Reference Consistent Reconstruction of 3D Cloth Surface

Norimichi Ukita†[1] and Takeo Kanade‡

†Graduate School of Information Science, Nara Institute of Science and Technology
‡The Robotics Institute, Carnegie Mellon University

Abstract

We propose a multiview method for reconstructing a folded cloth surface on which regularly-textured color patches are printed. These patches provide not only easy pixel-correspondence between multiviews but also the following two new functions. 1) Error recovery: errors in 3D surface reconstruction (e.g. errors in occlusion boundaries and shaded regions) can be recovered based on the spatio-temporal consistency of the patches. 2) Single-view hole filling: patches that are visible only from a single view can be extrapolated from the reconstructed ones based on the regularity of the patches. Using these functions for improving 3D reconstruction also produces the patch configuration on the reconstructed surface, showing how the cloth is deformed from its reference shape. Experimental results demonstrate the above improvements and the accurate patch configurations produced by our method.

Key words:
3D reconstruction, Cloth surface, Surface deformation acquisition, Graphs-cut

1. Introduction

Modeling the motion of non-rigid clothing is one of important topics in Computer Vision and Graphics: for surface reconstruction[1], body estimation under clothing[2, 3], and physical cloth simulation[4, 5]. Several studies have proposed ways to obtain the cloth model/parameters from the surface points of a cloth (see [6, 7], for example). Sample data of the cloth motion are required also for data-driven approaches without the physical cloth model (e.g. free cloth motion[8] and cloth motion driven by human motion[9]). Therefore, cloth surface reconstruction is a fundamental technology for all of the above applications.

We developed a 3D reconstruction method with the following properties that are crucial for cloth modeling:

Correctness Reconstruction error should be small.

High spatial density Spatially dense points are necessary because a cloth is completely non-rigid and its shape changes significantly even within a small area.

High temporal density Quick motion should be captured with a high frame-rate.

Completeness The surface of a cloth should be reconstructed as completely as possible to capture the whole motion of a cloth.

Configuration To capture the instantaneous motion of a cloth as well as its temporal deformation, each point on the reconstructed surface must correspond to its respective point on the reference surface (i.e. flat cloth with no tension). We call this correspondence a *configuration*. The configuration includes the orientation of each patch as well as its location. The configuration also enables time-coherent texture mapping (i.e. mapping any texture onto a deforming 3D surface).

These properties are classified into shape reconstruction (top four) and configuration acquisition. In our *reference configuration consistent* reconstruction, the inseparable relationships between them are used to improve their accuracy and robustness.

2. Related Work

General 3D reconstruction algorithms can be used for cloth surface reconstruction (e.g. dense and accurate reconstruction[21], one for a textureless object[23]). Recently, bundle adjustment[25, 26] and Graph-cut[15, 27] have been widely used for optimal solutions. These algorithms can obtain 3D points from multiview images, although some *incorrect* points are included and the *complete* shape cannot be captured due to occlusion and image processing errors such as multiview point correspondence. *High spatio-temporal density* can be acquired up to the spatio-temporal resolution of the cameras by frame-independent image-based reconstruction.

With frame-independent reconstruction, however, the *configuration* cannot be acquired because no correspondence between reconstructed shapes over time is obtained. Although surface point tracking[28, 29] provides us with temporal point correspondence, point correspondence over occlusion is difficult to obtain. Furthermore, to obtain the physical properties of the cloth (e.g. tension and spring parameters) and to achieve texture mapping from pictures on an image plane, motions of regularly distributed points and their deformation from the reference shape (i.e. flat cloth), namely the *configuration*, should be obtained.

The shape model (i.e. template mesh) of a garment provides several advantages, such as robust surface reconstruction and patch correspondences between the template and the reconstructed surface, by shape fitting. For example, shape matching with easily-identifiable parts of the garment (e.g. collars and cuffs) allows us to obtain temporally-coherent patch correspondences in the sequence of temporal meshes[24]. However, shape matching is inappropriate for several applications/scenarios; for example, 1) shape matching is not robust if easily-identifiable parts are occluded and/or no such parts exist in a target cloth and 2) the deformation from the reference flat shape cannot be obtained directly.

A coded pattern of color patches densely printed on the cloth improves the *correctness* and *spatial density* of multiview stereo. While most previous works using such a pattern concentrated only on these improvements (e.g. unique pattern generation for robust matching[31]), the pattern is also useful for obtaining the *configuration* as achieved by the White-Crane-Forsyth method[10]. In [10], the configuration is obtained by making patch correspondences between the coded pattern and the observed images based on a local combination of color patches.

For higher *correctness* and *spatial density*, active reconstruction using a programmable light source(s) (e.g. video projector) is effective: space-time matching[32], shape from defocus[33], and photometric stereo[34]. *High temporal density* can be guaranteed up to that of the camera by using high-frequency DLP projection[35]. However, each of these methods has several disadvantages. For example, temporal density is decreased because multiple images are required for each static shape[33] and the material of a target is limited[34]). Furthermore, an essential problem in these methods is the difficulty in extracting the color patches on the cloth due to projector lights, and therefore these methods cannot obtain the patch *configuration*.

Hole filling of the reconstructed shape is also important for obtaining the *complete* shape of the cloth because its complex folds produce large occlusions. While it is difficult to fill them even with a sophisticated filling method[36], prior knowledge of the target cloth allows us to obtain good results (e.g. using sample surfaces reconstructed in other frames[10, 24]). For example, in [10], completely-reconstructed (hole-free) surfaces observed at different moments are used as sample data for synthesizing hole filling results by surface deformation[11]. On the other hand, our method employs the patch configuration also for hole filling, which requires no sample surface data.

3. 3D Cloth Surface Reconstruction using Color Patches

3.1. White-Crane-Forsyth Method

Many methods have been proposed for reconstructing a 3D surface and its motion. Among them, a method proposed by White, Crane, and Forsyth[10] is the state-of-the-art multiview method using color patches printed on a cloth for reliability and precision. In their method, motion of the cloth with regularly-textured patches is observed from multiviews. While the method requires the printed patches, it is useful to acquire accurate cloth surfaces and parameters (e.g. tension, spring) for Vision and Graphics applications as mentioned in Introduction.

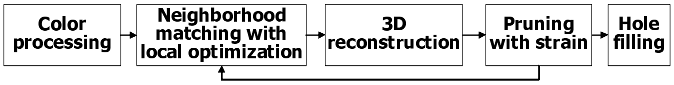


Figure 1: Five steps of the White-Crane-Forsyth method[10].

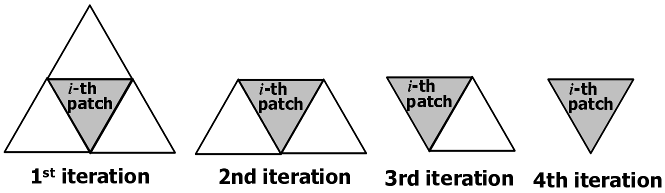


Figure 2: A combination of neighboring patches used for neighboring matching: a shaded patch depicts a patch of interest.

The method consists of the five steps below, whose flow is shown in Fig. 1:

Color processing: Sample values of each color are manually extracted from real images in advance. With the samples, each pixel color is determined based on its nearest neighbor in each observed image. Each patch is then segmented to find its neighbors (three neighbors for a triangle patch).

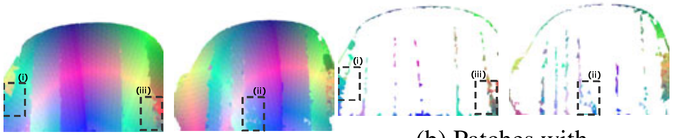
Neighborhood matching with local optimization: Patch correspondences over multiple views are established via those between each view and the cloth. A set of neighboring patches segmented in each image is compared with that on the cloth for matching. A patch of interest and its three neighbors are compared. ("1st iteration" in Fig. 2). Then the neighbors are decreased until 0. ("4th iteration" in Fig. 2). This matching works from flat regions, where many neighbors are observed in the image, to folded regions, where the combination of neighbors in the image does not match with that on the cloth. Patch correspondences between each image and the cloth gives us those between the multiview images.

More specifically, the correspondence between *i*-th image patch and *m*-th patch on the cloth is determined by locally optimizing a product over color similarities $c_{i,m}$ of neighbors: $c_{i,m} \prod_{l \in N_m} \max_{k \in n_i} c_{k,l}$, where n_i and N_m denote the image neighbors of *i*-th patch and the neighbors of *m*-th patch on the cloth.

3D reconstruction: The surface of the patches is reconstructed using triangulation of corresponding patches given by neighborhood matching.

Pruning with strain: Physically unrealistic strain should be avoided. Strain is computed from the distance between reconstructed 3D points; the original distance on the cloth is known. If any patches are pruned by this strain constraint, the process goes back to neighborhood matching with local optimization to remake correspondences of these patches again.

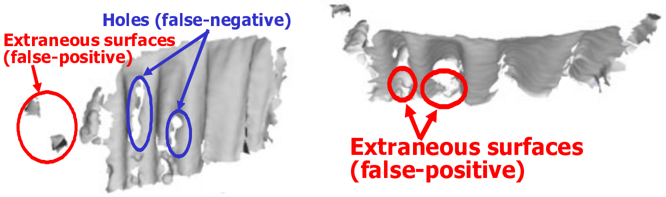
Figure 3: Input images. (Left) View1 and (Right) View2.



(a) Patch configuration images

(b) Patches with corresponding error

Figure 4: Results obtained by cloth reconstruction[10]. The zoom-in images of regions enclosed by broken lines are shown in Fig. 6.


Holes (false-negative)
Extraneous surfaces (false-positive)
Extraneous surfaces (false-positive)

(a) Left-frontal view

(b) Top view

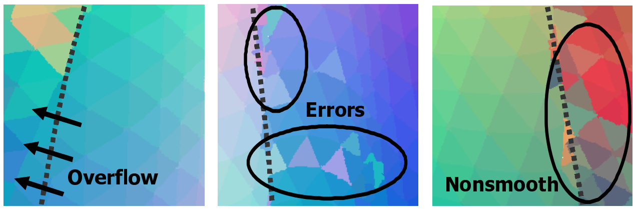
Figure 5: 3D surface reconstructed by the White-Crane-Forsyth method[10] using images captured at one frame; while the White-Crane-Forsyth method[10] can fill holes (false-negatives) by using the training data of the 3D surface as described in Sec. 2, this result was obtained by the images captured only at one frame. Red and blue circles show false-positives and false-negatives, respectively.

Hole filling: For filling a hole of patches in a frame of interest, occlusion-free surfaces of these patches in other frames are used. Surface deformation[11] interpolates the hole with the sample occlusion-free surfaces. Furthermore, temporal smoothing based on anisotropic diffusion[37] deforms time-varying surfaces while preserving fast non-rigid motion; conventional temporal smoothing is not appropriate because fast non-rigid motion might be blurred.

3.2. *Results by the White-Crane-Forsyth method and Their Problems*

We performed 3D cloth surface reconstruction using this method. A cloth with triangle patches, whose colors were red, green, blue, yellow, and magenta, was captured by cameras from two view points. Ther captured images are shown in Fig. 3, whose camera parameters were calibrated by the Zhang's method[12].

In each of the obtained patch configuration images shown in Fig 4 (a), gradually changing colors were given to the neighbors on the cloth. Neighbors with discontinuous colors indicate either occlusion boundaries or incorrect patch correspondences between the image and the cloth. Patches having incorrect correspondences (shown in Fig. 4 (b)) were extracted from Fig 4 (a). Most incorrect correspondences occurred at occlusion boundaries. These errors should be corrected for 3D reconstruction, while these results of neighborhood patch matching are better than those obtained by popular point matching; for comparison, the results of SIFT-based point matching[30] are shown in Fig. 7[2]. Indeed, these corresponding errors caused 3D reconstruction errors (i.e. large holes and extraneous elements) as shown in Fig. 5.



(i) (ii) (iii)

Figure 6: Detailed analysis of typical errors, which were extracted from (i), (ii), and (iii) in Fig. 4 (a). Each dotted line depicts an occlusion boundary.

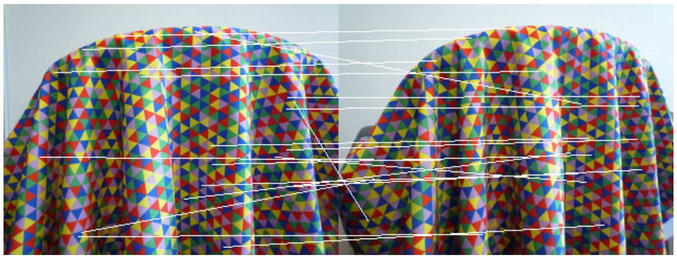


Figure 7: SIFT matching results. For visualization, a small number of the extracted points are shown. The success rate of matching was 9/17.

4. Detailed Analysis of the Problems and Their Solutions

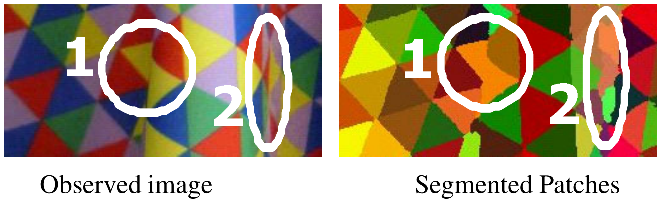
This section describes an analysis of what caused the problems with the White-Crane-Forsyth method[10], which gives us insights into solutions for resolving the problems. The detailed implementation of the solutions is described in the next section, Sec. 5.

4.1. *Implicit vs Explicit Occlusion and Ambiguity Handling*

Stepwise patch matching of the White-Crane-Forsyth method, shown in Fig. 2, allows us to separate flat regions with occlusion boundaries in heavily folded regions. Our method also employs this stepwise matching for obtaining initial patch correspondences. While this matching achieves implicit occlusion handling, overflows might occur as shown in (i) of Figs. 4 and 6 because the occlusions are not detected explicitly.

Indeed, complete detection of occlusion boundaries is difficult. In the examples shown in Fig. 8, patches with the same color were overlapped and regarded as one patch (circle "1" in the figure) and tiny patches were difficult to segment based

[2] In addition to SIFT matching, pixelwise color matching was also implemented. The results were not correct because SIFT features and other gradient-based features of the regularly-textured patches are very similar to each other.



Observed image

Segmented Patches

Figure 8: Patch segmentation errors in occlusion boundaries. Randomly selected colors were given to different patches in a right-hand image. 1) Patches with the same color are regard as one patch and 2) small patches are not extracted.

only on the color cue (circle "2"). These facts give us an insight into explicit occlusion handling; if a segmented patch is not represented clearly as a triangle, the region might be along an occlusion boundary. Removing such patches decouples three-dimensionally discontinuous surfaces in the image and allows us to impose a proper smoothness constraint on each surface.

As well as in occlusion boundaries, ambiguous textures (e.g. specular and shaded surfaces) violate multiview correspondences as shown in (ii) of Figs. 4 and 6. For example, the ambiguous textures might produce extraneous surfaces near real surfaces as shown in Fig. 4 (d). These reconstruction errors are difficult to remove based only on 3D proximity. We cope with this problem in a similar way to occlusion handling described above. Our method first makes correspondences between multiviews and the cloth without the ambiguous patches. Then the correspondences of the remaining ambiguous patches are constructed to fit the reconstructed patches.

4.2. Optimization with Local vs Global Matching

Neighborhood matching in this method[10] imposes a smoothness constraint only around a patch of interest. This local optimization might cause non-smooth correspondences (e.g. (iii) of Figs. 4 and 6)

Our method makes smooth correspondences by global optimization of multiview spatio-temporal cues. Assume that the i-th image patch (denoted by $\dot{p}_i$) corresponds to the m-th patch on the cloth (denoted by p_m). The patch configuration is optimized so that it satisfies the following consistencies:

C1. Uniqueness Each patch on the cloth matches at most one patch in each image. This consistency is identical to an occlusion term, which is used to avoid 1-to-many pixelwise correspondences between multiviews, proposed in [13].

C2. Color consistency $\dot{p}_i$ and p_m have the same color.

C3. Spatial consistency Neighbors of $\dot{p}_i$ correspond to neighbors of p_m.

C4. Multiview consistency $\dot{p}_i$ and its multiview correspondence in another view must have the same color. The multiview correspondence is given in the 3D reconstruction process.

C5. Temporal consistency Let $\dot{p}'_i$ denote the same point with $\dot{p}_i$ at the next capturing frame, which is obtained by a feature tracker[14]. $\dot{p}'_i$ must correspond to p_m.

An energy function that consists of all of the above consistencies is globally optimized by Graph-cut[15]. Though Graph-cut is widely used in multiview reconstruction where "pixelwise disparity" is optimized, "patch correspondences" between multiviews are optimized in our formulation. In both of the formulations, a smoothness term plays an important role; neighboring pixels should have the same disparity in multiview reconstruction and neighboring patches should be neighbors also in the other view in our formulation. In multiview reconstruction, however, the disparity gradually changes pixel by pixel in a curved surface. This causes difficulty in discrete optimization via Graph-cut. On the other hand, the optimization via Graph-cut fits into multiview patch correspondences, except patches around occlusion boundaries, which are removed in occlusion and ambiguity handling described above.

In our formulation, Graph-cut determines which patch on the cloth corresponds to which patch in each image. Although the original max flow algorithm used in Graph-cut can optimize only a binary energy function, α-expansion[16] achieves multi-label (i.e. patche IDs on the cloth in our formulation) optimization. It is also known that the max flow can obtain the global optimal solution only if the energy function satisfies submodularity[17]. For a non-submodular energy in our formulation, Quadratic Pseudo-Boolean Optimization (QPBO[18, 19]) provides a partially global optimal solution[3].

4.3. Cloth Strain Constraint vs Patch Configuration

In this method[10], a constraint with cloth strain prunes unrealistic patch correspondences in terms of a 3D distance between patches. This constraint cannot prune an extraneous surface if it is close to correct patches. Figure 5 (b) shows such errors.

Our global optimization also cannot make a correct correspondence in a patch with error in the preprocesses, namely color processing and 3D reconstruction. These two kinds of errors cause inconsistency in color consistency, C2, and multiview consistency, C4.

If other kinds of consistency are not satisfied in an image patch, $\dot{p}_i$, it is not easy to find a consistent patch ID because it is difficult to find which error(s) (e.g. patch segmentation, image-to-cloth patch correspondence, and 3D reconstruction) actually occurred in $\dot{p}_i$. For example, if $\dot{p}_i$ and another patch in the same image, $\dot{p}_j$, have a correspondence with p_m (i.e. C1 is not satisfied) and both of $\dot{p}_i$ and $\dot{p}_j$ satisfy all of other consistencies, we cannot determine which one has an error. What is worse is that making the new correspondence in $\dot{p}_i$ or $\dot{p}_i$ violates the results of global optimization; the change in the correspondence affects the correspondences in neighboring patches for satisfying spatial consistency, C3.

On the other hand, the error in the color/surface of $\dot{p}_i$ can be locally recovered if only C2/C4 is not satisfied in $\dot{p}_i$. This is because the change in the color/surface is not influential on the optimized correspondences in other patches.

[3] The max flow algorithm and QPBO were implemented in accordance with [20] and [19], respectively. Their source codes can be downloaded from the web site of the authors of these papers: http://www.cs.ucl.ac.uk/staff/V.Kolmogorov/software.html

4.4. *Shapes in Other Frames vs in a Frame of Interest*

The White-Crane-Forsyth method[10] employs sample surfaces and known patch correspondences between them in order to fill large holes in a reconstructed surface. While the hole filling method with sample surfaces[11] provides a reasonably interpolated surface, the sample surfaces must be reconstructed with no holes in other frames. This limitation rules out reconstruction from still images. An essential drawback of this hole filling is that patches corresponding to the hole, which are actually observed from one of the views in a frame of interest, are not used for hole filling. Our method improves accuracy of 3D reconstruction by integrating the surface reconstructed by multiviews and the single-view patches via the equidistant constraints between known correspondences on the cloth.

5. Our Cloth Reconstruction Method

5.1. *Detailed Implementation*

From the discussion in Sec. 4, our reconstruction method is designed as shown in Figs. 9 and 10. Compared with the White-Crane-Forsyth method, occlusion and ambiguity handling is added and neighborhood matching, pruning, and hole filling are augmented.

Color processing: Initial patches in each observed image are segmented in a similar way to the color processing in [10]. The results of pixelwise color detection are shown as "Detected color pixels" in the top-left box of Fig. 10.

Occlusion and ambiguity handling: As described in Sec. 4.1, evaluation of triangularity removes segmented image patches that are not similar to a triangle before initial reconstruction and neighborhood matching. "Segmented patches" in the top-left box of Fig. 10 shows the remaining segmented patches.

Specifically, for handling occlusion and ambiguity, the following two kinds of image patches are removed in our method:

Small patches If the number of pixels composing a patch is less than a threshold, this patch is removed.

Ambiguous neighbors In each segmented patch, pixels that are next to any other patches are counted. Three of them, each of which has one of the top three counts, are regarded as neighbors of this patch.

If a patch satisfies one or more of the following criteria, this patch is removed: 1) The number of pixels that are next to any of the neighbors is less than a threshold. 2) None of the neighbors includes this patch in its neighbors.

These simple criteria might over-remove patches. Our strategy is that these unreliable patches are initially neglected. Their correspondences are estimated 1) in other views where they are clearly observed or 2) by following the correspondences of other reliable patches.

3D reconstruction: Unlike the White-Crane-Forsyth method, the Epipolar constraint is used for an initial 3D surface ("Reconstructed 3D surface" in Fig. 10) at each frame. In our experiments, a stereo based method[21] reconstructed a dense 3D point cloud. Note that the point cloud is reconstructed simply by stereo reconstruction based on Epipolar geometry with no advantage of the known patch configuration on the cloth. The 3D surface patches of the point cloud are then computed by mesh reconstruction[22]. This surface reconstruction may generate artificial surface patches by connecting discontinuous points. By removing the surface patches that are far away from any of the reconstructed points, the initial reconstructed surface is obtained. Note also that multiview correspondences are not established in image patches removed by occlusion and ambiguity handling.

Neighborhood matching with global optimization: Initial patch correspondences among multiviews and the cloth ("Images with patch configuration" in Fig. 10) are established in the same way as the White-Crane-Forsyth method[10]. The optimization with Graph-cut is then achieved, whose details are described below:

Each patch on the cloth, p_m, has the following attributes: an ID m, a color cue $P^C(m)$, and the IDs of neighboring patches $P^N_l(m)$ aligned clockwise, where $l \in \{1, 2, 3\}$ denotes a neighbor. The same attributes are given also to each image patch in v-th view, $\dot{p}_{i,v}$: an image-patch ID i, a color cue $\dot{P}^C(i, v)$, and the image-patch IDs of neighbors $\dot{P}^N_k(i, v)$, where $k \in \{1, 2, 3\}$. In addition, $\dot{p}_{i,v}$ has the ID of its corresponding cloth patch, $\dot{P}^M(i, v)$.

For satisfying five consistencies C1−C5 mentioned in Sec. 4.2, an energy function E is minimized by Graph-cut:

$$
\begin{aligned}
E &= \textstyle\sum_v^{N^V} \sum_i^{N^I_v} (E^1(i,v) + E^2(i,v) + E^3(i,v) + E^4(i,v) + E^5(i,v)) && (1)\\
E^1(i,v) &= C^1 \textstyle\sum_{i' \in \bar{\boldsymbol{I}}} P^S(\dot{P}^M(i,v), \dot{P}^M(i',v)) && (2)\\
E^2(i,v) &= C^2 P^D(\dot{P}^C(i,v), P^C(\dot{P}^M(i,v))) && (3)\\
E^3(i,v) &= C^3 \textstyle\sum_k^3 \min_{l \in \{1,2,3\}} \left(P^D(\dot{P}^M(\dot{P}^N_k(i,v)), P^N_l(\dot{P}^M(i,v))) \right) && (4)\\
E^4(i,v) &= C^4 \textstyle\sum_{o \in \boldsymbol{O}} P^D(\dot{P}^M(i,v), \dot{P}^M(i_o,o)) && (5)\\
E^5(i,v) &= C^5 \textstyle\sum_{t' \in \boldsymbol{T}} P^D(\dot{P}^M(i,v), \dot{P}^M(i_{t'},v)) && (6)
\end{aligned}
$$

- C^1, C^2, C^3, C^4, and C^5 are weight constants.

- N^V and N^I_v denote the number of views and image patches in v-th view, respectively.

- $P^S(i, j)$ and $P^D(i, j)$ are the penalty functions that return 1 if $i = j$ and $i \neq j$, respectively. Otherwise 0.

- $\bar{\boldsymbol{I}}$ includes all patches except $\dot{p}_{i,v}$ in v-th view.

- $\boldsymbol{O}$ denotes a set of different views, each of which has a patch correspondence with $\dot{p}_{i,v}$. i_o is the ID of an image patch in o-th view that matches $\dot{p}_{i,v}$.

- $\boldsymbol{T}$ denotes a set of temporal variables, $t' \in \boldsymbol{T}$. An image captured at t' is used for evaluating C5 (temporal consistency). In our experiments, an image captured at t is compared with images captured at $t - 1$ and $t + 1$: $t' \in \boldsymbol{T} = \{t - 1, t + 1\}$. $i_{t'}$ is the ID of an image patch captured at t' that temporally matches $\dot{p}_{i,v}$.

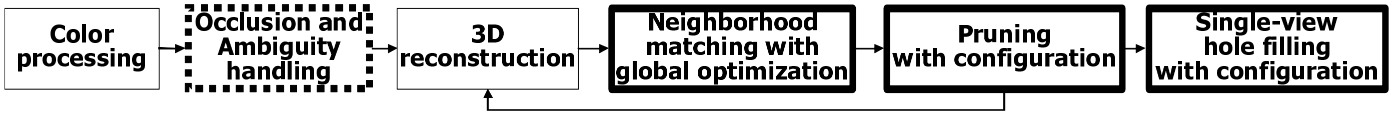

Color processing
Occlusion and Ambiguity handling
3D reconstruction
Neighborhood matching with global optimization
Pruning with configuration
Single-view hole filling with configuration

Figure 9: Six steps of our method. Boxes with dotted and thick lines indicate our new and augmented steps, respectively.

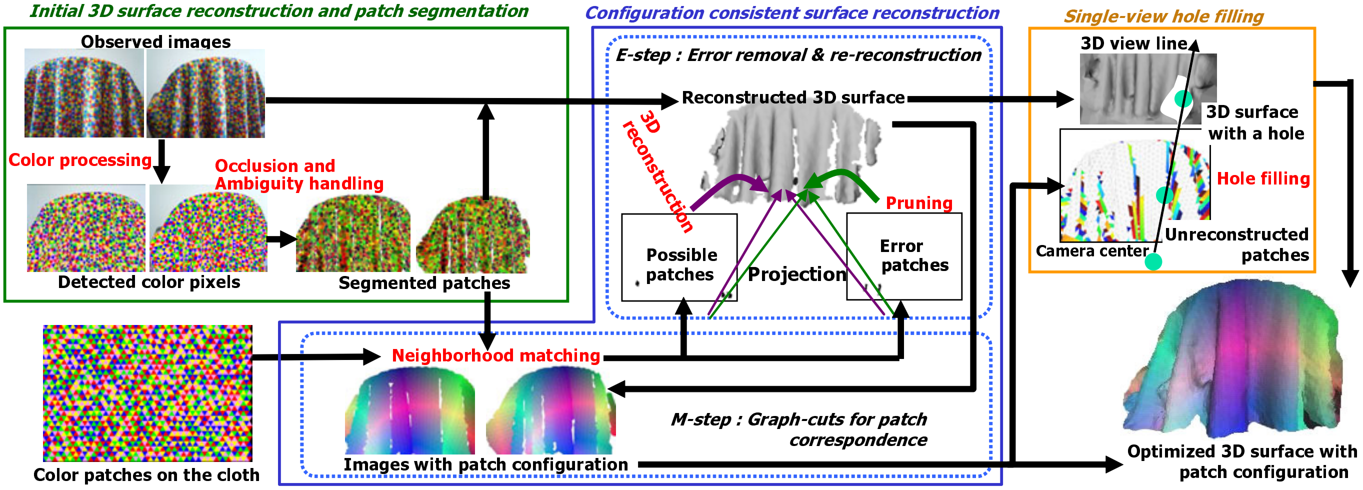

Initial 3D surface reconstruction and patch segmentation
Observed images
Color processing
Occlusion and Ambiguity handling
Detected color pixels
Segmented patches
Color patches on the cloth
Configuration consistent surface reconstruction
E-step : Error removal & re-reconstruction
Reconstructed 3D surface
3D reconstruction
Pruning
Possible patches
Projection
Error patches
Neighborhood matching
M-step : Graph-cuts for patch correspondence
Images with patch configuration
Single-view hole filling
3D view line
3D surface with a hole
Hole filling
Camera center
Unreconstructed patches
Optimized 3D surface with patch configuration

Figure 10: Overview of our method. Notes in red show the basic processes described in Sec. 5.1. Black arrows indicate process flows.

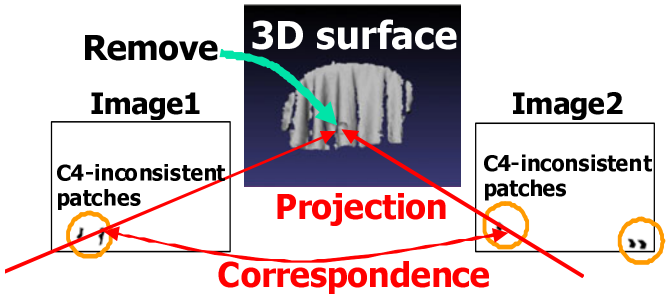
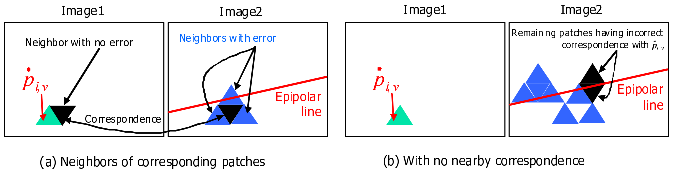


Figure 11: Pruning of patches with inconsistency in C4.



(a) Neighbors of corresponding patches

(b) With no nearby correspondence

Figure 12: Error recovery with the patch configuration. Blue patches depict possible correspondences. The epipolar line of $\dot{p}_{i,v}$ is depicted in image2. In image2 of (b), all patches with no correspondence are illustrated.

Pruning with configuration: Incorrect color and surface of $\dot{p}_{i,v}$, which has inconsistency in C2 and C4 respectively, can be found from the results of global patch optimization.

If only C2 is not satisfied, the second nearest neighbor color is selected in color processing.

If only C4 is not satisfied, the reconstructed surface of $\dot{p}_i$ is removed as illustrated in Fig. 11. After this pruning scheme, our method goes back to 3D reconstruction for recovering the surface of $\dot{p}_{i,v}$. The 3D error of $\dot{p}_{i,v}$ is recovered by one of the following two ways depending on whether or not any of the neighbors of $\dot{p}_{i,v}$ (denoted by $\dot{P}^N_k(i,v)$) is reconstructed with no error:

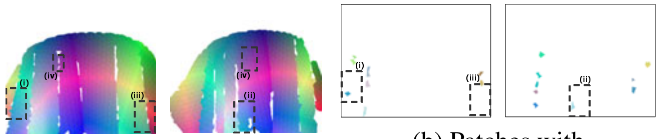
(a) Neighbors of correspondences If one or more of $\dot{P}^N_k(i,v)$ are reconstructed with no error, $\dot{p}_{i,v}$ might correspond to the neighbors of the corresponding patches of $\dot{P}^N_k(i,v)$. Then, $\dot{p}_{i,v}$ is compared only with these neighbors in other cameras (depicted by blue patches in image2 of Fig. 12 (a)) along the Epipolar line.

(b) With no nearby correspondence Otherwise, $\dot{p}_{i,v}$ remakes a correspondence so that it must not correspond to i) patches that incorrectly correspond to $\dot{p}_{i,v}$ in past matching processes and ii) patches with other correct correspondences. Possible patch correspondences of $\dot{p}_{i,v}$, are depicted by blue patches in image2 of Fig. 12 (b).

3D reconstruction of these patches are performed by [21] as in the initialization process, except that patch matching is evaluated only in the above possible patch correspondences.

Iteration of 3D reconstruction, neighborhood matching, and pruning is continued until no recoverable surface error is detected by way of the hard-EM algorithm. The hard-EM algorithm consists of error recovery with pruning and 3D reconstruction (E-step) and neighborhood matching (M-step) as follows.

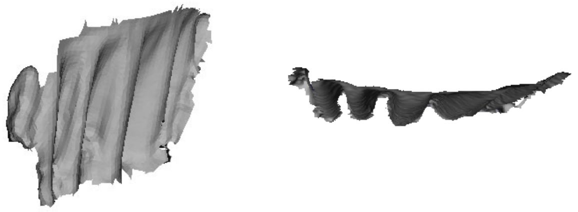
1. Estimate initial parameters, θ (i.e. cloth patch IDs of all image patches) in the same way as the White-Crane-Forsyth method[10]..
2. (E-step): Compute the best latent variables, $\mathbf{Z}$ (i.e. 3D points of the cloth surface), given θ. Note that patches



(a) Patch configuration images

(b) Patches with corresponding error

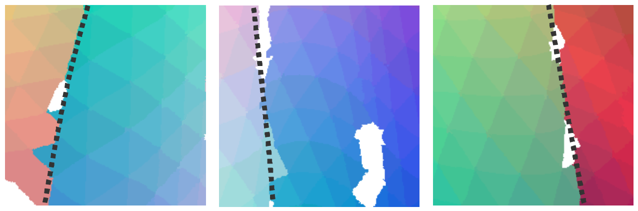
Figure 13: Results of our cloth reconstruction with configuration.



(a) Left-frontal view

(b) Top view

Figure 14: 3D surface reconstructed by our method (two views).



(i)

(ii)

(iii)

Figure 15: Detailed analysis of errors; compare them with Fig. 6, which show the errors in the same regions in Fig. 6.

with inconsistency only in C2 or C4 are used for 3D reconstruction after the first iteration.

3. (M-step): Update θ with the computed $\boldsymbol{Z}$ using Graph-cut.
4. Iterate E- and M-steps until no recoverable surface error is detected.

After the iteration is halted, patches in which one or more of the consistencies are not satisfied are removed from the reconstructed surface.

Hole filling: The surface corresponding to a patch that is observed only in a single view is estimated taking into account its view line and the smooth and equidistant relations with the neighboring reconstructed patches. The single-view patches are reconstructed from the fringe of the hole so that 1) the 3D position of the patch is located along its 3D view line, which is from the camera center to the patch on the image plane, and 2) the 3D distance from the hole patch to its neighboring reconstructed patch is approximately equal to that between two reconstructed patches. The detailed implementation of this hole filling is described below:

The mean of the 3D distances between all pairs of reconstructed neighboring patches (between their centers) is computed (denoted by $\bar{d}$). Given $\dot{p}_i^c$ that is observed only from c-th camera, let $L(i, v)$ denote the view line from the optical center of c-th camera to $\dot{p}_i^c$ on its 3D image plane. The 3D point of $\dot{p}_i^c$ is estimated so that it is on L and is $\bar{d}$ distant from its reconstructed neighbor (denoted by $\dot{p}_k^N(i, v)$):

- If $\bar{d} = d^{pl}$, where d^{pl} is a distance from L to the 3D point of $\dot{p}_k^N(i, v)$, the new 3D point corresponding to $\dot{p}_i^c$ is reconstructed in the foot perpendicular to L.
- If $\bar{d} < d^{pl}$, the new point is put in the foot perpendicular to L.
- If $\bar{d} > d^{pl}$, two points on L are $\bar{d}$ distant from the 3D point of $\dot{p}_k^N(i, v)$. In this case, the 3D point of $\dot{p}_i^c$ is interpolated by cubic spline interpolation. Then the one that is closer to the interpolated point is selected.

All the new 3D points are optimized so that the sum of $(\bar{d}-d^{pl})^2$ is minimized. 3D surface reconstruction[22] is then performed and the surface patches that are far away from any of the reconstructed points are removed, as in the initialization process. Finally, the optimized 3D surface and its patch configuration are reconstructed.

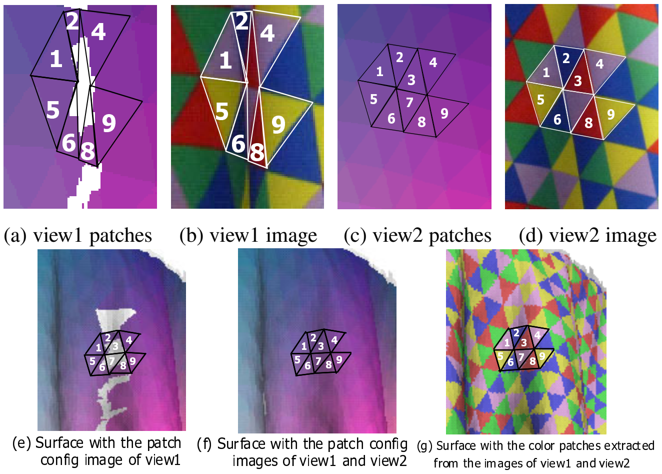
5.2. Improvements in Our Method

Experiments were conducted using our method with the same images used in Sec. 3.2. The following parameters were given to Graph-cut in all experiments in this paper: $C^1, C^2, C^3, C^4, C^5 = 3, 1, 1, 1, 1$. The temporal constraint C5 was not employed in the experiment with a pair of still images in this section.

Fig. 13 (a) and (b) shows the improved accuracy in patch configuration images and their error maps, respectively. The zoom-in images in Fig. 15 shows almost no overflow across the occlusion boundary. Fig. 14 shows the improvements in a reconstructed 3D surface; no big holes and no big false-positives. Compare our results with those produced by the White-Crane-Forsyth method[10], which are shown in Figs. 4, 5, and 6.

Figure 16 shows multiview compensation of the patch configuration around occlusion boundaries. While several patches (i.e. patches "3" and "7" in Fig. 16) could not be detected in view1 (i.e. (a) and (b) in Fig. 16), all patches were detected correctly in view2 (i.e. (c) and (d) in Fig. 16). These patches were then correctly and consistently located on the 3D cloth surface reconstructed by our single-view hole filling as shown in (f). For comparison, the 3D surface only with the patch configuration image of view1 is shown in (e), where the 3D surface was not colored by the IDs of the map patches. For visually verifying the correctness of the 3D surface and the obtained patch configuration, the color pixels extracted from view1 and view2 were reprojected onto the 3D surface; patches visible from view1 was textured by the color image of view1 and then the remaining patches were textured by the color image of view2. The result is shown in (g). It can be seen that the 3D surface was textured smoothly. This proves that the reconstructed surface is consistent with the obtained patch configuration.

As shown above, many of the improvements in our method are observed around occlusion boundaries. Across the occlusion boundaries, 3D surface patches reconstructed by stereo



(a) view1 patches (b) view1 image (c) view2 patches (d) view2 image

(e) Surface with the patch config image of view1 (f) Surface with the patch config images of view1 and view2 (g) Surface with the color patches extracted from the images of view1 and view2

Figure 16: Multiview compensation. (a) and (c) were extracted from (iv) in Fig. 13 (a), both of which show almost the same surface. (e) and (f) shows the 3D surfaces with the patch configuration reprojected from "view1" and "view1 and view2", respectively. (g) shows the 3D surface with the color images extracted from view1 and view2. Gray patches indicate the ones that were not observed from view1. A patch with the same digit indicates the correspondence.

should have a border with the patches reconstructed by single-view hole filling or the boundary line of the reconstructed surface. Hence, for further evaluation of improvements in our method, the images of the patch configuration with the boundary line were projected onto the reconstructed cloth surface. Figures 17 and 18 show the projection results of the White-Crane-Forsyth method and our method, respectively. In each figure, an input image, its patch configuration image, and 3D patches with the patch configuration are shown. First of all, occlusion boundary lines, B1 and B2, in the input image were manually drawn with black lines on the image of the patch configuration. Then the image of the patch configuration with the boundary lines was projected onto the 3D patches. While two patch configuration images (i.e. view1 and view2) were obtained in the experiments, only view1 was projected. 3D patches that were occluded from view1 were indicated by gray patches.

Comparison between Figs. 17 and 18 reveals the following:

- In view ii, 3D patches neighboring the boundary line, B2, (i.e. the right-hand side of B2) were not reconstructed by the White-Crane-Forsyth method because these patches were not reconstructed by stereo. Our method, on the other hand, reconstructed these patches, indicated by gray patches, by single-view hole filling.

- In view ii, it can be also seen that our method reconstructed 3D patches across B2 with high accuracy; B2 indicated by a black line is located just between patches colored by patch IDs (i.e. patches visible from view1) and gray patches (i.e. patches occluded from view1).

- While our method could reconstruct the patches around B2, those around B1, shown in view i, were not so accurate. This fact can be seen because patches occluded from view1 in the real shape were wider than those in the reconstructed result, which are indicated by gray patches in the figure; that is, the wrinkle in the real shape was deeper than the one in the reconstructed surface. This error might be caused because patches reconstructed by stereo were close to each other across the occluded patches so that the reconstructed patches were connected by simple interpolation by surface reconstruction[22] rather than single-view hole filling.

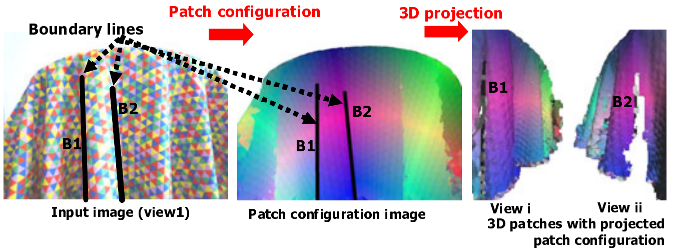


Figure 17: Reconstructed surface across occlusion boundaries: White-Crane-Forsyth method[10]. Gray patches indicate the ones that were not observed from view1.

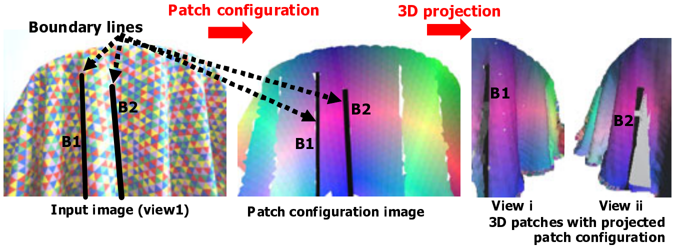


Figure 18: Reconstructed surface across occlusion boundaries: Our method.

Table 1: Computational time [sec] of each step in the proposed method. "3D reconstruction" and "neighborhood matching" are divided into "(point cloud reconstruction) + (surface reconstruction)" and "(initialization) + (graph-cut)", respectively.

Color processing	Occlusion and ambiguity handling	3D reconstruction	Neighbor matching by Graph-cut	Patch pruning	Hole filling
2	8	15 + 1	14 + 4	2	3

Finally, the computational time of each step was evaluated. As is well known, 3D reconstruction and Graph-cut have huge computational demands. Table 1 shows the computational time of each step in our method when it was applied to a stereo pair of 1024 × 960 pixels images shown in Fig. 3. It can be seen that the dominant time is spent by point cloud reconstruction and initialization of a patch configuration. Optimization of the patch configuration by Graph-cut needed a relatively small amount of time because optimization mainly affected patches only around occlusion boundaries.

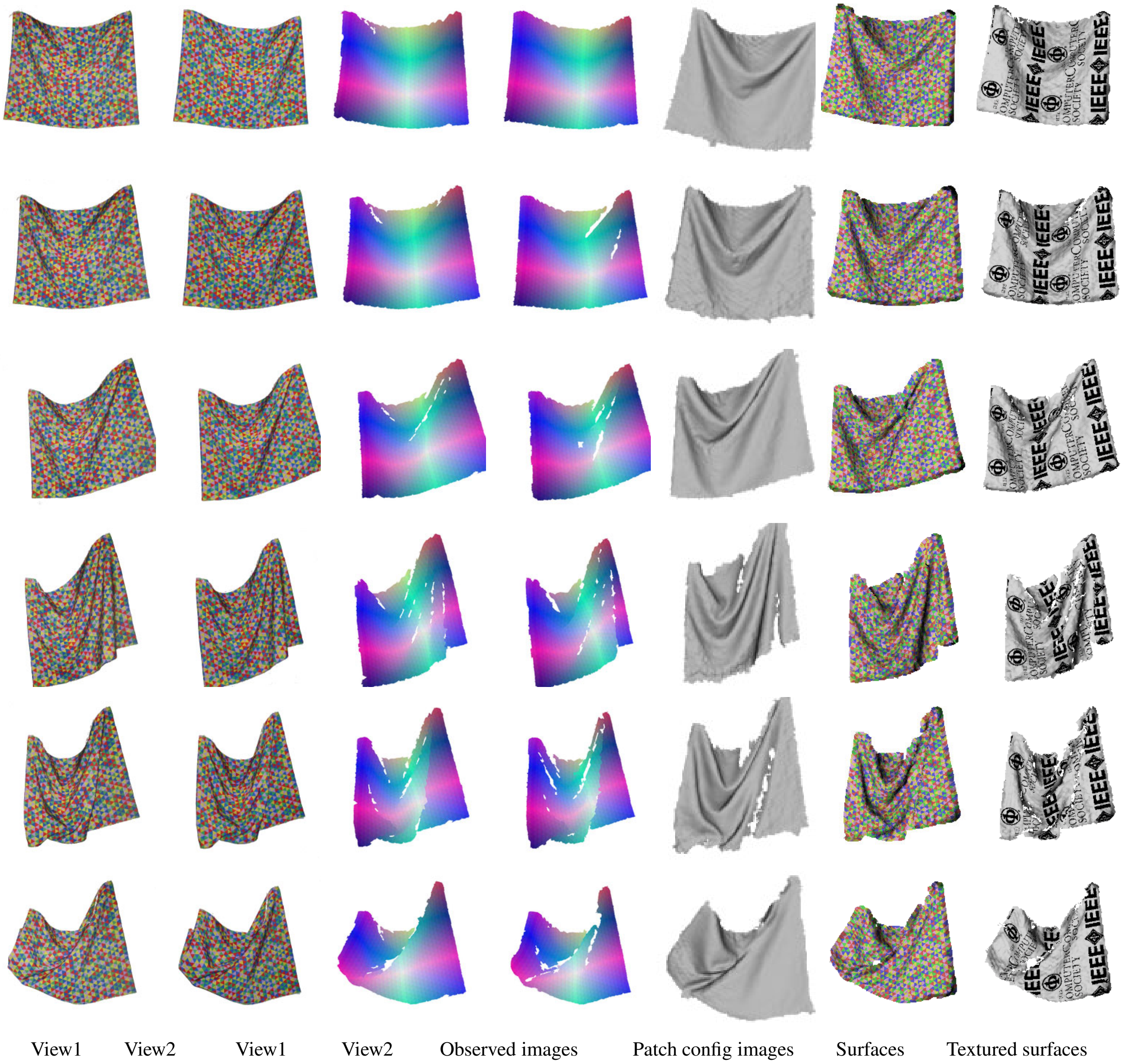


Figure 19: Results of our 3D cloth surface reconstruction with patch configuration: a swinging cloth. Each row shows images and results at the same moment.

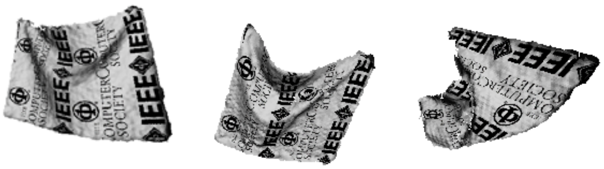
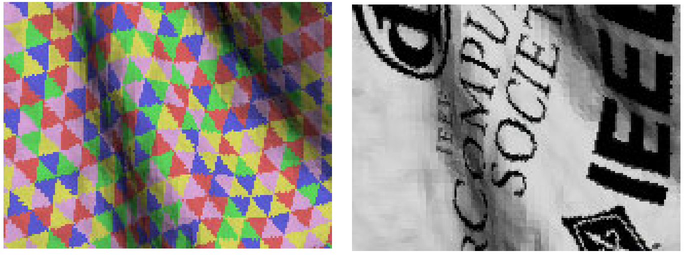


Figure 20: Different views of the reconstructed 3D surfaces with texture mapping.



(a) Texture from observed images

(b) Texture from the new texture

Figure 21: Zoom-in images of the textured surfaces.

6. Reconstruction from Sequences of a Moving Cloth

A moving cloth, which was used in the experiments described before, was captured in image sequences by a pair of synchronized cameras.

Figure 19 shows the results obtained from the two-view image sequences. Images in the sixth and seventh columns in Fig. 19 were generated by projecting texture images, (a) the detected colors in observed images and (b) a new texture, onto the reconstructed 3D surface. The textures were mapped from the triangles from the 2D texture image to those on the 3D surface by the Affine transform. Remember that our method removes surface patches that are far away from any of the reconstructed points. In the experiments, a small threshold for removing those 3D patches was given for verifying the results of point reconstruction; those 3D patches were removed while surface patch reconstruction[22] might fill the small holes of the reconstructed points.

Figure 20 shows the textured 3D surfaces that are observed from different views. This figure shows that the complex shape and patch configuration of the cloth were reconstructed simultaneously.

Figure 21 shows the zoom-in images of the reconstructed 3D surface with texture mapping. While jaggy effects were observed because no color blending was achieved, we can see the smooth textures on the 3D reconstructed surface.

The same experiments with another cloth, whose pattern was different from the one used in the experiments described above, were also performed. A bending arm with a sleeve mede of the printed cloth was captured. Figure 22 shows the results. Compared with the cloth surfaces in Fig. 19, a small number of patches, whose combination is prone to coincide with that on another area, on small surface areas were captured. It can be seen that the proposed method could capture the deformation of the 3D cloth surface. Figure 23 shows the detailed results (i.e. magnified images). From the results of texture mapping shown in Fig. 23, the following observations are confirmed:

Texture=Patch config image: The configuration along an occlusion boundary enclosed by a red rectangle could be estimated correctly.

Texture=Colors detected from observed images: The textured surface looks similar to the observed image; if the reconstructed surface is different from the real shape, the texture on the surface differs from the observed image.

Texture=New texture: The surface is textured smoothly.

For comparison, the magnified images of the results obtained by the White-Crane-Forsyth method[10], which can also obtain the 3D surface and the patch configuration on it, are shown in Fig. 24. Note again that these results were obtained with no hole filling using the sample surfaces of the target cloth, which should be collected in advance. Since the cloth had very few large wrinkles, the 3D surfaces could be reconstructed with few errors. However, the patch configuration in Fig. 24 includes typical errors of the White-Crane-Forsyth method (e.g. incorrect patch configuration around occlusion boundaries).

Experimental results in Fig. 25 demonstrate the effectiveness of our method for occlusion. The occlusion observed in these experiments was much significant than those in other experiments shown above. The magnified images of the results are shown in Fig. 26. An obstacle located in front of a target cloth made it impossible to capture several patches from two views simultaneously and track them. That makes it difficult to obtain the 3D surface by stereo and the temporally-consistent patch configuration by patch tracking. The target cloth was draped on the arm and was moved like the sleeve of Japanese kimono.

It can be seen that the configuration of the occluded patches could be obtained. The smoothly textured 3D surfaces ("Textured surfaces" in Figs. 25 and 26) prove the correctness of the obtained patch configuration. It can be also seen that the 3D surface of the occluded patches could be obtained by single-view hole filling. The regions reconstructed by single-view hole filling were enclosed by red and blue rectangles. Fig. 26 shows the magnified images of these regions.

For comparison, the results obtained by the White-Crane-Forsyth method[10] are shown in Fig. 27. In addition to false-negative surfaces due to occlusion, where the cloth was observed only from one view, typical errors of the White-Crane-Forsyth method are observed in Fig. 27, whereas our method could obtain the smooth patch configuration on the 3D surface.

7. Concluding Remarks

We propose a method for reconstructing the 3D surface of a folded cloth by cameras. Regularly-textured color patches printed on the cloth surface are employed to 1) provide explicit occlusion and ambiguity handling in a single view and 2) acquire the patch configuration on the reconstructed surface. The patch configuration is acquired by Graph-cut so that each patch

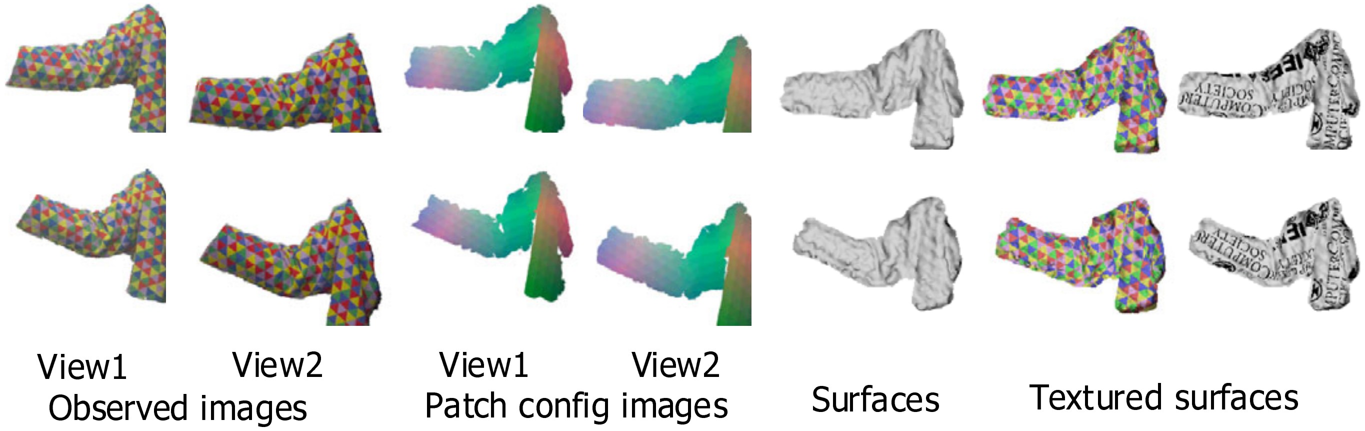


Figure 22: Results of our 3D cloth surface reconstruction with patch configuration: sleeve.

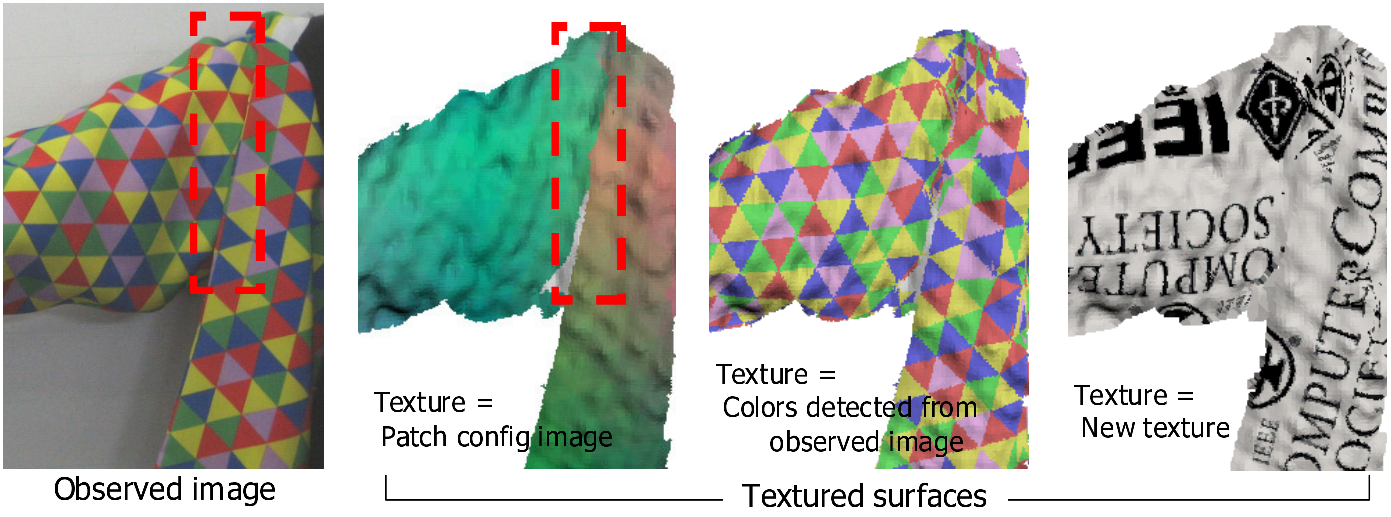


Figure 23: Detailed results of our 3D cloth surface reconstruction with patch configuration: sleeve.

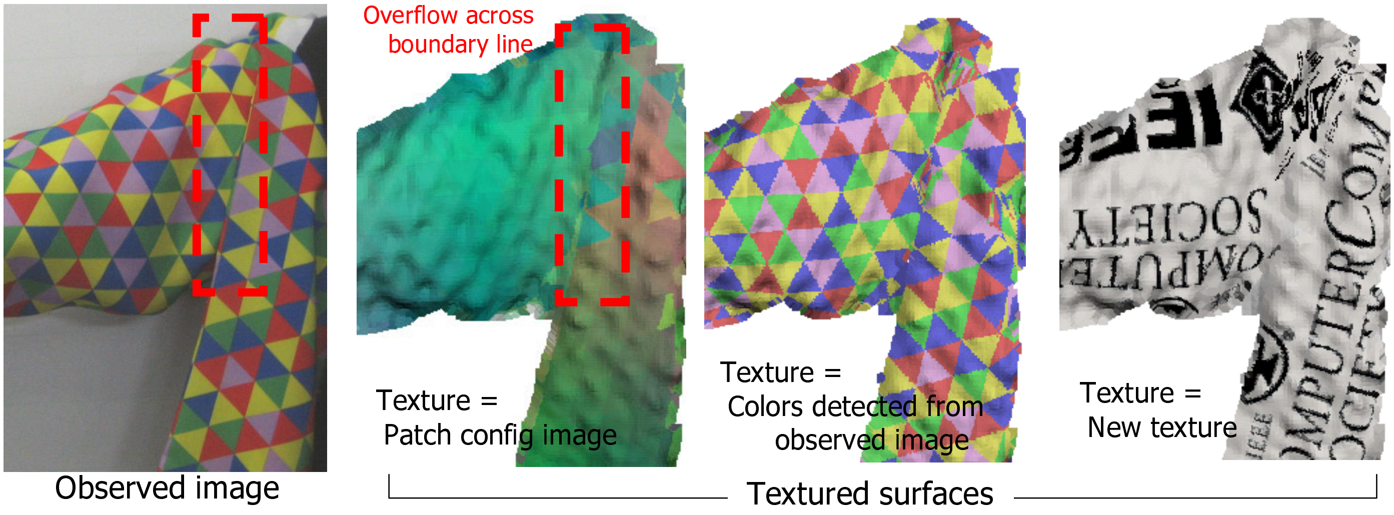


Figure 24: Detailed results of the White-Crane-Forsyth method: sleeve.

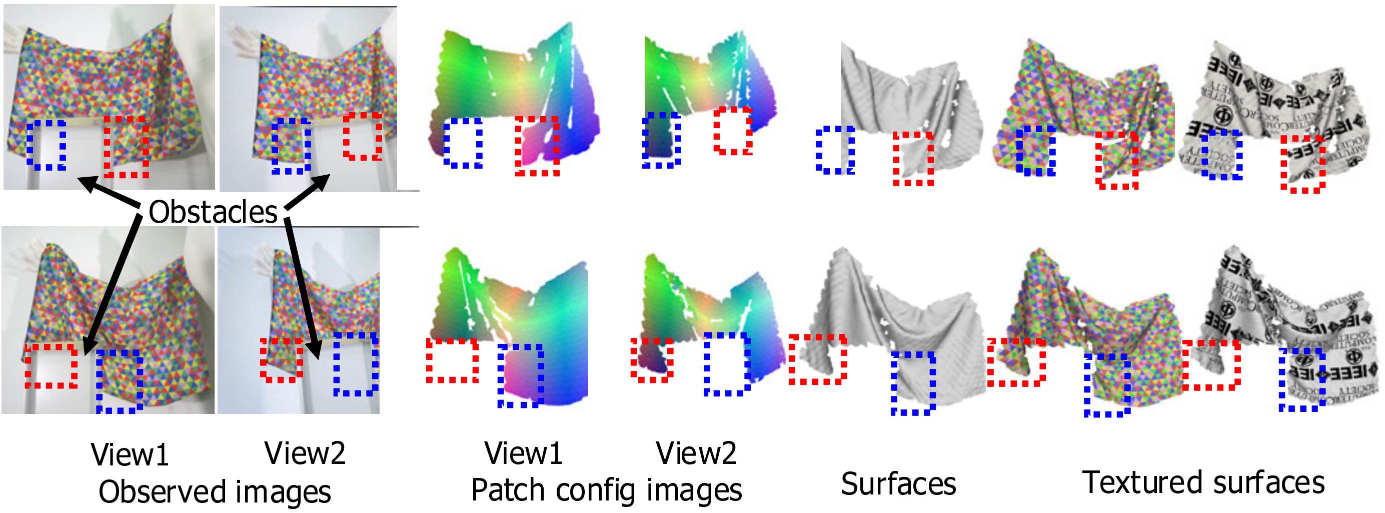


Figure 25: Results of our 3D cloth surface reconstruction with patch configuration: kimono sleeve. In each row, red and blue rectangles enclose the same regions.

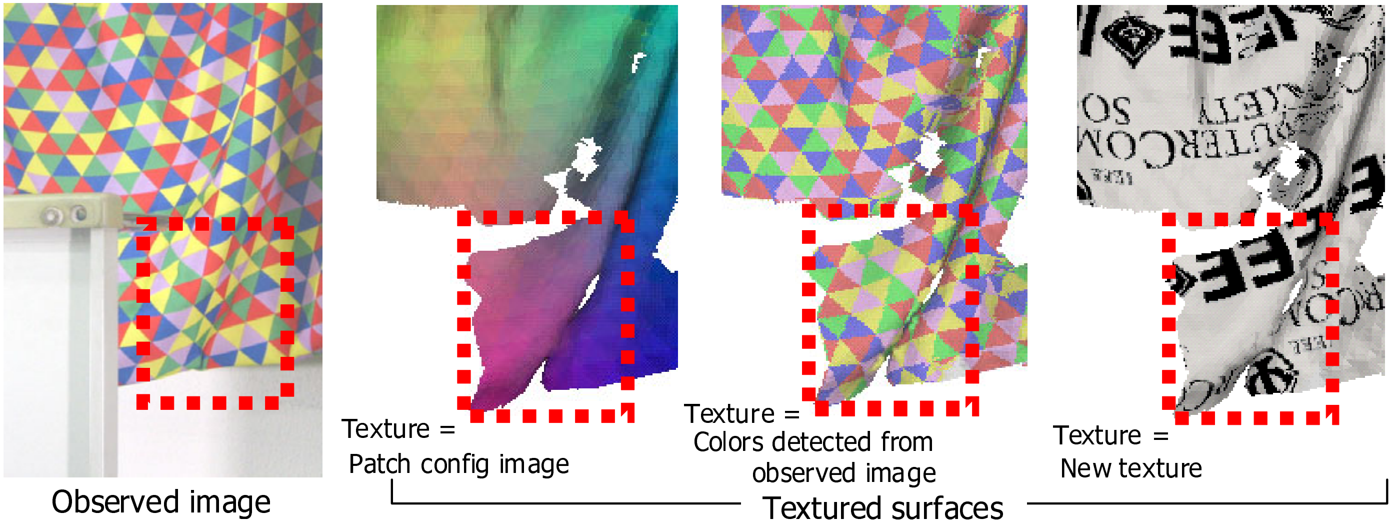


Figure 26: Detailed results of our 3D cloth surface reconstruction.

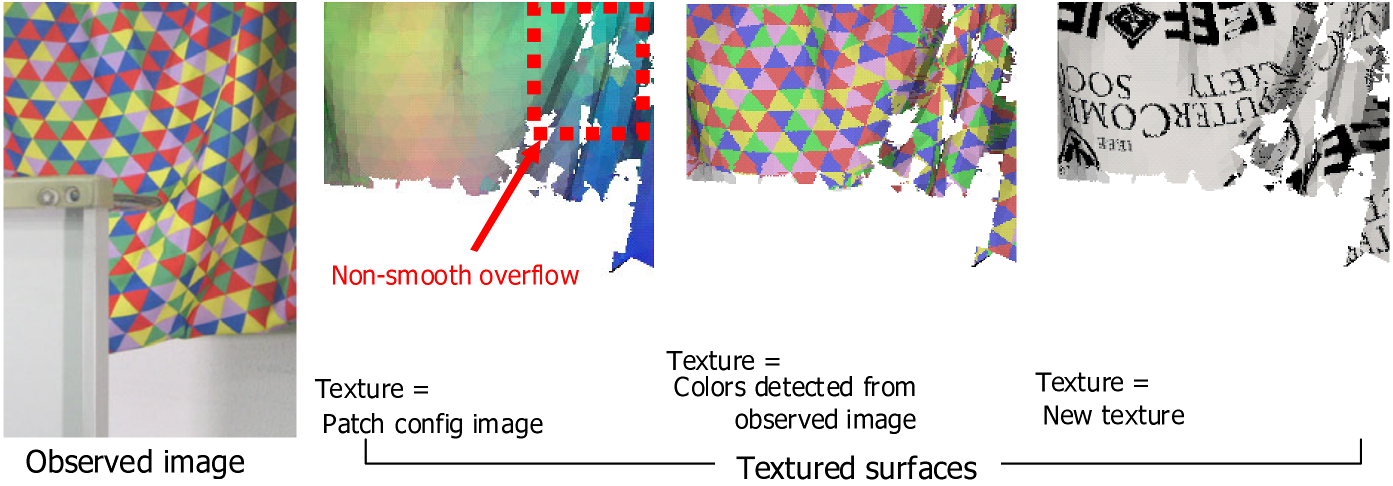


Figure 27: Detailed results of the White-Crane-Forsyth method: kimono sleeve.

on the reconstructed surface is consistent with its neighboring patches and its projection patches in all observed images. With the patch configuration, reconstruction error recovery and single-view reconstruction can be achieved.

Our method improves the correctness and completeness of an existing image-based 3D reconstruction algorithm by the patch configuration, while high spatio-temporal density is guaranteed up to the spatio-temporal resolution of the cameras by the frame-independent image-based reconstruction.

Future work includes reconstruction with many cameras, garment motion capture, and actual applications such as cloth parameter estimation and cloth and garment motion database. For these issues, integration with template-based methods (e.g. [24]) is also important for robust reconstruction.

References

[1] M. Salzmann, J. Pilet, S. Ilic, and P. Fua, "Surface Deformation Models for NonRigid 3D Shape Recovery," *PAMI*, Vol.29, No.8, pp.1481–1487, 2007.

[2] B. Rosenhahn, U. Kersting, K. Powell, R. Klette, G. Klette, H.-P. Seidel, "A system for articulated tracking incorporating a cloth model," *Machine Vision and Applications*, Vol.18, No.1, pp.25–40, 2007.

[3] A. O. Balan and M. J. Black, "The Naked Truth: Estimating Body Shape Under Clothing," *ECCV*, 2008.

[4] D. Baraff and A. P. Witkin, "Large Steps in Cloth Simulation," *SIGGRAPH*, 1998.

[5] R. Bridson, R. Fedkiw, and J. Anderson, "Robust treatment of collisions, contact and friction for cloth animation," *SIGGRAPH*, pp.594–603, 2002.

[6] K. S. Bhat, C. D. Twigg, J. K. Hodgins, P. K. Khosla, Z. Popovic, and S. M. Seitz, "Estimating Cloth Simulation Parameters from Video," Eurographics/SIGGRAPH *SCA*, 2003.

[7] N. Jojic and T. S. Huang, "Estimating cloth draping parameters from range data," *International Workshop on Synthetic-Natural Hybrid Coding and 3-D Imaging*, 1997.

[8] M. Salzmann, R. Urtasun, and P. Fua, "Local Deformation Models for Monocular 3D Shape Recovery," *CVPR*, 2008.

[9] F. Cordier and N. M.-Thalmann, "A Data-Driven Approach for Real-Time Clothes Simulation," *Computer Graphics Forum*, Vol.24, No.2, pp.173–183, 2005.

[10] R. White, K. Crane, and D. Forsyth, "Capturing and Animating Occluded Cloth," *SIGGRAPH*, 2007.

[11] R. W. Sumner, M. Zwicker, C. Gotsman, and J. Popovic, "Mesh-based inverse kinematics," *SIGGRAPH*, 2005.

[12] Z. Zhang, "A Flexible New Technique for Camera Calibration," *PAMI*, Vol.22, No.11, pp.1330–1334, 2000.

[13] V. Kolmogorov and R. Zabih, "Computing Visual Correspondence with Occlusions via Graph Cuts," *ICCV*, 2001.

[14] C. Tomasi and T. Kanade, "Detection and Tracking of Point Features," Carnegie Mellon University Technical Report, CMU-CS-91-132, 1991.

[15] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," *PAMI*, Vol.23, No.11, pp.1222–1239, 2001.

[16] Y. Boykov, O. Veksler, and R. Zabih, "Markov Random Fields with Efficient Approximations," *CVPR*, 1998.

[17] V. Kolmogorov and R. Zabih, "What Energy Functions Can Be Minimized via Graph Cuts?," *PAMI*, Vol.26, No.2, pp.147–159, 2004.

[18] P. L. Hammer, P. Hansen, and B. Simeone, "Roof duality, complementation and persistency in quadratic 0-1 optimization," *Mathematical Programming*, Vol.28, pp.121–155, 1984.

[19] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer, "Optimizing binary MRFs via extended roof duality," *CVPR*, 2007.

[20] Y. Boykov, V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision," *PAMI*, Vol.26, No.9, pp.1124–1137, 2004.

[21] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust Multi-View Stereopsis," *CVPR*, 2007.

[22] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson Surface Reconstruction," *SGP*, 2006.

[23] D. Bradley, T. Boubekeur, and W. Heidrich, "Accurate Multi-View Reconstruction Using Robust Binocular Stereo and Surface Matching," *CVPR*, 2008.

[24] D. Bradley, T. Popa, A. Sheffer, W. Heidrich, and T. Boubekeur, "Markerless Garment Capture," *SIGGRAPH*, 2008.

[25] M. I. A. Lourakis and A. A. Argyros, "Is Levenberg-Marquardt the Most Efficient Optimization Algorithm for Implementing Bundle Adjustment?," *ICCV*, 2005.

[26] G. F. Zhang, J. Y. Jia, T. T. Wong, and H. J. Bao, "Recovering consistent video depth maps via bundle optimization," *CVPR*, 2008.

[27] O. J. Woodford, P. H. S. Torr, I. D. Reid, and A. W. Fitzgibbon, "Global stereo reconstruction under second order smoothness priors," *CVPR* 2008.

[28] Y. Furukawa and J. Ponce, "Dense 3D motion capture from synchronized video streams,", *CVPR*, 2008.

[29] N. Ahmed, C. Theobalt, P. Dobrev, H.-P. Seidel, and S. Thrun, "Robust Fusion of Dynamic Shape and Normal Capture for High-quality Reconstruction of Time-varying Geometry," *CVPR*, 2008.

[30] D. Pritchard and W. Heidrich, "Cloth Motion Capture," *Eurographics*, 2003.

[31] V. Scholz, T. Stich, M. Keckeisen, M. Wacker, and M. Magnor, "Cloth Motion Capture Using Color-Coded Patterns," *Computer Graphics Forum*, Vol.24, No.3, pp.439–448, 2005.

[32] L. Zhang, B. Curless, and S. M. Seitz, "Spacetime Stereo: Shape Recovery for Dynamic Scenes," *CVPR*, 2003.

[33] L. Zhang and S. Nayar, "Projection Defocus Analysis for Scene Capture and Image Display," *SIGGRAPH*, 2006.

[34] C. Hernadez, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla, "Non-rigid Photometric Stereo with Colored Lights," *ICCV*, 2007.

[35] S. G. Narasimhan, S. J. Koppal, and S. Yamazaki, "Temporal Dithering of Illumination for Fast Active Vision," *ECCV*, 2008.

[36] J. Davis, S. R. Marschner, M. Garr, and M. Levoy, "Filling Holes in Complex Surfaces using Volumetric Diffusion," *3DPVT*, 2002.

[37] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *PAMI*, Vol.12, No.7, pp.629–639, 1990.