as different speed. The average point-to-ray distances of various configurations are reported in Table 1. We also plot the curves of point-to-ray distance with respect to different σ for$1ms$ and $4ms$ delays (see Fig. 7). The distance variation at each σ value is caused by different motions. From these measurements, we can see that when σ equals to the trigger delay, we have the smallest amount mis-alignment (the point-to-ray distances are smaller than $0.3mm$ in all cases). Our interpolation is particularly useful when the trigger delay is large. For example, in the case of $4ms$ delay, the mis-alignment is $2.23mm$ without interpolation (*i.e.*, $\sigma = 0$). Our interpolation significantly brings down the average error.

3D reconstruction and tracking results. Here we show the 3D reconstruction and correspondence tracking results on three types of scenes: facial expressions, hand interactions, and deformed paper/cloth. Video results are available in the supplementary material.

Figure 8 shows our results on capturing various facial expressions. We use the UV-blue ink to draw dot patterns on the tester's face. We apply our template-fitting algorithm to recover the 3D face mesh. Here we use the 3D face template provided by R3DS Wrap4D. Other popular face template, such as 3DMM [72], can also be used. In the first few frames, we use the rest pose (*i.e.*, neural expression) for initializing the marker positions on the template. As the rest pose can be easily registered to the template (even without using the marker guidance), we can project the fluorescent images to the registered template to initialize marker locations on the template. In the rest of the video, we will use these points to compute the marker term (Eqn. 4) for template fitting. By using the fluorescent markers, our 3D reconstruction results are more accurate. The tracked correspondences are more reliable and robust.

Figure 9 shows results of various hand scenes, including gestures and hand-object interactions. In these scenes, we use UV-blue ink to draw dot patterns on the hand and UV-red ink to paint on the object. The hue difference between the two types of ink help us separate the hand and object, which can significantly improve the template fitting. Fig. 10 shows a comparison of hand template fitting results with vs. without separating the object in hand. We can the direct fitting result (*i.e.*, without separating the object) is highly inaccurate, since the object points are also being considered in the fitting.

For hand scenes, we use the MANO [45] model as the 3D template. Similar to capturing the face, we also use the rest pose in the first few frames for registering marker points on the template. As we have multiple viewpoints and we use marker guidance, our approach can handle occlusions very well (*e.g.*, the crossed-hand scene). Again, our 3D reconstruction and correspondence tracking are highly accurate and robust (see video results). By using the interpolation scheme, our recovered 3D hand model and tracked features can align with the marker-free view very well.

Figure 11 shows results of a deformed paper. For paper and cloth, we assume they are rectangular shaped and thus use a 3D finite plane as template. Either UV-blue or UV-red in can be used to draw on these targets. In the experiment

Figure 9. **Hand interactions**. From top to bottom, we show marker-free images, UV images with markers, recovered 3D hand and object point clouds, and recovered 3D hand meshes warped on the marker-free views.

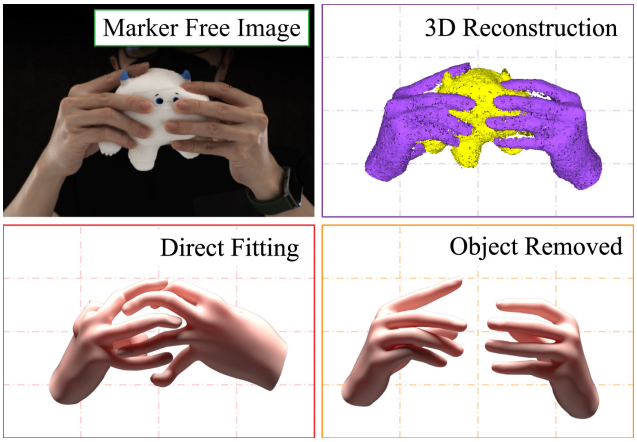


Figure 10. Comparison of hand template fitting results with vs. without separating the object point cloud.

shown Fig. 11, we draw a 13×14 grid and use their corners as features. By imposing a grid with the same size on the plane template, the feature points are easily initialized on the template mesh. Unlike the face and hand scene that require rest pose for initializing marker features on the pre-defined template, the features are directly imposed on the paper template, due to the simplicity of the template and with knowledge of the pattern. Our results demonstrate that our tracking is robust even with strong distortions. To validate the accuracy of our 3D reconstruction and view alignment, we show rendering results with novel textures on the marker-free images. We first apply the novel texture on our recovered surface mesh. We then project the re-textured surface back to the reference view to generate images with the novel texture. We can see that the distortions of the novel texture are highly consistent with the surface distortions, which indicates that both our 3D reconstruction and view alignment are reliable and accurate.

Tracking results with vs. without markers. In order to show the effectiveness of our fluorescent markers on lack-of-texture surfaces, we compare the correspondence tracking results that use our markers versus without using the markers (*i.e.*, directly apply the tracking algorithm on the marker-free reference view).

Fig. 12 and Fig. 13 show the comparison results on the paper and face scenes (video comparison results are available in the supplementary material). We use the tracking algorithm provided in the commercial software R3DS Wrap4D2. We can see that our tracked correspondences are more reliable and robust, especially on the paper scene, which lacks texture in large area. In the paper scene, the features are initialized as grid pattern. Under the marker-free view, most of the features will lose track within 20 frames (the entire video has $\sim$300 frames). In the face scene, as human skin has good amount of local features that can be tracked, the tracking result under the marker-free view are reliable for many feature points. However, feature mismatch or lose-track happens at times (see highlighted features in the zoom-in view in Fig. 13). With our fluorescent markers, the tracking result is more robust, with the mismatch rate largely reduced.

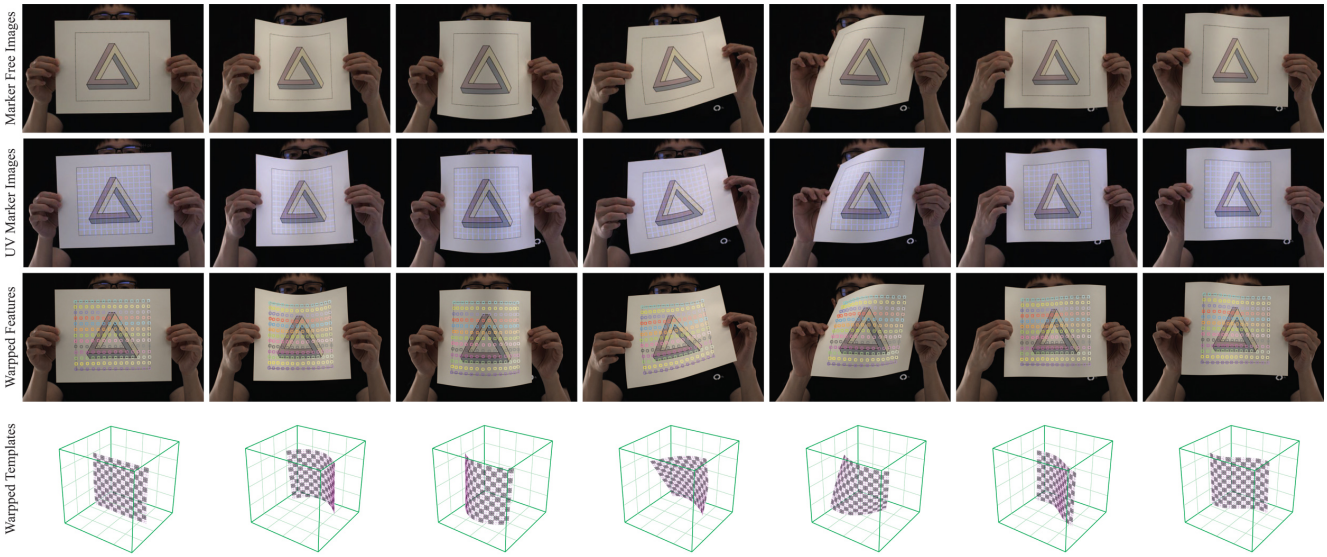


Figure 11. **Deformed paper**. From top to bottom, we show marker-free images, UV images with markers, tracked color-coded correspondences, and recovered 3D surface.

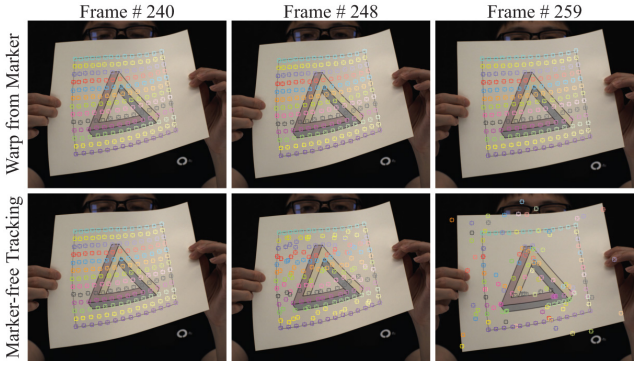


Figure 12. Comparisons of feature tracking results with vs. without using fluorescent markers (paper scene).

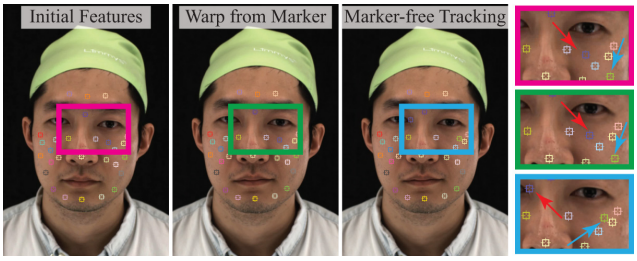


Figure 13. Comparisons of feature tracking results with vs. without using fluorescent markers (face scene). In the three zoom-in views, we highlight the features that are mis-matched in the marker-free tracking.

5. Conclusions

In summary, we have presented a method leveraging the fluorescent markers that are only visible under UV light for 3D reconstructing and tracking of deformable surfaces. In contrast to existing methods, our technique works with textureless surfaces. We developed a practical imaging system for simultaneously capturing videos with and without markers via time multiplexing. We developed a template-based algorithm for 3D reconstruction and tracking and an interpolation scheme to align the recovered 3D model and tracked correspondences back to the markerless views. We have performed experiments on challenging deformable scenes, including facial expressions, gestures, hand-object interaction, and deformed paper/cloth. Results demonstrate that by using the fluorescent markers, we can obtain high-quality 3D reconstruction and robust feature tracking. Our method can be used in a studio setting for high-quality digital media production, as well as provide datasets with ground-truths for training deformable object reconstruction and tracking algorithms.

References

[1] *2023 TLVs AND BEIs* . American Conference of Governmental Industrial Hygienists, 2023.

[2] Marina Alterman, Yoav Y. Schechner, and Aryeh Weiss. Multiplexed Fluorescence Unmixing. In *Proceedings of the IEEE International Conference on Computational Photography (ICCP)*, 2010.

[3] Yuta Asano, Misaki Meguro, Chao Wang, Antony Lam, Yinqiang Zheng, Takahiro Okabe, and Imari Sato. Coded Illumination and Imaging for Fluorescence Based Classification. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Proceedings of the European Conference on Computer Vision*, 2018.

[4] Adrien Bartoli and Toby Collins. Template-Based Isometric Deformable 3D Reconstruction with Sampling-Based Focal Length Self-Calibration. In *Proceedings of the IEEE Confer-*

ence on Computer Vision and Pattern Recognition (CVPR), June 2013.

[5] Adrien Bartoli, Yan Gérard, François Chadebecq, Toby Collins, and Daniel Pizarro. Shape-from-Template. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(10):2099–2118, 2015.

[6] Adrien Bartoli and Andrew Zisserman. Direct Estimation of Non-Rigid Registrations. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2004.

[7] Jan Bednařík, Pascal Fua, and Mathieu Salzmann. Learning to Reconstruct Texture-Less Deformable Surfaces from a Single View. In *Proceedings of the International Conference on 3D Vision (3DV)*, September 2018.

[8] Bernd Bickel, Mario Botsch, Roland Angst, Wojciech Matusik, Miguel Otaduy, Hanspeter Pfister, and Markus Gross. Multi-Scale Capture of Facial Geometry and Motion. In *Proceedings of ACM SIGGRAPH*, 2007.

[9] Henryk Blasinski and Joyce Farrell. Computational Multispectral Flash. In *Proceedings of the IEEE International Conference on Computational Photography (ICCP)*, 2017.

[10] Henryk Blasinski, Joyce Farrell, and Brian Wandell. Simultaneous Surface Reflectance and Fluorescence Spectra Estimation. *IEEE Transactions on Image Processing*, 29:8791–8804, 2020.

[11] Aljaž Božič, Michael Zollhöfer, Christian Theobalt, and Matthias Nießner. DeepDeform: Learning Non-rigid RGB-D Reconstruction with Semi-supervised Data. 2020.

[12] Derek Bradley, Tiberiu Popa, Alla Sheffer, Wolfgang Heidrich, and Tamy Boubekeur. Markerless Garment Capture. *ACM Transactions on Graphics*, 27(3):99, 2008.

[13] Chen Cao, Qiming Hou, and Kun Zhou. Displaced Dynamic Expression Regression for Real-Time Facial Tracking and Animation. *ACM Transactions on Graphics*, 33(4), July 2014.

[14] Ajad Chhatkuli, Daniel Pizarro, Adrien Bartoli, and Toby Collins. A Stable Analytical Framework for Isometric Shape-from-Template by Surface Integration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(5):833–850, 2017.

[15] Gyeongmin Choe, Srinivasa G. Narasimhan, and In So Kweon. Simultaneous Estimation of Near IR BRDF and Fine-Scale Surface Geometry. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[16] Gyeongmin Choe, Jaesik Park, Yu-Wing Tai, and In So Kweon. Exploiting Shading Cues in Kinect IR Images for Geometry Refinement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[17] Toby Collins and Adrien Bartoli. Realtime Shape-from-Template: System and Applications. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, pages 116–119, 2015.

[18] R. Danźřek, E. Dibra, C. Öztireli, R. Ziegler, and M. Gross. DeepGarment: 3D Garment Shape Estimation from a Single Image. *Computer Graphics Forum*, 36(2):269–280, May 2017.

[19] Ying Fu, Antony Lam, Yasuyuki Matsushita, Imari Sato, and Yoichi Sato. Interreflection Removal Using Fluorescence. In *Proceedings of the European Conference on Computer Vision*, 2014.

[20] Ying Fu, Antony Lam, Imari Sato, Takahiro Okabe, and Yoichi Sato. Separating Reflective and Fluorescent Components Using High Frequency Illumination in the Spectral Domain. In *Proceedings of the IEEE International Conference on Computer Vision*, 2013.

[21] Pablo Garrido, Michael Zollhöfer, Dan Casas, Levi Valgaerts, Kiran Varanasi, Patrick Pérez, and Christian Theobalt. Reconstruction of Personalized 3D Face Rigs from Monocular Video. *ACM Transactions on Graphics*, 35(3), May 2016.

[22] Vincent Gay-Bellile, Adrien Bartoli, and Patrick Sayd. Direct Estimation of Nonrigid Registrations with Image-Based Self-Occlusion Reasoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):87–104, 2010.

[23] Vladislav Golyanik, Soshi Shimada, Kiran Varanasi, and Didier Stricker. HDM-Net: Monocular Non-rigid 3D Reconstruction with Learned Deformation Model. In *Virtual Reality and Augmented Reality*, pages 51–72, 2018.

[24] Daniel C. Gray, William Merigan, Jessica I. Wolfing, Bernard P. Gee, Jason Porter, Alfredo Dubra, Ted H. Twietmeyer, Kamran Ahmad, Remy Tumbar, Fred Reinholz, and David R. Williams. In vivo fluorescence imaging of primate retinal ganglion cells and retinal pigment epithelial cells. *Optics Express*, 14(16):7144–7158, Aug 2006.

[25] Shuai Han, Yasuyuki Matsushita, Imari Sato, Takahiro Okabe, and Yoichi Sato. Camera Spectral Sensitivity Estimation from a Single Image under Unknown Illumination by Using Fluorescence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2012.

[26] Nils Hasler, Mark Asbach, Bodo Rosenhahn, Jens-Rainer Ohm, Hans-Peter Seidel, Leif Kobbelt, Torsten Kuhlen, Til Aach, and Rüdiger Westermann. Physically Based Tracking of Cloth. In *Proceedings of the11th International Fall Workshop on Vision, Modeling, and Visualization (VMV)*, 2006.

[27] Heiko Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 807–814. IEEE, 2005.

[28] Takayuki Honda, Takayuki Hamamoto, and Daisuke Sugimura. Low-Light Color Image Super-Resolution Using RGB/NIR Sensor. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2018.

[29] Takayuki Honda, Daisuke Sugimura, and Takayuki Hamamoto. Multi-Frame RGB/NIR Imaging for Low-Light Color Image Super-Resolution. *IEEE Transactions on Computational Imaging*, 6:248–262, 2020.

[30] Matthias B. Hullin, Martin Fuchs, Ivo Ihrke, Hans-Peter Seidel, and Hendrik P. A. Lensch. Fluorescent Immersion Range Scanning. *ACM Transactions on Graphics*, 27(3):1–10, August 2008.

[31] Matthias B. Hullin, Johannes Hanika, Boris Ajdin, Hans-Peter Seidel, Jan Kautz, and Hendrik P. A. Lensch. Acqui-

sition and Analysis of Bispectral Bidirectional Reflectance and Re-radiation Distribution Functions. *ACM Transactions on Graphics*, 29(4), July 2010.

[32] Muhammad Iqbal. *An Introduction to Solar Radiation*. Academic Press, 1983.

[33] Marieke E. Klijn and J́urgen Hubbuch. Application of ultraviolet, visible, and infrared light imaging in protein-based biopharmaceutical formulation characterization and development studies. *European Journal of Pharmaceutics and Biopharmaceutics*, 165:319–336, 2021.

[34] Dilip Krishnan and Rob Fergus. Dark Flash Photography. *ACM Transactions on Graphics*, 28(3), Jul 2009.

[35] Joseph R. Lakowicz. *Principles of Fluorescence Spectroscopy*. Springer, 2006.

[36] Antony Lam and Imari Sato. Spectral modeling and relighting of reflective-fluorescent scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[37] Qi Liu-Yin, Rui Yu, Lourdes Agapito, Andrew Fitzgibbon, and Chris Russell. Better Together: Joint Reasoning for Non-rigid 3D Reconstruction with Specularities and Shading. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2016.

[38] Abed Malti, Adrien Bartoli, and Toby Collins. A Pixel-Based Approach to Template-Based Monocular 3D Reconstruction of Deformable Surfaces. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 1650–1657, 2011.

[39] Jalees Nehvi, Vladislav Golyanik, Franziska Mueller, Hans-Peter Seidel, Mohamed A. Elgharib, and Christian Theobalt. Differentiable Event Stream Simulator for Non-Rigid 3D Tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2021.

[40] Dat Tien Ngo, Jonas Östlund, and Pascal Fua. Template-Based Monocular 3D Shape Recovery Using Laplacian Meshes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):172–187, 2016.

[41] Dat Tien Ngo, Sanghyuk Park, Anne Jorstad, Alberto Crivellaro, Chang D. Yoo, and Pascal Fua. Dense Image Registration and Deformable Surface Reconstruction in Presence of Occlusions and Minimal Texture. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[42] Daniel Pizarro and Adrien Bartoli. Feature-Based Deformable Surface Detection with Self-Occlusion Reasoning. *International Journal of Computer Vision*, 97:54–70, 01 2010.

[43] A. Pumarola, A. Agudo, L. Porzi, A. Sanfeliu, V. Lepetit, and F. Moreno-Noguer. Geometry-Aware Network for Non-rigid Shape Prediction from a Single View. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[44] A. Richards and R. Leintz. Forensic Reflected Ultraviolet Imaging. *Journal of Forensic Identification*, 63:46–69, 01 2013.

[45] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied Hands: Modeling and Capturing Hands and Bodies Together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 36(6), Nov. 2017.

[46] Mathieu Salzmann, Raquel Urtasun, and Pascal Fua. Local Deformation Models for Monocular 3D Shape Recovery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[47] Imari Sato, Takahiro Okabe, and Yoichi Sato. Bispectral Photometric Stereo Based on Fluorescence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2012.

[48] David S. Schacter, Mario Donnici, Evgeny Nuger, Matthew Mackay, and Beno Benhabib. A Multi-Camera Active-Vision System for Deformable-Object-Motion Capture. *Journal of Intelligent and Robotic Systems*, 75:413–441, January 2014.

[49] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[50] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016.

[51] George Gabriel Stokes. On the Change of Refrangibility of Light. *Proceedings of The Royal Society of London*, 6:195–200, January 1854.

[52] Daisuke Sugimura, Takuya Mikami, Hiroki Yamashita, and Takayuki Hamamoto. Enhancing Color Images of Extremely Low Light Scenes Based on RGB/NIR Images Acquisition With Different Exposure Times. *IEEE Transactions on Image Processing*, 24(11):3586–3597, 2015.

[53] Robert W Sumner, Johannes Schmid, and Mark Pauly. Embedded deformation for shape manipulation. In *ACM siggraph 2007 papers*, pages 80–es. 2007.

[54] Huixuan Tang, Xiaopeng Zhang, Shaojie Zhuo, Feng Chen, Kiriakos N. Kutulakos, and Liang Shen. High Resolution Photography with an RGB-Infrared Camera. In *Proceedings of the IEEE International Conference on Computational Photography (ICCP)*, 2015.

[55] Tali Treibitz, Zak Murez, B. Greg Mitchell, and David J. Kriegman. Shape from Fluorescence. In *Proceedings of the European Conference on Computer Vision*, 2012.

[56] Aggeliki Tsoli and Antonis. A. Argyros. Patch-Based Reconstruction of a Textureless Deformable 3D Surface from a Single RGB Image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, October 2019.

[57] Giovanni Verri, Catia Clementi, Daniela Comelli, Sharon Cather, and Francesca Piqueé. Correction of Ultraviolet-Induced Fluorescence Spectra for the Examination of Polychromy. *Applied Spectroscopy*, 62:1295–1302, 2008.

[58] Jiayi Wang, Franziska Mueller, Florian Bernard, Suzanne Sorli, Oleksandr Sotnychenko, Neng Qian, Miguel A. Otaduy, Dan Casas, and Christian Theobalt. RGB2Hands: Real-Time Tracking of 3D Hand Interactions from Monocular RGB Video. *ACM Transactions on Graphics*, 39(6), November 2020.

[59] Jian Wang, Tianfan Xue, Jonathan T. Barron, and Jiawen Chen. Stereoscopic Dark Flash for Low-light Photography. In *Proceedings of the IEEE International Conference on Computational Photography (ICCP)*, 2019.

[60] Oliver Wang, James Davis, Erika Chuang, Ian Rickard, Krystle De Mesa, and Chirag Dave. Video Relighting Using Infrared Illumination. *Computer Graphics Forum*, 2008.

[61] Tao Wang, Haibin Ling, Congyan Lang, Songhe Feng, and Xiaohui Hou. Deformable Surface Tracking by Graph Matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[62] Michael H. West, Robert E. Barsley, John Frair, and Fred Hall. Ultraviolet Imaging System (RUVIS) and Detection of Trace Evidence and Wounds on Human Skin. *Journal of Forensic Identification*, 40(5):249, 1990.

[63] Ryan White, Keenan Crane, and D. A. Forsyth. Capturing and Animating Occluded Cloth. *ACM Transactions on Graphics*, 26(3), July 2007.

[64] Alexander Wilkie, Andrea Weidlich, Caroline Larboulette, and Werner Purgathofer. A Reflectance Model for Diffuse Fluorescent Surfaces. In *Proceedings of the International Conference on Computer Graphics and Interactive Techniques*, November 2006.

[65] Zhihao Xia, Jason Lawrence, and Supreeth Achar. A Dark Flash Normal Camera. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021.

[66] Jinhui Xiong, Jian Wang, Wolfgang Heidrich, and Shree Nayar. Seeing in Extra Darkness Using a Deep-Red Flash. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021.

[67] Hiroki Yamashita, Daisuke Sugimura, and Takayuki Hamamoto. RGB-NIR Imaging with Exposure Bracketing for Joint Denoising and Deblurring of Low-Light Color Images. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[68] Takane Yokoi, Kenji Suzuki, and Koichiro Oba. Ultraviolet Light Imaging Technology and Applications. In *Electron Image Tubes and Image Intensifiers II*, volume 1449, pages 30 – 39. International Society for Optics and Photonics, SPIE, 1991.

[69] Cherry Zhang and Imari Sato. Image-Based Separation of Reflective and Fluorescent Components Using Illumination Variant and Invariant Color. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2866–2877, 2013.

[70] Yuxiao Zhou, Marc Habermann, Weipeng Xu, Ikhsanul Habibie, Christian Theobalt, and Feng Xu. Monocular Real-time Hand Shape and Motion Capture using Multi-modal Data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[71] Michael Zollhöfer, Matthias Nießner, Shahram Izadi, Christoph Rehmann, Christopher Zach, Matthew Fisher, Chenglei Wu, Andrew Fitzgibbon, Charles Loop, Christian Theobalt, and Marc Stamminger. Real-Time Non-Rigid Reconstruction Using an RGB-D Camera. *ACM Transactions on Graphics*, 33(4), July 2014.

[72] Michael Zollhöfer, Justus Thies, Darek Bradley, Pablo Garrido, Thabo Beeler, Patrick Péerez, Marc Stamminger, Matthias Nießner, and Christian Theobalt. State of the Art on Monocular 3D Face Reconstruction, Tracking, and Applications. *Eurographics*, 2018.