Texture Illumination Separation for Single-Shot Structured Light Reconstruction

Minh Vo, Srinivasa G. Narasimhan, and Yaser Sheikh

Abstract—Active illumination based methods have a trade-off between acquisition time and resolution of the estimated 3D shapes. Multi-shot approaches can generate dense reconstructions but require stationary scenes. Single-shot methods are applicable to dynamic objects but can only estimate sparse reconstructions and are sensitive to surface texture. We present a single-shot approach to produce dense shape reconstructions of highly textured objects illuminated by one or more projectors. The key to our approach is an image decomposition scheme that can recover the illumination image of different projectors and the texture images of the scene from their mixed appearances. We focus on three cases of mixed appearances: the illumination from one projector onto textured surface, illumination from multiple projectors onto a textureless surface, or their combined effect. Our method can accurately compute per-pixel warps from the illumination patterns and the texture template to the observed image. The texture template is obtained by interleaving the projection sequence with an all-white pattern. The estimated warps are reliable even with infrequent interleaved projection and strong object deformation. Thus, we obtain detailed shape reconstruction and dense motion tracking of the textured surfaces. The proposed method, implemented using a one camera and two projectors system, is validated on synthetic and real data containing subtle non-rigid surface deformations.

Index Terms—Single-shot, decomposition, separation, illumination, texture, mixture

1 Introduction

Structured light systems have been the method of choice to obtain dense and accurate surface models of complex objects for many industrial applications such as inspection of manufactured parts, biometrics, etc. Over the years, various types of illumination coding strategies have been developed and are mainly classified into two categories: multi-shot [13], [21], [43] and single-shot [25], [42], [48]. Multi-shot methods can estimate per-pixel depth map for various types of objects using temporal coding of the illumination patterns but require the scene to be stationary during the image acquisition. By decoding the spatial structure embedded in the illumination pattern, single-shot methods are applicable to dynamic objects but generate low spatial resolution reconstructions. Moreover, single-shot systems are susceptible to high frequency textured objects which distorts the appearance of the light pattern (see Fig. 1 for examples). Such different requirements of multi-shot and single-shot methods lead to a trade-off between the spatial and the temporal resolution of the 3D shape and is hampering wider success of structured light systems.

Another well-known limitation of the current structured light approaches is the prohibition of the concurrent use of multiple projectors because of the complex appearances in the mixed illumination regions. With the emergence of the virtual reality and tele-presence engines [5], [32], there is an increasing need for surround structured light systems where the use of multiple projectors is a must. Unfortunately, instead of producing high spatial resolution reconstructions in the overlapping regions, large holes are obtained. Currently, the problem is mitigated using additional hardware [6]. However, this approach alters the calibrated extrinsic between individual units and may require online recalibration of the whole system to integrate individual 3D shapes into one piece.

In this work, we present a structured light system[1] that deals with the two aforementioned issues. First, despite being a single-shot method, it can estimate high spatial resolution shape information of highly textured objects. Second, it allows concurrent use of multiple projectors without additional hardware. The proposed method is single-shot in the sense that it does not use the illuminated images in the previous frames to temporally decode the illumination pattern. The projector generates the necessary high frequency illumination patterns for dense shape estimation.

Due to the use of multiple projectors, the observed image not only contains the mixed appearance between the surface texture and the illumination but also the mixture of the illumination from different projectors and their combinations. Our solution follows the analysis-by-synthesis approach where possible combinations of the surface texture and the projected illumination patterns are used to recreate the observed mixed appearance image. More concretely, we develop an optimization framework that reliably estimates warps of the illumination patterns visible in the observed image regions and a reference texture template to compose the observed image. The texture template is obtained by periodically interleaving the projection sequence with an all-white pattern. The warping functions are computed starting from a sparse set of correspondences between the camera and the projectors. These initial matches are

• *The authors are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: {mpvo, srinivas, yaser}@cs.cmu.edu.*

Manuscript received 9 Jan. 2015; revised 6 May 2015; accepted 27 May 2015. Date of publication 9 June 2015; date of current version 13 Jan. 2016. Recommended for acceptance by R. Yang.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2015.2443775

1. http://www.cs.cmu.edu/~ILIM/projects/IL/TextIllumSep

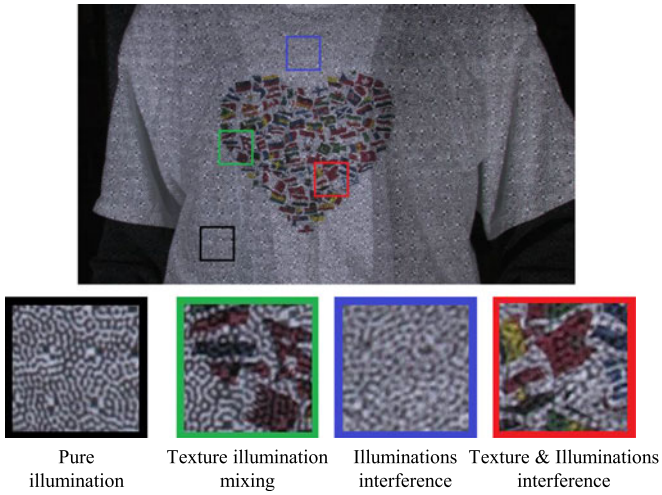


Fig. 1. Conventional single-shot structured light systems often fail for highly textured objects. For multiple projectors system, problematic effects such as the mixture of albedo variations and the interference between high frequency illumination patterns arises, which makes it difficult to establish reliable and dense camera-projector correspondences.

greedily propagated into textured areas where spatial correspondences cannot be directly estimated. Our separation scheme allows the use of multiple projectors to accurately capture the 3D shape of large objects. Fig. 1 shows a typical example of surface texture that our method can handle.

As a result of the separation, the dense correspondences between projectors and camera in the textured, and (or) mixed illumination regions are established. Subsequently, these correspondences are triangulated to reconstruct the high resolution 3D shape of the object. In addition, we obtain dense tracking inside the texture region in the presence of the illumination pattern.

Since the method computes warps to a template, it does not exhibit drift over time. Moreover, unlike conventional single-shot structured light systems whose performances degrade as surface texture frequency increases, our method achieves better decomposition accuracy with higher frequency texture. Finally, despite being presented for single-shot approaches, this method can also be used in conjunction with multi-shot systems by spatially modulating the illumination patterns with a random pattern [49]. Structured light systems using our decomposition method can take advantage of the consumer trends in resolution increases for both camera and projector and thus, avoid the need for expensive high quality time-of-fight sensors.

The proposed method is implemented on a structured light system consisting of one camera and two projectors, where four possible types of appearance could be observed: pure illumination from individual projector, illumination from two projectors, surface texture and illumination from individual projector, and illumination from two projectors and surface texture. We demonstrate dense and accurate decomposition and reconstruction results on both synthetic and real data with non-rigidly deforming objects.

2 Related Work

Active illumination for shape estimation mainly focuses on designing coded patterns that are robust to occlusion, depth discontinuity [7], [9], [36], [43], [51], defocus [2], or global illumination [10], [11], [13], [18], [19], [20]. A thorough review of the current state of the art techniques is given in [35]. Generally, the structured light system has to select appropriate illumination patterns depending on its temporal or spatial resolution requirements. Since multi-shot methods can robustly generate high spatial resolution but require stationary scenes, motion compensation schemes have been developed to handle slowly moving objects [46]. Another common approach is to interleave the patterns for structure estimation with patterns optimized for computing motion [26].

In contrast with multi-shot approaches, single-shot methods sacrifice spatial resolution for high temporal resolution reconstruction and usually cannot deal with textured objects. In the context of single-shot structured light reconstruction, the problem of textured objects is almost unexplored with the exception of the work from Koninckx et al. [27], [28]. Their methods rely on a feedback loop that changes patterns according to the error in the decoding process and heuristic rules to set the color codes depending on different textured surfaces. In principle, this method treats the surface texture as a nuisance and designs illumination patterns robust to the texture. Conversely, our method considers the texture as an additional source of information that needs to be recovered along with the 3D shapes.

While single camera-projector systems are most popular, there have been few attempts to deploy the multiple systems at once because of the distorted appearance of multiple illumination patterns superimposed on each other. Recently, Furukawa et al. [16] present a two projectors-one camera system that projects de Bruijn stripe pattern of different colors from different projectors to reduce the interference between those patterns. Sagawa et al.[34] extend the idea further and obtain higher resolution shape by interpolating the de Bruijn sequence. Inherit from these works, a surround structured light system is introduced [24]. Due to the use of sparse color grid, all these methods produce sparse shape reconstruction and are not robust to color objects. Another notable surround structured light reconstruction is presented by Lanman et al. [29] where the planar mirrors are elegantly arranged to simulate the projector patterns coming from different views. There are also efforts to create the surround 3D reconstruction systems for immersive visualization and tele-presence using the Kinects I depth sensor [5], [6], [32]. Generally, Kinect-based methods rely on hole filling or smoothing algorithms [32] and external mechanical vibrator to make their light patterns blurred when viewed by the other Kinects [6].

Although the theme of image separation has not been applied to structured light system, its history dates back to the early 90s when researchers extract motion between different depth layers from image sequences [22], [23], [45], [47]. Their key observation is that the background and the reflection layers undergo different motions due to their different distances to the transparent layer, and hence, can be separated. These methods require multiple images with motions of various directions and they only focus on motion separation, not image layer restoration.

The problem of image restoration is later investigated under the scope of Blind Source Separation [12]. A typical scenario is when the object is seeing through transparent

surfaces such as a window glass. The solutions to this problem usually assume a fixed camera and stationary scene and employ additional hardware such as polarization filters along with general statistic tools such as Independent Component Analysis [14], [39], Blind Deconvolution [37], [38], or image statistic such as image gradient sparsity [17].

A similar type of image separation is performed while estimating intrinsic images [15], [30], [40]. Unlike the multi-layer blind separation framework, intrinsic image has only two components: reflection and shading. Prevailing intrinsic image decomposition algorithms assume that natural images contain piecewise-smoothness reflectance and smoothly varying shading. Recently, by exploiting the difference in the distribution of the two intrinsic components, Yu and Brown [50] present a fast separation algorithm that produces state of the art results with a single image. In contrast with these methods, our work decomposes the high frequency illumination patterns and texture from the observed mixed appearance image. Such high frequency mixture breaks the smoothness assumption of both transparent/semi-reflection removal and intrinsic images.

Another image separation direction is the direct/global separation framework [1], [33] where the extracted direct component can be used for structure recovery techniques [18]. Our method is different from these approaches in the sense that it is a single-shot system and it does not co-locate the camera and the projector. The preliminary version of our work appears in Vo et al. [44].

3 Decomposition Framework

3.1 Mathematical Formulation

Consider an object being illuminated by one or more projectors. The brightness $I(x, y)$ at location (x, y) in the observed image is modeled as a multiplication of a texture image $I_T(x, y)$ and the illumination image $I_L(x, y)$ at that location:

$$I(x, y) = I_T(x, y)I_L(x, y). \tag{1}$$

The texture image I_T is the image observed if the projector illuminates an all-white pattern on the object. The illumination image I_L is the incident lighting pattern. This equation can explain different types of mixed appearance: illumination light from one projector onto textured surface, illumination from multiple projectors onto a textureless surface, or illumination lights from multiple projectors onto a textured surface. When the object is being illuminated by multiple light sources, the incident light is the combination of all lighting coming from individual projectors.

Equation (1) is ill-posed because it has more unknowns, e.g., $I_T(x, y)$ and $I_L(x, y)$, than equations. Thus, we must rely on additional knowledge of the illumination and texture source. Since the illumination image is a projection of the known projector patterns, these patterns serve as our referenced templates. The reference texture template can be obtained by interleaving the projection sequence with an all-white pattern.

As the object deforms, the appearance of both the texture image I_T and the illumination image I_L change accordingly. This suggests that we must be able to warp the texture template and projecting patterns to the current their current hidden appearances in order for Equation (1) still to hold true. Because image deformation is high-dimensional and non-linear, analytic forms that describe consistent deformation behavior over the entire image do not exist. Thus, we model the warping function locally within a small patch to approximate for these distortions. More specifically, we employ an affine warping function f, a constant gain a_T and an offset b_T to map patches in the texture template T to the texture image I_T:

$$I_T(x, y) = a_T T(f(x, y)) + b_T, \tag{2}$$

The constant terms, a_T and b_T, are defined patch-wise and help compensating for the intensity changes between the two images due to changes in surface normal, light directions and ambient illumination. The warping function f is written as:

$$f(x_k, y_l) = \begin{bmatrix} x_0 + p_0 + p_2 k + p_3 l \\ y_0 + p_1 + p_4 k + p_5 l \end{bmatrix}, \tag{3}$$

where (k, l) are the row and column of the texture template patch centered at point (x_0, y_0), and $p_{0..5}$ are the warping parameters.

Unlike the texture template warping where the deformation between two time instances can be small, the appearance of a patch in the illumination pattern is quite different when observed in the camera image due to perspective transformation. Assuming photometrically calibrated projectors, we adopt a homography warping g, a set of constant gain a_{Li}, and offset b_{Li} to transform the illumination pattern L_i to the illumination image I_L:

$$I_L(x, y) = \sum_{I_l=1}^{N} a_{Li} L_i(g_i(x, y)) + b_{Li}, \tag{4}$$

where N is the number of projectors visible at pixel x, y in the camera image. The role of a_{Li} and b_{Li} are also to compensate the intensity changes between the illumination patterns and the illumination image. Empirically, we found the homography warping more robust than affine warping used in Vo et al. [44]. This warping function h is expressed as:

$$g_i(x_k, y_l) = \begin{bmatrix} \dfrac{x_0 + q_{i0} + q_{i2} k + q_{i3} l}{1 + q_{i6} k + q_{i7} l} \\ \dfrac{y_0 + q_{i1} + q_{i4} k + q_{i5} l}{1 + q_{i6} k + q_{i7} l} \end{bmatrix}, \tag{5}$$

where (k, l) denotes the row and column of the texture template patch centered at point (x_0, y_0), and $q_{i0,..i7}$ are the warping parameters of the ith projector.

We minimize the following cost function over a patch of size $(2M + 1) \times (2M + 1)$ centered at the decomposing point (x_0, y_0) for the warp parameters $p_{0..5}$, $q_{i0..i7}$ and their photometric compensation coefficients a_T, b_T, a_{Li}, b_{Li}:

$$\sum_{k=-M}^{M} \sum_{l=-M}^{M} \left[I_T(x_k, y_l) I_L(x_k, y_l) - I(x_k, y_l) \right]^2, \tag{6}$$

Once the optimization converges, we quantify the decomposition by the normalized cross correlation (NCC) score of the patch in the observed image and the patch synthesized by Equation (1). A decomposition is deemed successful if the ZNCC score is greater than predefined threshold, set to 0.9 for all experiments.

Decomposition complexity. The complexity of Equation (6) grows linearly with the number of projectors. For a typical setup of two projectors and one camera, at every patch, the

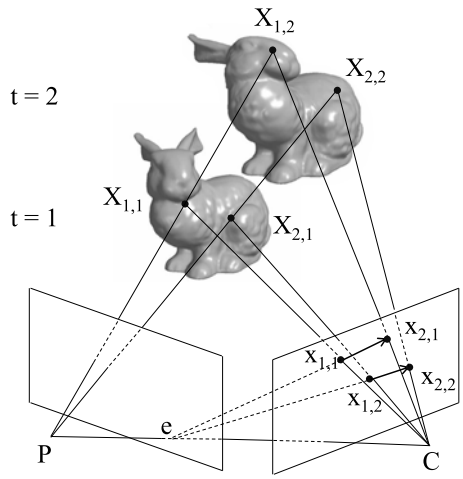


Fig. 2. Illumination flow constraint. Two rays emanating from the projector P hit the object at $[X_{1,1}, X_{2,1}]$ and $[X_{1,2}, X_{2,2}]$ at time instance 1 and 2, respectively. Their corresponding projection to camera C are $[x_{1,1}, x_{2,1}]$ and $[x_{1,2}, x_{2,2}]$. The illumination flow vectors, $\vec{x}_{1,1} - \vec{x}_{1,2}$ and $\vec{x}_{2,1} - \vec{x}_{2,2}$, intersect at the epipole e of the projector. This geometry constraint reduces the dimensionality of the illumination flow by one.

number of unknowns are 17 (eight parameters for I_T and nine parameters for I_L) for texture and one illumination pattern mixing, 18 parameters for two projector illumination patterns mixing, and 26 (eight parameters for I_T and 18 parameters for I_L) for texture and two illumination patterns mixing. For our local decomposition scheme, the ratio of the number of constraints, i.e., the number of pixels in a patch, over the number of unknowns is small, which makes solving for these warp parameters challenging. Thus, proper exploitation of additional geometric constraints and a good initial warp estimation are crucial to the decomposition scheme.

3.2 Illumination Flow Constraint

For a dynamic object illuminated by a stationary structured light system, its observed image changes at every time instance. Besides the motion of the textured surface observed on this image, there is also an apparent motion due to the illumination pattern. Interestingly, the flow direction of the illumination pattern and the texture are radically different from each other. Unlike the motion flow which can be in any directions according to the movement of points on the object surface, the direction of individual illumination flow is only along specific lines. Fig. 2 illustrates this phenomenon. Because the illumination flow lies on the projection of the ray emanating from the projector, their directions must be along the epipolar lines of the camera-projector system. Furthermore, the illumination flow field encodes the changes in depth of rays emanating from the light source that hit the object. Because establishing spatial correspondences between camera and projector is much more difficult than estimating temporal correspondences, especially in the wide baseline scenario, any structured light systems can gain benefit from the temporal coherency of the illumination flow for shape reconstruction.

For our scenario, the illumination flow constraint allows us to simplify Equation (5) to:

$$g_i(x_k, y_l) = \left[\begin{array}{c} \frac{x_0 + q_{i0} d_x + q_{i1} k + q_{i2} l}{1 + q_{i5} k + q_{i6} l} \\ \frac{y_0 + q_{i0} d_y + q_{i3} k + q_{i4} l}{1 + q_{i5} k + q_{i6} l} \end{array} \right], \quad (7)$$

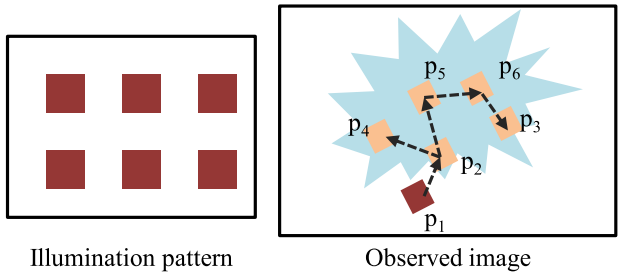


Fig. 3. Correspondence propagation. The blue textured region is illuminated by the square pattern. Since the appearance of the squares in the textured region have changed, direct matching between the observed image and the projector pattern is not possible. The correspondences are propagated from the p_1, whose match in the projector pattern is easily found, to other squares inside the textured region. For the propagation path showed by dotted arrows, p_1 is the immediate parent of the child p_2, p_2 is the immediate parent of the child p_4, and etc.

where the center of the patch (x_0, y_0) is forced to lie on the epipolar line parameterized by normalized direction vector (d_x, d_y) of that line in the projector image. Compared to Equation (5), for each projector, this equation reduces the dimensions of the optimization in Equation (6) by one and allows us to search for the decomposed illumination patch only along the epipolar line.

4 Greedy Correspondence Growing

Due to perspective distortion in the illumination image I_L and the large deviation from the texture template T for fast moving objects, good initial guesses for the warping parameters are required to optimize Equation (6). Even with random patterns [13], the local minimum of the cost function still exist due to the repetitive nature of the spatial encoding illumination pattern. We solve the initialization problem greedily. Starting from a few correspondences established in pure illumination or low textured regions, the results are gradually expanded to the mixed regions by optimizing for the warp parameters in Equation (6). Fig. 3 gives an overview of this propagation process.

4.1 Types of Mixed Appearance

Solving for the warp parameters in Equation (6) requires knowing the mixture type of the decomposed point. Let the point being decomposed be the "child" and the point initializing the child's warping parameters be the "parent". The choice of possible appearance mixture of the child point depends on the type of mixture of its immediate parent in the propagation chain. We allow at most one change in the mixture's component with respect to its parent. For example, for the case of two projectors and one camera, if the parent composes of the light from one projector and the texture surface, its child can only be created from the pure illumination from that projector, the illumination and the texture, or mixture of two illumination patterns and the textured surface. Although this assumption can be violated such as when the illumination and texture boundaries are at the same location, empirically they rarely happen and hence, are ignored. Because evaluating all the possible choices of the child mixture type is computationally expensive, we give more priority in the processing of the mixture type that is the same as the parent first. Table 1 shows all allowed mixture type for this structured light configuration.

TABLE 1
Determination of Mixture Type

Parent type	Possible child type
L_1	L_1, TL_1, L_1L_2
L_2	L_2, TL_2, L_1L_2
L_1L_2	$L_1, L_2, L_1L_2, TL_1L_2$
TL_1	L_1, TL_1, TL_1L_2
TL_2	L_2, TL_2, TL_1L_2
TL_1L_2	$TL_1, TL_2, L_1L_2, TL_1L_2$

TL_1: texture and projector 1 mixture
TL_2: texture and projector 2 mixture
L_1L_2: projector 1 and 2 mixture
TL_1L_2: texture, projector 1 and 2 mixture

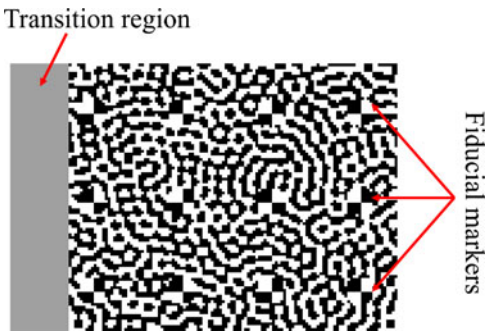


Fig. 4. Part of the illumination pattern. The fiducial markers are embedded into the random pattern to provide a sparse set of correspondences between the camera and the projector.

4.2 Initialization of Warp Parameters

We initialize the warping parameters in two steps.

Step 1. Grow a sparse set of correspondences between the camera and the projectors using a greedy correspondence growing algorithm [8]. This greedy growing strategy and the use of a random illumination pattern allows us to establish dense correspondences everywhere except for the mixed appearance regions.

Step 2. For a pixel that is close to the mixed region boundary, we exhaustively search its local neighbors for patches on the illumination patterns and texture template that minimizes the cost defined in Equation (6). The choice of possible combination of patches are given from Section 4.1. The mixed boundary is defined as locations where the direct correspondences between camera and projectors computed in Step 1 fail.

Depending on the motion of the objects, the range of the exhaustive search is set apriori. Since the deformation between the template patches and the ones in the pure texture and illumination images could be large, we pre-warp them using the warping parameters of the parent point before examining their contribution to the cost function for different hypothesized mixture types.

4.3 Optimization

The initial warping parameters obtained in Section 4.2 are refined by minimizing Equation (6) with a standard Gauss-Newton method. Given the complexity of the decomposition, two heuristics are applied to ensure accurate and convergence of the optimization.

Coordinate descent. Even with good initial guess, optimizing Equation (6) could be challenging due to its high dimensionality, e.g., two illumination patterns and the textured surface mixing requires optimizing for 26 parameters per patch. We mitigate the problem by the coordinate descent scheme where we alternate between optimizing for the warping parameters of one source, say texture, and fixing the rest, say illumination pattern one and two.

Coarse to fine decomposition. Since both the texture and illumination images have high spatial frequencies, a coarse to fine scheme of multiple Gaussian pyramid results in severe aliasing artifacts. However, since the texture surface usually has lower frequency content than the illumination, the texture image must be decomposed at an appropriate scale to avoid drifting in low frequency textured regions. Thus, we try different sizes of the local patch, arranged in descending order, to handle the differences in scale of texture image. The process is terminated as soon as a certain patch size yields successful decomposition.

5 Results on Synthetic Dataset

The performance of our approach is validated on synthetic cloth sequences containing a range of texture frequencies. The cloth composed of 64,000 vertices is generated using the OpenCloth engine [31] and can deform in subtle and non-rigid ways. Its physical size is $750{\times}400\ mm^2$ and is placed approximately 1250 mm away from the camera, which resembles our real experimental setup. The non-rigidity of cloth makes dense decomposition and shape reconstruction challenging. The all-white pattern is projected once in 30 frames. For all of our experiments, this interleaving interval gives good trade-off between the temporal resolution of the results and the moving speeds of the objects.

We set the camera and the projector resolution to 1920×1080 and 1280×800, respectively. The decomposed patch size is varied from 21×21 to 15×15 with a step size of 2 to account for different scale of the surface texture. We split the region of interest into sub-regions and independently execute them to take advantage of the multi-core architecture of modern computer. Since the runtime of the algorithm depends on the size of the patch and the size of the image to be decomposed, we report the runtime in points per second unit. Decomposing texture-illumination, illumination-illumination, and illumination-illumination-texture take on average 2,009, 2,152, 1,535 points every second on a Quad-core i7 CPU, respectively. The 3D shapes are estimated by triangulating the correspondences obtained after the decomposition. The results are presented without any post-processing.

For a succinct description of the result, we use the following notations: TL_i means texture and illumination pattern ith mixing; L_1L_2 means two illumination patterns mixing ; TL_1L_2 means texture and two illumination patterns mixing.

5.1 Illumination Pattern

Fig. 4 shows one of our static bandpass random binary illumination pattern [13]. The size of the speckle in this pattern can be tuned to provide suitable contrast for illuminating objects of different size. Fiducial checkerboard markers are uniformly seeded at every 32 pixels inside this pattern to provide set of sparse spatial correspondences. These correspondences

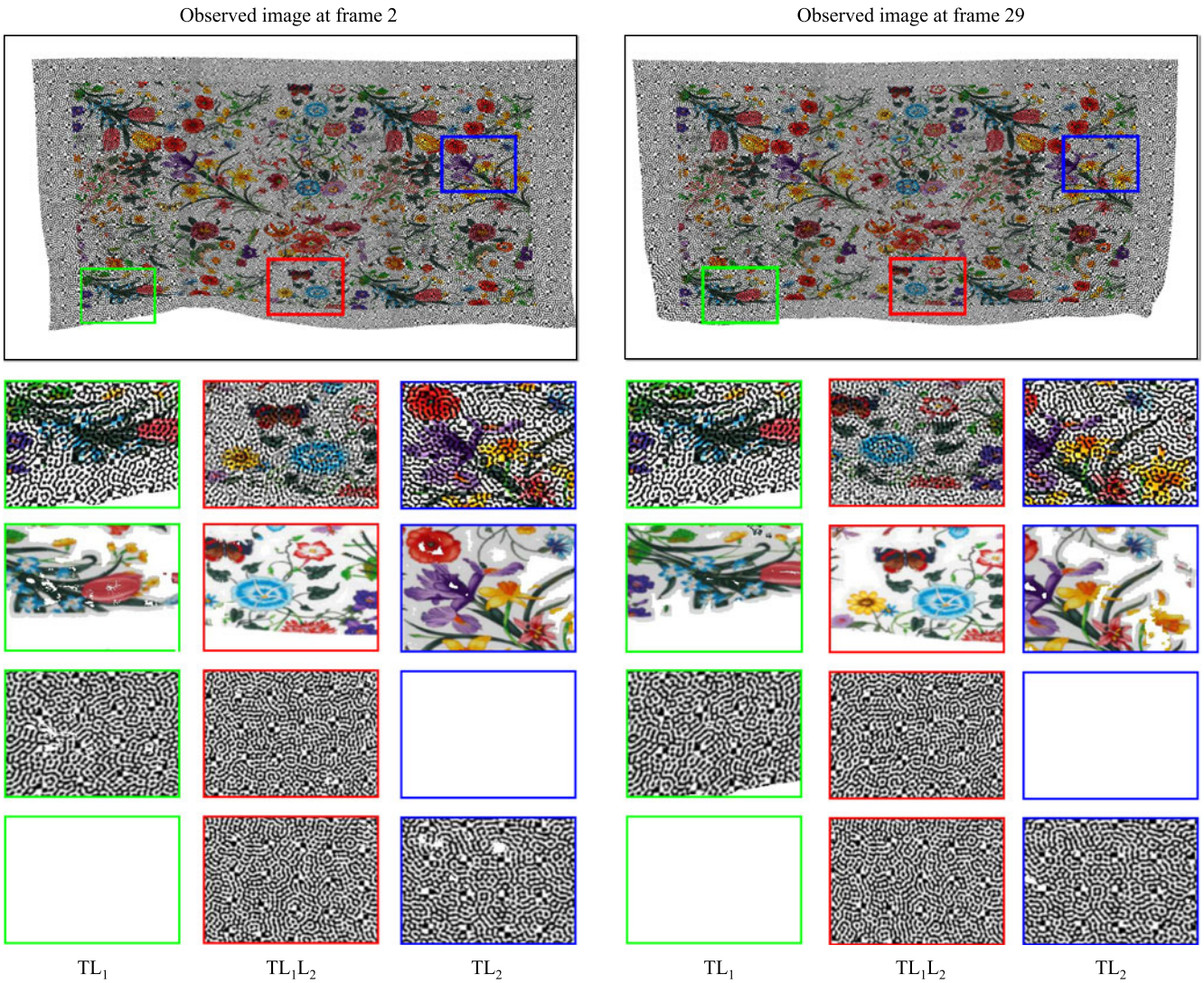


Fig. 5. Decomposition of the texture and illumination images from a synthetic flower cloth sequence with interleaving pattern projected every 30 frames. We show insets of three types of mixture: texture-illumination mixing, illumination interference, and texture and illumination interference. The same regions of the observed images are magnified to show how its intricate appearance changes over time. From the first row to the fourth row are the magnified regions of the observed image, the estimated texture, the estimated illumination for projector 1, and the estimated illumination for projector 2. Note that the small white spots in decomposed illuminations, an indication of decomposition failure, at frame 2 are not expanding and have been fixed in frame 29.

are computed using template matching along epipolar lines. Because the distance between these markers is usually magnified in the camera image, these markers do not cause ambiguities in the propagation process. For ease of implementation, a transition region is added to the side of the pattern to avoid the boundary issue when only a partial illumination patch lies on the illumination pattern, which commonly happens when multiple light patterns are mixed.

5.2 Qualitative Evaluation

We show our decomposition results on different texture frequency cloth: flower in Fig. 5 and bear in Fig. 6. The bear texture is very similar to the illumination pattern. This extreme case violates the assumptions of any methods powered by independent component analysis or smoothness prior. Thus, such methods are not applicable. The small but noticeable defects in the decomposition of frame 2 do not expand during the greedy propagation and are fixed in frame 29. Hence, there is little-to-no drift in the estimated warping functions. Visually, the method achieves better decomposition for higher frequency texture as there are less visible defects in the highly textured bear than the flower cloth.

5.3 Quantitative Evaluation

Fig. 7 shows our depthmap estimation of the flower and bear cloth at frame 29th, respectively. Without our separation, the correlation score between the projector patch and its corresponding patch in the observed image is very low in the mixed appearance regions. The depth estimated in these regions is eliminated, which results in large holes in the 3D shape. By separating the pure texture and illumination images, the proposed method accurately estimates the 3D shape with the exceptions of places where strong foreshortening occurs. For a particular point, its depth is retained only if its reprojection onto the projectors and camera images is less than 1 pixel error.

To better quantify the decomposition error, we compare the correspondences obtained after the decomposition with

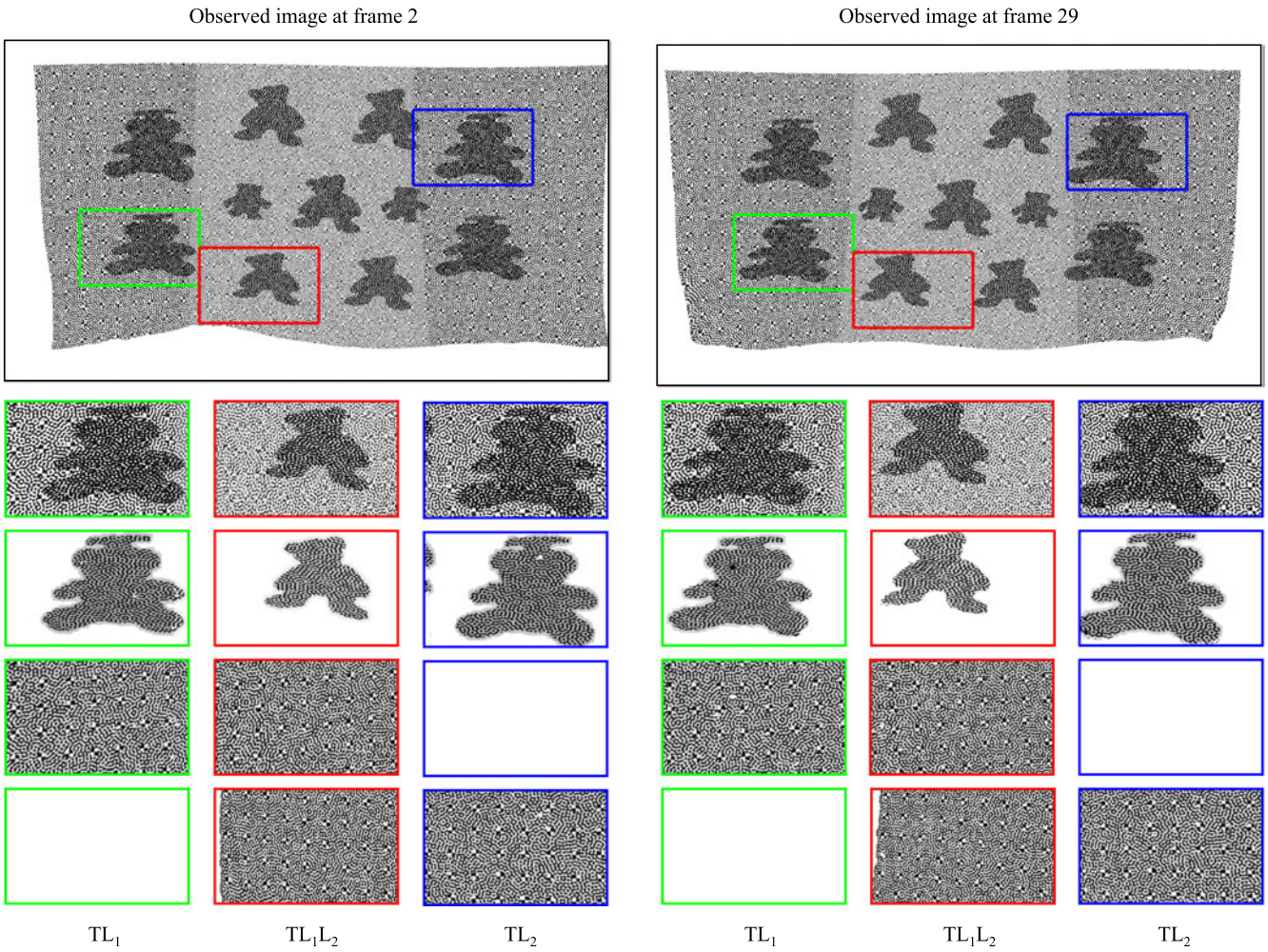


Fig. 6. Decomposition of the texture and illumination images from a synthetic bear cloth sequence with interleaving pattern projected every 30 frames. We show insets of three types of mixture: texture-illumination mixing, illumination interference, and texture and illumination interference. The same regions of the observed images are magnified to show how its intricate appearance changes over time. From the first row to the fourth row are the magnified regions of the observed image, the estimated texture, the estimated illumination for projector 1, and the estimated illumination for projector 2. Our method achieves better decomposition for higher frequency texture as white spots showing missing data are less visible.

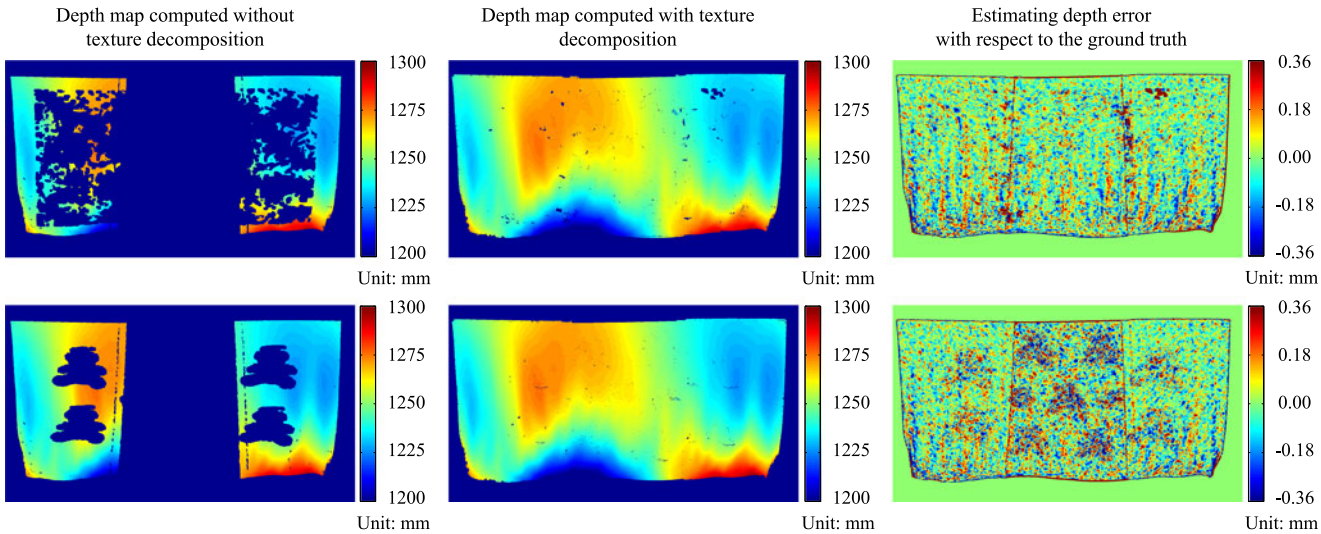
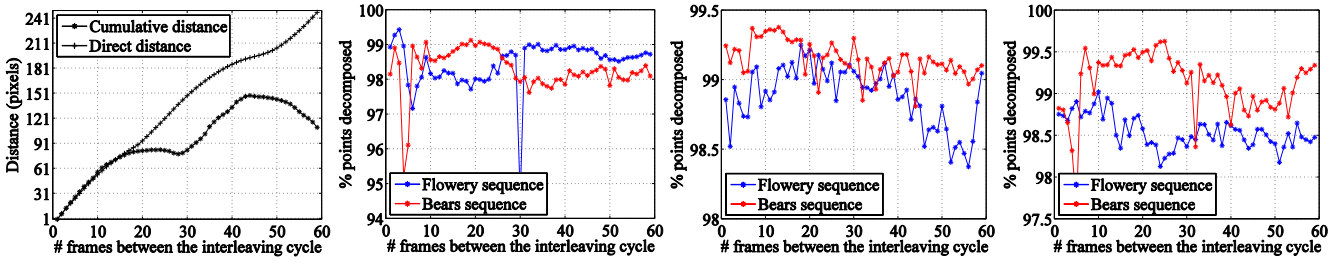
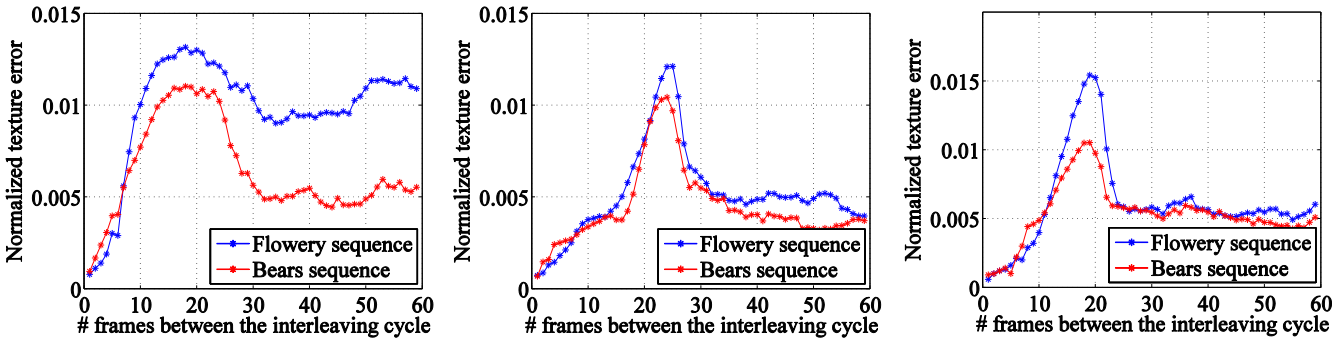


Fig. 7. Estimated depthmap of synthetic bear cloth sequence with interleaving pattern projected every 30 frames at the 29th frame . The depth map estimated with our decomposition scheme not only shows its completeness but its high accuracy with respect to ground truth depth.



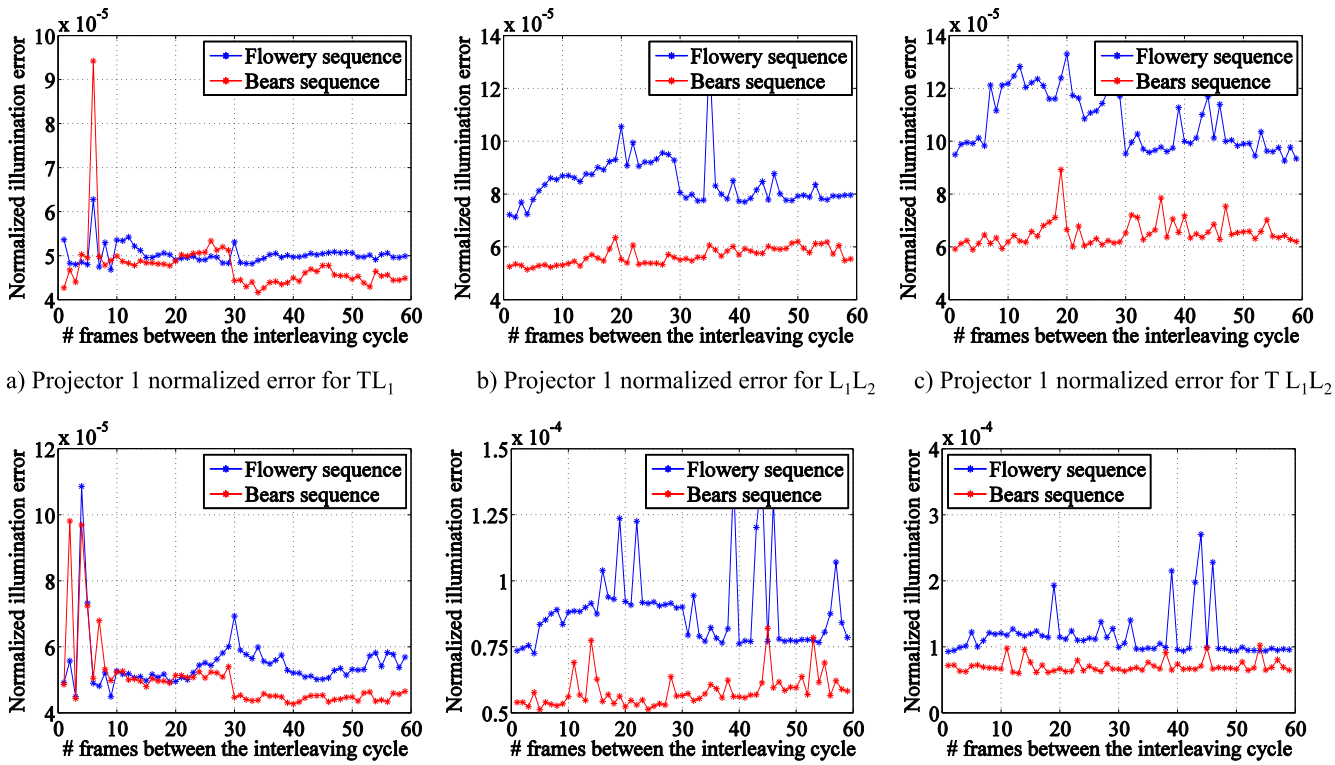
a) Warp distance b) Fraction of decomposed T L c) Fraction of decomposed L_1L_2 d) Fraction of decomposed T L_1L_2

Fig. 8. Warping distance and the percentage of points successfully decomposed with respect to the interleaving period. The larger the distance means difficult decomposition. The decomposed points are categorized into three types: texture and illumination, two illumination patterns, and texture and two illumination patterns. The number of points successfully decomposed is stable over time. Because of its higher frequency texture, the bear cloth has higher percentage of points decomposed than the flower cloth.



a) Normalized error for TL_1 b) Normalized error for TL_2 c) Normalized error for T L_1L_2

Fig. 9. The accuracy and robustness of the texture decomposition with respect to the interleaving period. The normalized texture errors are classified into different type of mixtures: texture-illumination pattern 1, texture-illumination pattern 2, and texture-illumination-illumination. Because of its higher frequency texture, decomposing the bear cloth sequence yields better accuracy than the flower cloth.



a) Projector 1 normalized error for TL_1 b) Projector 1 normalized error for L_1L_2 c) Projector 1 normalized error for T L_1L_2

d) Projector 2 normalized error for TL_2 e) Projector 2 normalized error for L_1L_2 f) Projector 2 normalized error for T L_1L_2

Fig. 10. The accuracy and robustness of the illumination decomposition with respect to the interleaving period. The normalized illumination errors for individual projectors are classified into different type of mixtures: texture-illumination,illumination-illumination, and texture-illumination-illumination. Because of its higher frequency texture, decomposing the bear cloth sequence yields better accuracy than the flower cloth.

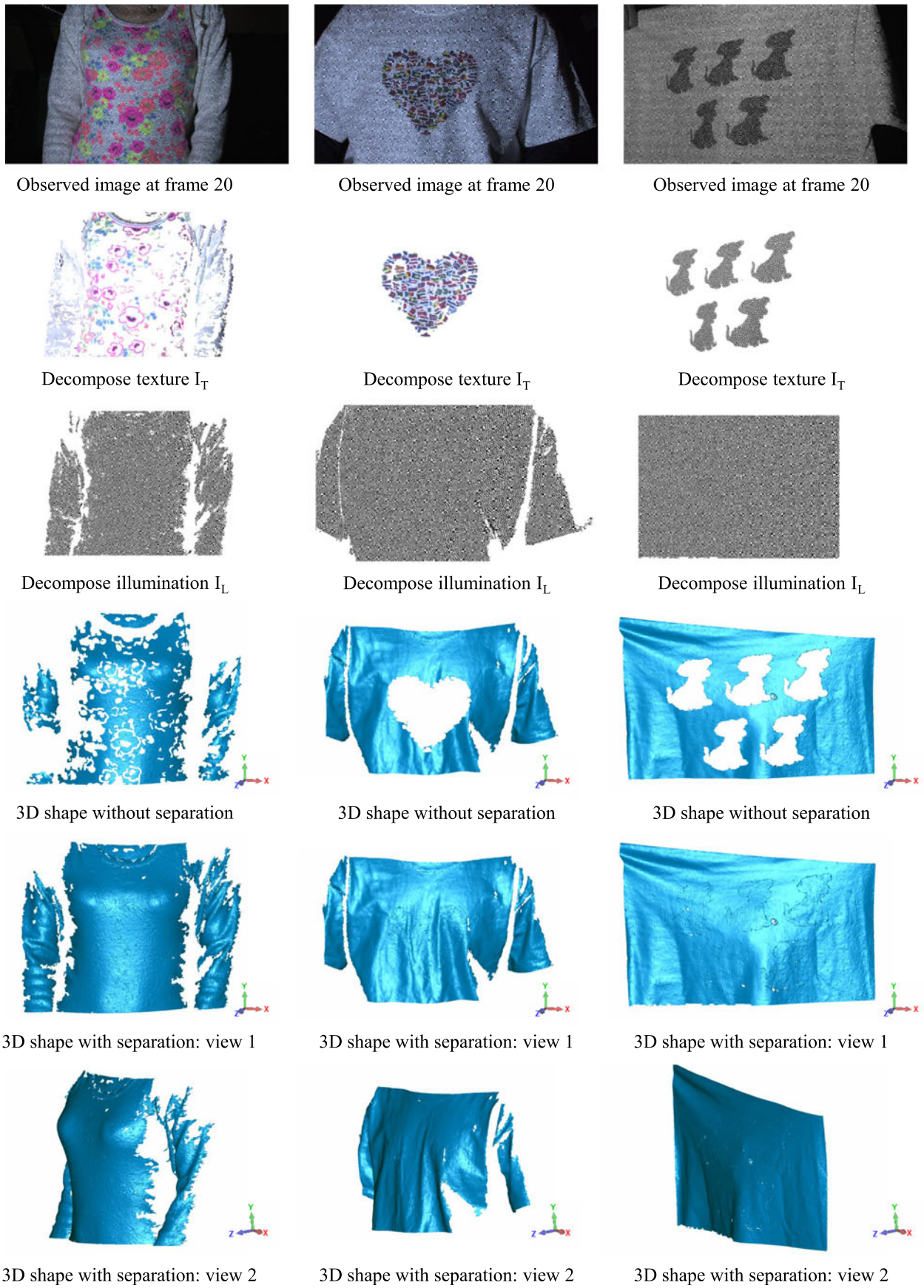


Fig. 11. Texture-Illumination decomposition of the real cloth sequences with 30 frames interleaving. Due to the relatively low texture frequency of the flower shirt, its texture image cannot be complete recovered. Yet, all the sequences, the illumination images are well computed. Compared to the 3D shape without the separation scheme, detailed surface information is recovered after texture and illumination separation. Please refer to our project website for the videos of the results.

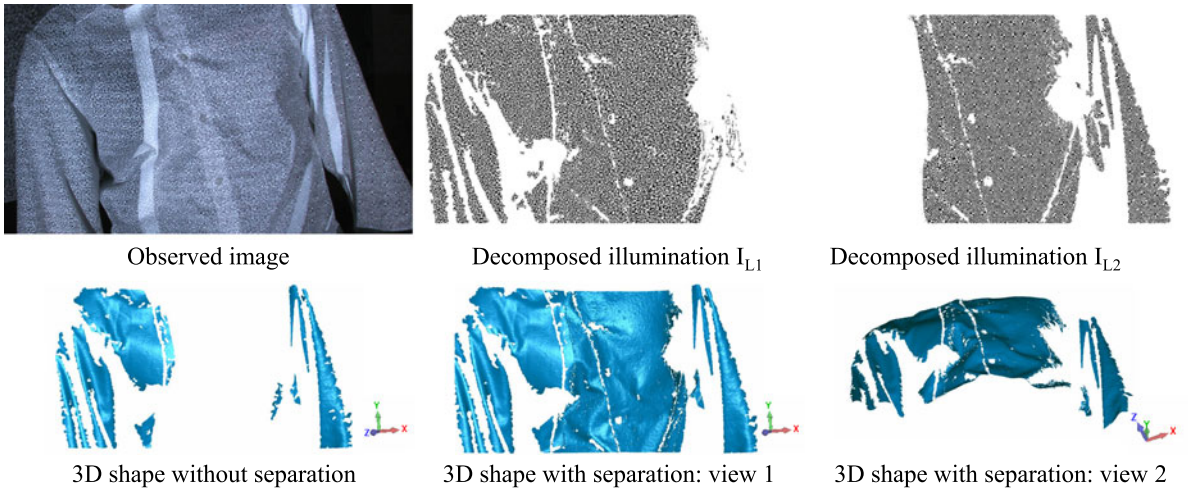


Fig. 12. Illumination decomposition and 3D shape of the real cloth sequences. Estimating the 3D shape of the cloth in the mixed illumination regions is not possible without the decomposition scheme. After the two illumination patterns are separated, the details of the folding cloth are clearly revealed. The results are without any post-processing. Please refer to our project website for the videos of the results.

those estimated on the synthesized pure illumination and texture images. The pure illumination image is created by projecting the illumination pattern onto a textureless cloth for each individual projector while keeping the others off. Texture image is obtained by texturemapping the cloth with binarized Perlin-noise pattern generated by error-diffusion dithering and projecting white light on the cloth. This binarized random pattern is known to give accurate tracking results [3], [26]. We compute the spatial correspondences between camera and projector on the pure illumination image and the temporal correspondences on the pure texture image using patch-based matching methods [4]. These correspondences are the best possible estimation at a given frame and hence, serve as ground truth.

We define our error metric as the normalized error in decomposing texture and illumination images:

$$\frac{1}{N}\sqrt{\left(\frac{\widehat{x_i}-x_i}{W}\right)^2+\left(\frac{\widehat{y_i}-y_i}{H}\right)^2}, \tag{8}$$

where N is the number of points inside the mixed appearance regions, $(\widehat{x_i},\widehat{y_i})$, and (x_i,y_i) are the ground truth and the estimated correspondence locations, respectively. Depending on whether the error of the illumination or texture is being evaluated, (W,H) could be either the resolution of the projector or camera image.

We compute two metrics for measuring the difference between the current image and the reference template: direct warp distance with respect to the template and the cumulative warp distance computed along the deformation path from the reference template to the current image. Since the motion of the cloth is quite repetitive, the second distance metric is needed to measure the difficulty in estimating the warping parameters.

Fig. 8 shows the warp distance and percentage of points successfully decomposed for different type of appearance mixtures for the bear and flower cloth sequences along with the distance of the current frame to its reference template as a function of the interleaving period. Despite long interleaving period, the fraction of points decomposed does not experience significant drops. The texture decomposition error and the fraction of correspondences estimated for flower sequence are not as good as for the bear sequence. This indicates that the algorithm performs better with higher frequency texture. The explanation for this phenomenon is similar to the optical flow problem: tracking highly textured surfaces exhibits less drift.

We show the texture decomposition accuracy of our method for the bear and flower cloth sequences on different type of mixed regions in Fig. 9. Since the error does not increase dramatically over time, our method does not require frequent interleaving. The maximum normalized error for texture and multiple illumination patterns mixing is slightly higher than the mix of the texture and only one illumination pattern. This is expected because of its high warping complexity. While the bear cloth achieves better accuracy than the flower cloth when only one projector is visible to it, the difference appears to be smaller when both projectors are visible. This is because the multiple illuminations create more image gradient in the observed images, only very accurate estimation of texture patch can re-synthesize the observed image.

Fig. 10 shows our accuracy and robustness in decomposing the illumination. Despite some sudden spikes, the effect of the interleaving period on the accuracy is minimal. Except for the mixing between projector 2 and the texture, estimating the warping parameters for the bear cloth is twice more accurate than the flower cloth. This behavior follows the trend of the previous graphs.

6 Results on Real Dataset

We conduct several experiments with real cloth sequences of different texture frequencies for three different scenarios: one projector and textured object, two projectors and textureless object, and two projectors and textured object. In the first case, we interleave the all-white pattern every 30 frames. In the second case, since the two illumination patterns have provided the warping templates, no interleaving is needed. In the third case, we project the

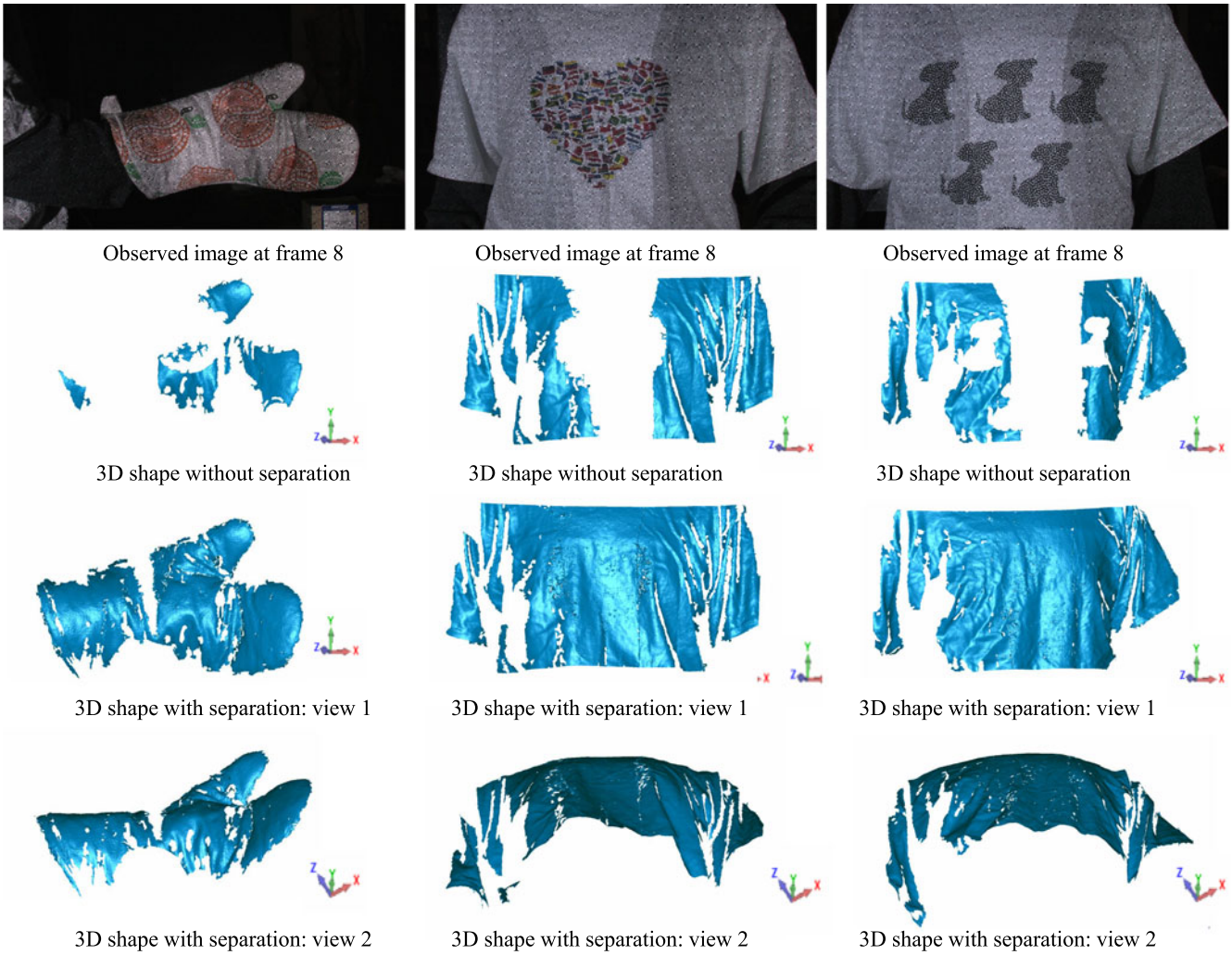


Fig. 13. Estimated 3D shape for the cloth sequences. The interleaving pattern is projected every 10 frames. No post-processing is applied. Without the decomposition, estimated 3D shapes contain large holes that hole-filling methods cannot yield reasonable results. Conversely, our decomposition scheme allows accurate 3D surface estimation with clearly visible 3D shape of the folded cloth. Please refer to our project website for the videos of the results.

interleaving pattern every 10 frames because of the complexity of the scene appearance. For all of our experiments, the scenes are illuminated using a 1280×800 DLP View Sonic projector and the images are acquired by the Canon XH-G1s HD 1920×1080 camera operating at 30fps. The camera and the projectors are calibrated using the method of Vo et al. [43]. The results are presented without any post-processing.

6.1 Texture and Illumination Mixing

We validate our algorithm for the texture and illumination separating task on three textured cloth sequences: flower, flag, and dog, arranged in increasing order of the texture frequency. Fig. 11 shows the decomposition results. While the decomposed texture image can be imcompleted for low texture frequency object, such as the flower shirt, such texture does not cause confusion in the illumination decoding as the estimated illumination image contains few empty regions. For high frequency texture object, such as the dog cloth, both the texture and illumination image are well decomposed. Interestingly, even for seemingly textureless regions, as in the sleeves of the flower shirt, the composite of the rough but textureless cloth and the illumination explains the observed image well, which does not happen with pure illumination alone. This enables us to recover the surface shape in such regions.

Since the flower shirt has relatively low texture frequency, without texture and illumination separation, its 3D shape is best-estimated among the three sequences. However, the abrupt changes in color at the boundary of the flowers still alter the appearance of the illumination pattern and the estimated 3D shape cannot be completed. As the texture frequency increases, the chance of reconstructing the surface shape in the mixed region decreases as shown for the flag and dog cloths. With our separation algorithm, the 3D shape is accurately recovered even in the mixed appearance regions for all three sequences. We obtain subtle 3D deformation, as vividly shown in the flag shirt sequence. The shape information in such large holes cannot be accurately recovered with any hole filling methods. These results, shown at frame 20 of the 30 frames interleaving period, indicates that the decomposition algorithm suffers from little drifting over time. Thus,

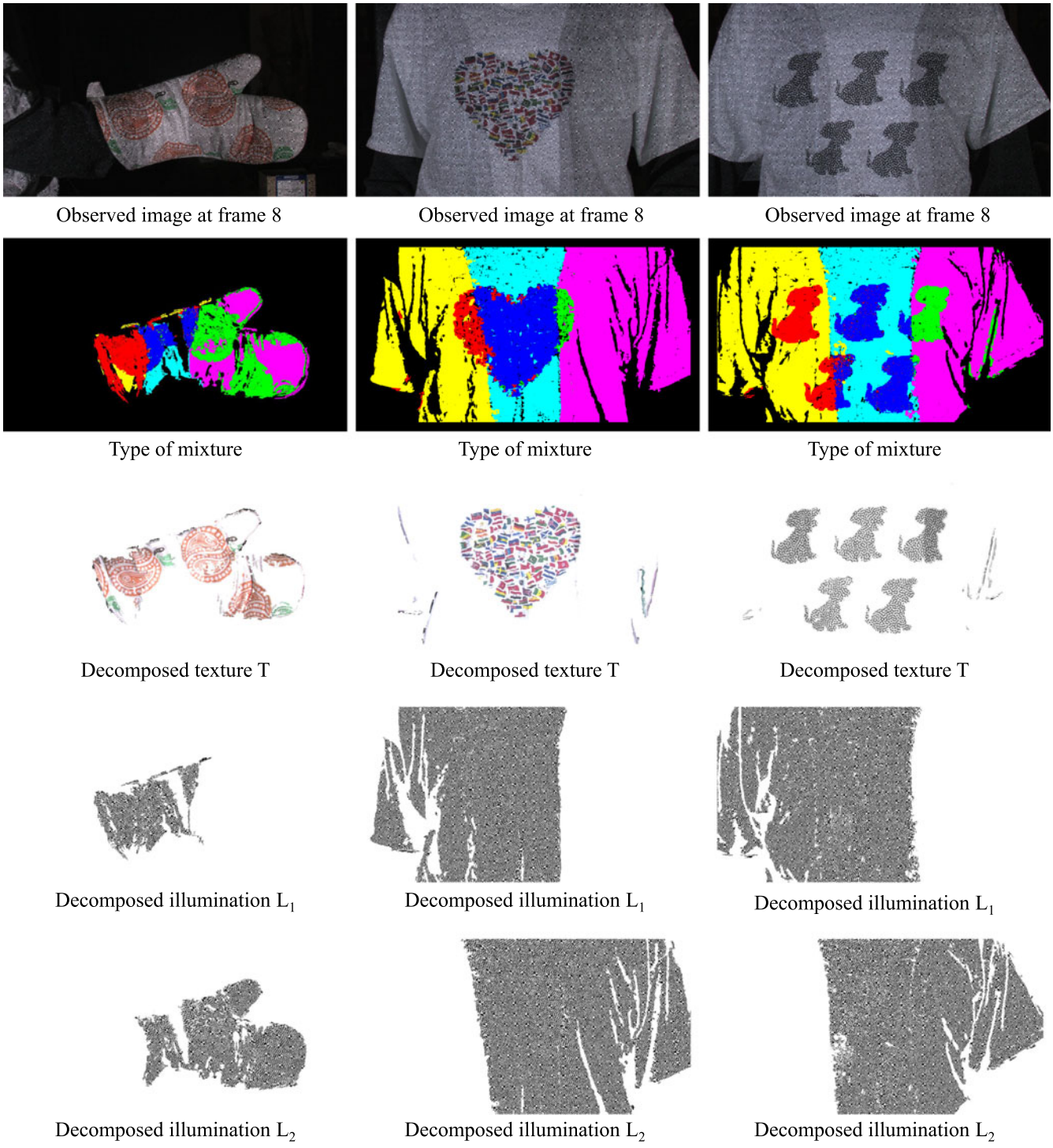


Fig. 14. Decomposition of the real cloth sequences. The interleaving pattern is projected every 10 frames. The scenes contain six different type of appearances: pure illumination from projector 1 (yellow), pure illumination from projector 2 (pink), projector 1 and texture (red), projector 2 and texture (green), projector 1 and projector 2 (light blue), and projector 1, projector 2, and texture (blue).

infrequently interleaving is possible so that high temporal resolution of the estimated 3D surface can be computed.

6.2 Two Illumination Mixing

We also validate the decomposition algorithm on the textureless shirt illuminated by two projectors, as shown in Fig. 12, where the superposition of the illumination patterns is a well-known problem. In this setup, while we experimented with two different high frequencies illumination pattern, the same patterns can be used as well, as shown in the next section. Since the illumination patterns have provided the reference templates for the warp function, no interleaving is needed. As can be seen, heavily distorted appearance of the illumination patterns in their overlapping regions makes shape recovery impossible. However, after our decomposition, dense and accurate shape information can be computed in a single-shot. The decomposition fails when occlusion or strong foreshortening occurs.

Fig. 15. Benefit of texture separation for tracking. Applying the tracking algorithm directly on the observed image results erroneous optical flow. Conversely, after texture separation, the same tracking algorithm can faithfully show how the cloth is moving.

6.3 Texture and Two Illumination Mixing

Fig. 14 shows our decomposition results for the glove, flag, and dog sequences. The dog sequence contains the highest texture frequency. As encoded in the mixture type image, the scenes contain six different type of appearances: pure illumination from projector 1 (yellow), pure illumination from projector 2 (pink), projector 1 and texture (red), projector 2 and texture (green), projector 1 and projector 2 (light blue), and projector 1, projector 2, and texture (blue). Such complication of the mixture types makes image decomposition challenging. Consequently, while there could be places where the mixture types are not visually correct, the majority of them happen in either the transition region of the illumination pattern (see Fig. 4), the small but homogeneous textured regions, or the occluding boundary. As can be seen in the decomposed images, despite the simple imaging model, our approach can handle complex appearances of real world textured objects illuminated by the two projectors. We believe this is due to the local block decomposition strategy which is robust to global lighting variation.

Fig. 15 shows the benefit of texture separation for surface tracking. Because of the mixed appearance, direct application of the patch based tracking algorithm [4] on the observed image does not yield plausible optical flow. The directions of the flows are noisy and they do not any dominant directions. Conversely, when the same tracking algorithm is executed on the decomposed texture image, the direction of the flow is less noisy and they faithfully present the true motion of the cloth. Such accurate tracking could be helpful for applications such as shape registration.

Fig. 13 shows the 3D shape of those cloth sequences. Without our separation scheme, the estimated 3D shapes are largely incomplete with obvious holes. In contrast, after our decomposition, the shape of the cloth is much better represented. While we could not obtain good reconstruction at the texture and illumination mixing regions where there are abrupt intensity changes at transition region the illumination patterns, the estimated shapes in the texture and two illuminations mixing are visually plausible. Our estimated shape of folded cloth, which cannot be obtained by hole-filling methods, appears naturally.

Fig. 16 shows the 3D shape obtained from the Kinect I sensor. For fair comparison with the performance of the proposed method shown in the flag sequence (see Fig. 14), the same subject stands at a similar distance to the sensor. Visually, the quality 3D shape from our method outperforms that of the Kinect. It is noteworthy that unless smoothing filter is applied to raw Kinect results, the mesh generation fails as the surface normal computed from raw point cloud is noisy.

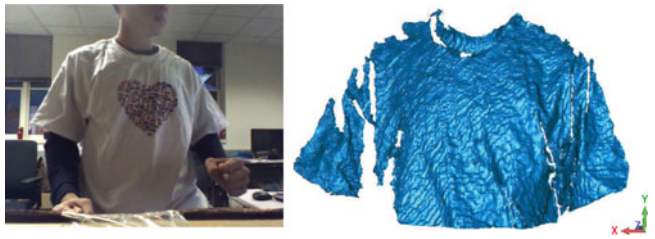


Fig. 16. 3D shape from the Kinect. Note that smoothing filters has been applied to generate the mesh from raw point cloud data.

7 Discussion

Since we use the texture warping parameters estimated in the previous frame as initialization, we can robustly warp an observed patch to its reference template that is temporally far away. Thus, only infrequent projection of the interleaving white frame is needed and the obtained results have high temporal resolution. This strategy is analogous to the approach of Tian and Narasimhan et al. [41] who use less distorted patches to estimate globally optimal sets of warping parameters. Clearly, since the interleaving sequence is dependent on the motion of the object, the interleaving period has to be adapted to different applications. Nevertheless, in the era 60-fps consumer-grade cameras, there is no need to interleave every other frame.

In spite of the greedy growing strategy, erroneous matches cannot propagate long as examined and thresholded by the cost function in Equation (6). Hence, our method avoids both the global ambiguity of the illumination pattern and being stuck in regions where occlusion, surface discontinuity, or severe foreshortening occurs. Moreover, since only a few seed points are needed initially, the good correspondences in the current frame can quickly propagate to non-decomposable regions in the earlier frames. Such automatic recovery from previous failures is another important property of our method.

For objects with high frequency texture everywhere, the proposed algorithm will fail. Yet, such cases are rare and most objects have a mixture of low and high frequency texture regions (see Fig. 11). While our separation method may also fail for objects with low frequency texture, the camera-projector correspondence can be established to reconstruct 3D shape for those cases.

While we only show the results for a single deforming object, our algorithm is applicable to general scenes containing multiple objects. As long as the seed points, i.e., correspondences established in low frequency textured regions, are available on each object, these correspondences can propagate to the entire object. Nevertheless, because of the nature of the patch decomposition approach, our algorithm cannot handle well the textured regions at the occluding boundary.

Theoretically, the decomposition model of Equation (1) can be generalized to more projectors where the type of mixture grows exponentially. However, the increase in the number of projector faces a diminishing return scenario when random illumination pattern from multiple projectors

are averaged out. In such cases, the contrast of the observed mixed regions become too low and our decomposition fails. Nevertheless, in practice, even for a surround structured light system, as in [24], there are at most two projectors to illuminate any point on the object surface.

Due to the sequential nature of our decomposition algorithm, it does not run in realtime. Unless knowledge about the surface geometry is known, reasonable initial guess of the optimization is not possible. Hence, the only feasible way to speed up the method is through more efficient warping method [4].

Acknowledgments

This research was supported in parts by an ONR Grant N00014-11-1-0295, the US National Science Foundation (NSF) Grant IIS-1317749, and an NSF Grant No. 1353120.

References

[1] S. Achar, S. Nuske, and S. G. Narasimhan, "Compensating for motion during direct-global separation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 1481–1488.

[2] S. Achar and S. G. Narasimhan, "Multi focus structured light for recovering scene shape and global illumination," in *Proc. 13th Eur. Conf. Comput. Vis.*, 2014, pp. 205–219.

[3] B. Atcheson, W. Heidrich, and I. Ihrke, "An evaluation of optical flow algorithms for background oriented Schlieren imaging," *Exp. Fluids*, vol. 46, no. 3, pp. 467–476, 2009.

[4] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework," *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, 2004.

[5] S. Beck, A. Kunert, A. Kulik, and B. Froehlich, "Immersive group-to-group telepresence," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 4, pp. 616–625, Apr. 2013.

[6] D. A. Butler, S. Izadi, O. Hilliges, D. Molyneaux, S. Hodges, and D. Kim, "Shake'N'Sense: Reducing interference for overlapping structured light depth cameras," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2012, pp. 1933–1936.

[7] D. Caspi, N. Kiryati, and J. Shamir, "Range imaging with adaptive color structured light," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 5, pp. 470–480, May 1998.

[8] J. Cech, J. Sanchez-Riera, and R. Horaud, "Scene flow estimation by growing correspondence seeds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 3129–3136.

[9] H. J. Chen, J. Zhang, D. J. Lv, and J. Fang, "3-D shape measurement by composite pattern projection and hybrid processing," *Opt. Express*, vol. 15, no. 19, pp. 12318–12330, 2007.

[10] T. Chen, H. Lensch, C. Fuchs, and H. P. Seidel, "Polarization and phase-shifting for 3D scanning of translucent objects," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2007, pp. 1–8.

[11] T. Chen, H. P. Seidel, and H. P. Lensch, "Modulated phase-shifting for 3D scanning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2008, pp. 1–8.

[12] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. New York, NY, USA: Wiley, 2002.

[13] V. Couture, N. Martin, and S. Roy, "Unstructured light scanning to overcome interreflections," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 1895–1902.

[14] H. Farid and E. H. Adelson, "Separating reflections from images by use of independent component analysis," *J. Opt. Soc. Amer. A,*, vol. 16, no. 9, pp. 2136–2145, 1999.

[15] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, "On the removal of shadows from images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 59–68, Jan. 2006.

[16] R. Furukawa, R. Sagawa, H. Kawasaki, K. Sakashita, Y. Yagi, and N. Asada, "One-shot entire shape acquisition method using multiple projectors and cameras," in *Proc. 4th Pacific-Rim Symp. Image Video Technol.*, 2010, pp. 107–114.

[17] K. Gai, Z. Shi, and C. Zhang, "Blind separation of superimposed moving images using image statistic," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 19–32, May 2011.

[18] J. Gu, T. Kobayashi, M. Gupta, and S. K. Nayar, "Multiplexed illumination for scene recovery in the presence of global illumination," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 692–698.

[19] M. Gupta, Y. Tian, S. G. Narasimhan, and L. Zhang, "A combined theory of defocused illumination and global light transport," *Int J. Comput. Vis.*, vol. 98, no. 2 , pp. 146–167, 2011.

[20] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan, "A practical approach to 3D scanning in the presence of interreflections, subsurface scattering and defocus," *Int J. Comput. Vis.*, vol. 102, no. 3, pp. 33–55, 2012

[21] M. Gupta and S. K. Nayar, "Micro phase shifting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 813–820.

[22] M. Irani, B. Rousso, and S. Peleg, "Computing occluding and transparent motions," *Int. J. Comput. Vis.*, vol. 12, no. 1, pp. 5–16, 1994.

[23] A. Jepson and M. Black, "Mixture models for optical flow computation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 1993, pp. 760–761.

[24] N. Kasuya, R. Sagawa, R. Furukawa, and H. Kawasaki, "One-shot entire shape scanning by utilizing multiple projector-camera constraints of grid patterns," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshop*, 2013, pp. 299–306.

[25] H. Kawasaki, R. Furukawa, R. Sagawa, and Y. Yagi, "Dynamic scene shape reconstruction using a single structured light pattern," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2008, pp. 1–8.

[26] S. Konig and S. Gumhold, "Image-based motion compensation for structured light scanning of dynamic surfaces," *Int J. Intell. Syst. Tech. Appl.*, vol. 5, no. 3, pp. 434–441, 2008.

[27] T. P. Koninckx, A. Griesser, and L. Van Gool, "Real-time range scanning of deformable surfaces by adaptively coded structured light," in *Proc. 4th Int. Conf. 3D Digital Imaging Model.*, 2003, pp. 293–300.

[28] T. P. Koninckx and L. Van Gool, "Real-time range acquisition by adaptive structured light," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 432–445, Mar. 2006

[29] D. Lanman, D. Crispell, and G. Taubin, "Surround structured lighting: 3-D scanning with orthographic illumination," *Comput. Vis. Image Understanding*, vol 113, no. 11, pp. 1107–1117, 2009.

[30] G. Ye, E. Garces, Y. Liu, Q. Dai, and D. Gutierrez, "Intrinsic video and applications," *ACM Trans. Graph.*, vol. 33, no. 4, p. 80, 2014.

[31] M. Movania. (2011). Opencloth [Online]. Available: https://code.google.com/p/opencloth/

[32] A. Maimone and H. Fuchs, "Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras," in *Proc. IEEE Int. Symp. ISMAR*, 2011, pp. 137–146.

[33] S. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar, "Fast separation of direct and global components of a scene using high frequency illumination," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 935–944, Jul. 2006.

[34] R. Sagawa, R. Furukawa, and H. Kawasaki, "Dense 3D reconstruction from high frame-rate video using a static grid pattern," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 9, pp. 1733–1747, Sep. 2014.

[35] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," *Pattern Recognit.*, vol. 43, no. 8, pp. 2666–2680, 2010.

[36] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2003, pp. 195–202.

[37] Y. Y. Schechner, N. Kiryati, and R. Basri, "Separation of transparent layers using focus," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1998, pp. 1061–1066.

[38] Y. Y. Schechner, J. Shamir, and N. Kiryati, "Blind recovery of transparent and semireflected scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2000, pp. 38–43.

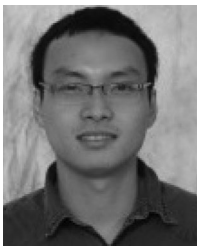
[39] Y. Y. Schechner, J. Shamir, and N. Kiryati, "Polarization and statistical analysis of scenes containing a semireflector," *J. Opt. Soc. Amer. A,*, vol. 17, no. 2, pp. 276–284, 2000.

[40] M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol 27, no. 9, pp. 1459–1472, Sep. 2005.

[41] Y. Tian and S. G. Narasimhan, "Globally optimal estimation of nonrigid image distortion," *Int. J. Comput. Vis.*, vol. 98, no. 3, pp. 279–302, 2012.

[42] A. O. Ulusoy, F. Calakli, and G. Taubin, "One-shot scanning using de Bruijn spaced grids," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2009, pp. 1786–1892.

[43] M. Vo, Z. Wang, B. Pan, and T. Pan, "Hyper-accurate flexible calibration technique for fringe-projection-based three-dimensional imaging," *Opt. Express*, vol 20, no. 15, pp. 16926–16941, 2012.
[44] M. Vo, S. G. Narasimhan, and Y. Sheikh, "Separating texture and illumination for single-shot structured light reconstruction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshop*, 2014, pp. 433–440.
[45] J. Wang and E. Adelson, "Representing moving images with layers," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 625–638, Sep. 1994.
[46] T. Weise, B. Leibe, and L. Van Gool, "Fast 3d scanning with automatic motion compensation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2007, pp. 1–8.
[47] Y. Weiss and E. Adelson, "A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 1996, pp. 321–326.
[48] S. Yamazaki, A. Nukada, and M. Mochimaru, "Hamming color code for dense and robust one-shot 3d scanning," in *Proc. Brit. Mach. Vis. Conf.*, 2011, pp. 96.1–96.9.
[49] Z. Yang, Z. Xiong, Y. Zhang, J. Wang, and F. Wu, "Depth acquisition from density modulated binary patterns," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 23–28.
[50] L. Yu and M. Brown, "Single image layer separation using relative smoothness," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2014, pp. 2752–2759.
[51] L. Zhang, B. Curless, and S. M. Seitz, "Rapid shape acquisition using color structured light and Multi-pass dynamic programming," in *Proc. 1st IEEE Int. Symp. 3D Data Process., Vis., Transmiss.*, 2002, pp. 24–36.



Minh Vo received the BS degree in electrical engineering from the Catholic University of America in 2012. He is currently working toward the PhD degree in robotics at Carnegie Mellon University. His research focuses on developing novel illumination and imaging systems and large scale dynamic scene reconstruction in the wild. He is a student member of the IEEE.

Srinivasa G. Narasimhan received the PhD degree from Columbia University in 2003. He is an associate professor in the Robotics Institute, Carnegie Mellon University. His group focuses on novel techniques for imaging, illumination and light transport to enable applications in vision, graphics, robotics, and medical imaging. His works have received several awards: Marr Prize Honorable Mention Award (2013), FORD URP Award (2013), Best Paper Runner up Prize (ACM I3D 2013), Best Paper Honorable Mention Award (IEEE ICCP 2012), Best Paper Award (IEEE PROCAMS 2009), the Okawa Research Grant (2009), the US National Science Foundation (NSF) CAREER Award (2007), Adobe Best Paper Award (IEEE Workshop on Physics Based Methods in Computer Vision, ICCV 2007) and IEEE Best Paper Honorable Mention Award (IEEE CVPR 2000). His research has been covered in popular press including *NY Times*, BBC, *PC* magazine and *IEEE Spectrum* and is highlighted by NSF and NAE. He is the coinventor of programmable headlights, Aqualux 3D display, assorted-pixels, motion-aware cameras and low-power outdoor-Kinect. He cochaired the International Symposium on Volumetric Scattering in Vision and Graphics in 2007, the IEEE Workshop on Projector-Camera Systems (PROCAMS) in 2010, and the IEEE International Conference on Computational Photography (ICCP) in 2011, coedited a special journal issue on computational photography, and serves on the editorial board of the *International Journal of Computer Vision* and as area chair of top computer vision conferences (CVPR, ICCV, ECCV).

Yaser Sheikh received the BS degree from the Ghulam Ishaq Institute of Engineering Science and Technology in 2001 and the PhD degree from the University of Central Florida advised by Prof. Mubarak Shah in 2006. He is an associate professor at the Robotics Institute, Carnegie Mellon University, with appointments in the Mechanical Engineering Department and the Quality of Life Technology Center. His research interests span computer vision, computer graphics, and robotics, and are focused on the machine perception of social behavior. He has won *Popular Science's* Best of Whats New Award, the Honda Initiation Award (2010), Best Paper Awards at WACV (2012), SAP (2012), SCA (2010), and ICCV THEMIS (2009), and the Hillman Fellowship for Excellence in Computer Science Research (2004). He has more than 50 publications in leading conferences in computer vision, computer graphics, and machine learning, and holds four patents on his research. He has served as a senior committee member at leading conferences in computer vision, computer graphics, and robotics including SIGGRAPH (2013, 2014), CVPR (2014, 2015), ICRA (2014), and ICCP (2011). His research is sponsored by various government research offices, including NSF and DARPA, and several industrial partners including the Intel Corporation, the Walt Disney Company, Nissan, Honda, Toyota, and the Samsung Group. His research has been covered by various media outlets including The *New York Times*, *Popular Science*, BBC, MSNBC, *New Scientist*, slashdot, and *WIRED*.