

Cloth representation by shape from shading with shading primitives

Feng Han and Song-Chun Zhu
Departments of Computer Science and Statistics
University of California, Los Angeles
Los Angeles, CA 90095
hanf@cs.ucla.edu, sczhu@stat.ucla.edu

Abstract

Cloth is a complex visual pattern with flexible 3D shape and illumination variations. Computing the 3D shape of cloth from a single image is of great interest to both computer graphics and vision researches. However, the acquisition of 3D cloth shape by Shape from Shading (SFS) is still a challenge. In this paper, we present a two-layer generative model for representing both the 2D cloth image and the 3D cloth surface. The first layer represents all the folds on cloth, which are called “shading primitives” in [4], and thus captures the overall “skeleton structures” of cloth. We learn a number of typical 3D fold primitives using some training images obtained through photometric stereo. The 3D fold primitives yield a dictionary of 2D shading primitives for cloth images. The second layer represents non-fold parts with very smooth (often flat) surface or shading, which interpolates the primitives in the first layer with a smoothness prior like conventional SFS. Then we present an algorithm called “cloth sketching” to find all the shading primitives on cloth image and simultaneously recover their 3D shape by fitting to the 3D fold primitives. Our sketch representation can be viewed as a 2-layer Markov random field (MRF), and it introduces some prior knowledge on the folds and has lower dimension and is more robust than the traditional shape-from-shading representation which assumes a MRF model on pixels. We show a number of experiments with satisfactory results in comparison to previous work.

1. Introduction

Cloth is a complex visual pattern with flexible 3D shape and shading variations. A compact representation for 2D cloth images and 3D cloth surfaces is important for many applications in both computer graphics, e.g. cloth animation, and computer vision, e.g. human understanding, tracking, and non-photorealistic human portrait and cartoon sketch.

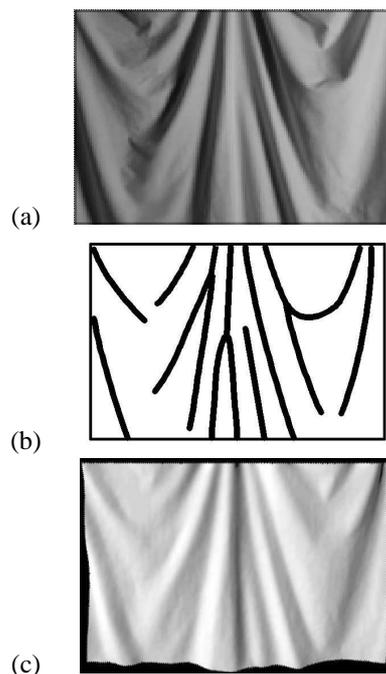


Figure 1. (a). One cloth hung on wall under some lighting. (b). Sketches of folds on the cloth. (c). The computed 3D surface of the cloth.

In the graphics literature, cloth is always represented by a mesh surface with a large number of polygons for geometric based, physical based and particle based cloth modeling and simulation techniques [7, 2].

In computer vision, our objective is to compute the shape of cloth from a single image using mostly the shading information. In the literature, people proposed some shading representations for the folds of cloth [8] and developed methods for detecting the folds from 2D images [4, 5].

However, computing the cloth surfaces using shape-

from-shading (SFS) techniques [13, 11] is still a challenge. The representation underlying the SFS techniques is a Markov random field on the lattice of pixels with smoothness prior to regularize the ill-posed problem (i.e. under-constrained). Such smoothness prior (MRF) only characterizes the changes among nearby pixels and is too weak to model the global information. We show some results on cloths using some highly ranked shape-from-shading algorithms [13] in Fig.10.

In this paper, we present a two-layer generative model for representing both the 2D cloth image and the 3D cloth surface. The first layer represents all the folds on cloth, which are called “shading primitives” in [4], and thus captures the overall “skeleton structures” of cloth. An example is shown in Fig.1. We collect a number of 3D cloths surfaces using photometric stereo (see Fig.6) and manually sketch the various types of folds on the 3D surfaces. Thus we learn a number of typical 3D fold primitives using these training images (see Fig.7). The 3D fold primitives yield a dictionary of 2D shading primitives for cloth images as shown in Fig.8. We represent the 3D folds using an illumination cone model [1]. The second layer represents non-fold parts with very smooth (often flat) surface or shading, which interpolates the primitives in the first layer with a smoothness prior like conventional SFS [13]. Our sketch representation can be viewed as a 2-layer Markov random field (MRF), and it introduces some prior knowledge on the folds and has lower dimension and is more robust than the traditional shape-from-shading representation.

Then we present an algorithm called “cloth sketching” to find all the shading primitives on cloth image and simultaneously recover their 3D shape by fitting to the 3D fold primitives. With the 3D shape of the folds being boundary conditions, we compute the surface of the non-fold part using the shape-from-shading method on the second layer lattice of pixels. We show a number of experiments with satisfactory results in comparison to previous work

The organization of the paper is as follows: Section (2) presents a two-layer representation for both 2D image and 3D surface. Section (3) and Section (4) discuss the learning and inference issues for the new model to do cloth sketching and reconstruction. Then we show the experimental results and comparison in Section (5). The paper is concluded in Section (6) with a summary and future work.

2 Cloth representation

2.1 Two-layer model

Let Λ be the lattice, \mathbf{I} the image, and \mathbf{S} the surface height map defined on Λ . The lattice is divided into two disjoint parts: the pixels on the folds and the rest pixels without

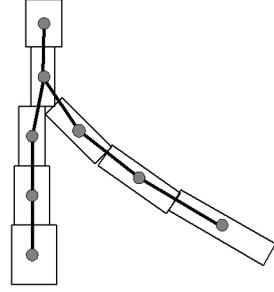


Figure 2. A subgraph of G consisting of fold primitives.

folds,

$$\Lambda = \Lambda_{\text{fd}} \cup \Lambda_{\text{nfd}}.$$

Whether a pixel is on Λ_{fd} or Λ_{nfd} will be inferred in computation. Thus both the image and the surface are divided into two parts,

$$\mathbf{I} = (\mathbf{I}_{\text{fd}}, \mathbf{I}_{\text{nfd}}), \quad \mathbf{S} = (\mathbf{S}_{\text{fd}}, \mathbf{S}_{\text{nfd}}).$$

The image \mathbf{I}_{fd} and surface \mathbf{S}_{fd} are represented by a number of low dimensional fold primitives. We denote these primitive by a set

$$V = \{\pi_i = (\ell_i, \theta_i^{\text{geo}}, \theta_i^{\text{pht}}, \gamma_i^{\text{tpl}}), i = 1, 2, \dots, K\}.$$

Each 3D fold primitive π_i is selected from a learned dictionary Δ (to be introduced shortly), and is specified by four sets of attributes:

1. A label ℓ_i indexing the type of the 3D fold primitive in the dictionary. Fig. 4 shows three type of folds.
2. The geometric transformation θ_i^{geo} for location, orientation, scale (size) and deformation (shape) of the fold primitive.
3. The photometric attributes for illumination θ_i^{pht} : lighting direction and surface albedo.
4. Each primitive is connected to other primitives to form a graph. This is represented by the topological attributes γ_i^{tpl} , which is a set of addressing pointers to the primitives connected with current primitive.

These fold primitives connect with each other like a chain without over-lapping to generate each fold in \mathbf{S}_{fd} , while the fold lattice Λ_{fd} is covered by a number of windows corresponding to these fold primitives as Fig.2 illustrates.

Using each fold primitive as a vertex and denoting neighboring structure among these fold primitives by an edge set

$$E = \{e = (p, g) : \pi_p, \pi_g \in V\},$$

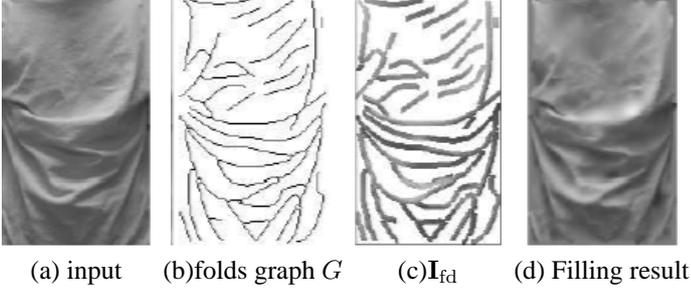


Figure 3. Filling in I_{nfd} by using I_{fld} as boundary condition.

we can further represent the fold layer as an attribute graph $G = (V, E)$. Figure 2 shows an example subgraph.

For each 3D primitive π_i , it can generate an image patch $R_{\Lambda(\pi_i)}$ on window $\Lambda(\pi_i)$ based on its attributes:

$$R_{\Lambda(\pi_i)}(x, y) = B(\ell_i, \theta_i^{\text{geo}}, \theta_i^{\text{pht}}),$$

where $B()$ is the Lambertian reflectance model process.

Then we can generate the whole pixels in Λ_{fld} as,

$$I_{\text{fld}}(x, y) = R_{\Lambda(\pi_i)}(x, y), \forall (x, y) \in \Lambda_{\text{fld}}(\pi_i).$$

As to the pixels in Λ_{nfd} , they can be filled in by using the pixels in Λ_{fld} as boundary condition with some smoothness prior.

$$I_{\text{nfd}}(x, y) = \arg \max p(I_{\text{nfd}}(x, y) | I_{\text{fld}}(x, y), \beta),$$

where β is the parameter to control smoothness prior. Figure 3 shows such an example.

Similarly, each 3D primitive π_i can also generate a depth patch $D_{\Lambda(\pi_i)}(x, y)$ in 3D. So the S_{fld} can be generated as,

$$S_{\text{fld}}(x, y) = D_{\Lambda(\pi_i)}(x, y), \forall (x, y) \in \Lambda_{\text{fld}}(\pi_i).$$

As to the 3D shape of non-fold areas $S_{\text{nfd}}(x, y)$, they can be computed by traditional SFS by using $S_{\text{fld}}(x, y)$ as boundary condition. With this recovered surface plus lighting, we can have the other way to generate I_{nfd} . Usually, the result is almost the same as the above method.

2.2 Two-layer representation based SFS model for cloth reconstruction

Assuming a Lambertian reflectance model and S_{fld} and S_{nfd} share the same illumination and constant surface albedo, with this two-layer representation, we can formulate computing cloth surface S by Shape from Shading in Bayesian framework as:

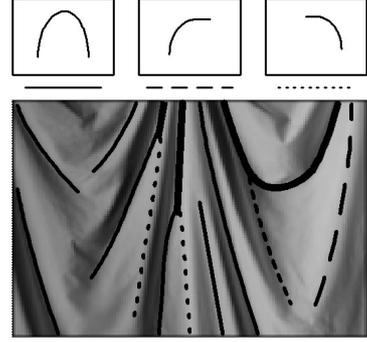


Figure 4. Three types of folds defined on cloths are shown at the top. The folds in the cloth image are marked with different format of lines to show the type.

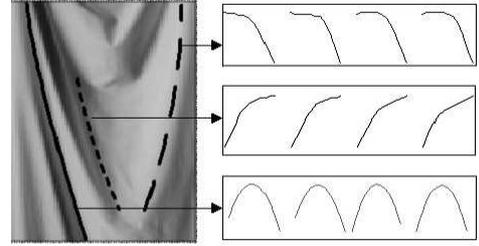


Figure 5. Some typical surface cross-section profiles for the three type of folds showed in Fig. 4

$$\begin{aligned} S &\sim p(S|I) \\ &\sim p(S_{\text{fld}})p(I_{\text{fld}}|S_{\text{fld}})p(I_{\text{nfd}}|S_{\text{nfd}}, S_{\text{fld}}). \end{aligned}$$

In this model, the fold part of image I is explained by an unknown number of low dimensional 3D fold primitives. The non-fold part of image I is explained by S_{nfd} , using S_{fld} as boundary condition. $p(S_{\text{fld}})$ represents the spatial regularity for the folds, which is defined on the attribute graph G .

3 Learning

3.1 Learning 3D fold primitives

To learn the 3D fold primitives to represent the 3D shape of folds, we divide all the folds into three types as shown in Figure 4. The first type is the regular folds seen from front

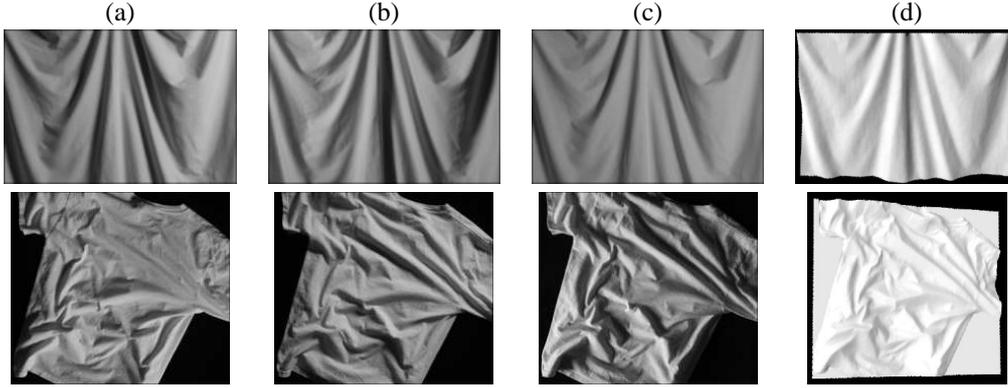


Figure 6. (a), (b), (c) are three images out of the sequence used to reconstruct the 3D cloth shape in (d).

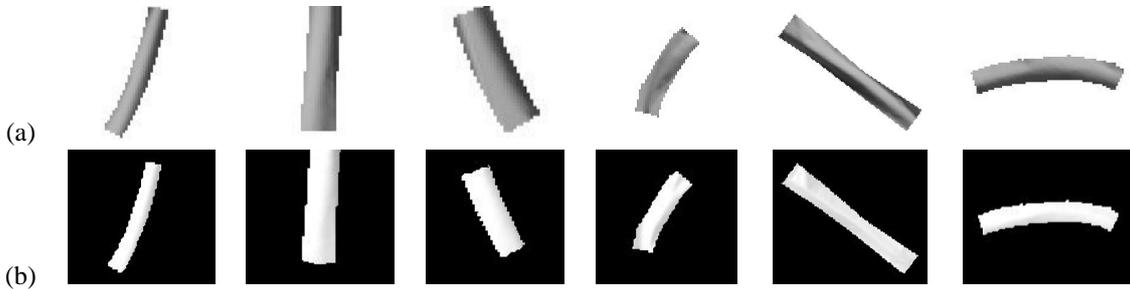


Figure 7. (a). The 2D appearance of 6 extracted fold patches. (b). The 3D shape of 6 extracted fold patches rendered in OpenGL.

view, while the other two are those half-folds seen from side view. Figure 5 shows some typical cross-section profiles of these three types of 3D folds based on the 3D surface of the cloth in Figure 1 obtained by photometric stereo [10]. To get some 3D cloth surfaces as training data, we use the photometric stereo algorithm in [10]. For each of the sample cloths, we take a sequence of images (~ 20) under different lighting conditions. Two sample cloths are shown in Figure 6, in which (a), (b), (c) are three images from the sequence used to get the cloth surface in (d). Based on the 3D data of sample cloths, we build up an interface program to help manually extract fold patches as training data to learn the 3D fold primitives. Some of typical extracted fold patches are shown in Figure 7. It can be clearly seen from Figure 5 that the 3D shapes are consistent for each type of 3D folds. Therefore, we use PCA to represent the shapes of these three types of fold primitives.

Thus, we have a dictionary with three types of 3D primitives to represent folds,

$$\Delta = \{B_1, B_2, B_3\}.$$

Each primitive B_i is represented by the coefficient of the eigenfunctions of the PCA model. The mean shape and eigenfunctions for 3D primitives B_i are learned from the training patches extracted from the 3D surfaces of sample cloths. Figure 8 shows the learned mean fold shape under different viewing directions and lighting conditions. It shows that these 3D fold primitives generate some 2D shading primitives on images.

In the attributed graph G of folds, each vertex is a 3D fold primitives B_i from the dictionary Δ , but goes under some translation, rotation, scaling and deformation (changing the coefficients of eigenvectors) of the unit fold primitive with mean shape as height map.

3.2 Learning spatial regularity prior model for folds $p(G)$

The folds on cloths are not randomly spreading in space. Instead, they follow some spatial regularities, not only for each individual fold, but also for the relative spatial rela-

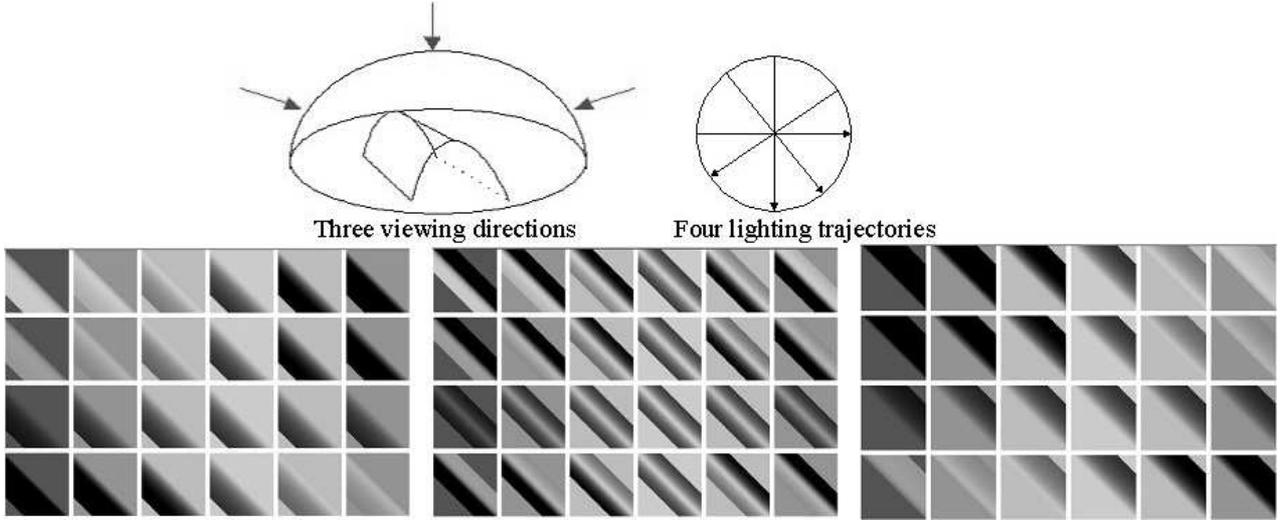


Figure 8. The rendering results for the learned mean fold shape under different viewing directions and lighting conditions.

tions among folds.

For each individual fold, the overall shape should be smooth in 3D space without sudden change. To enforce this regularity, we use a Markov chain model to force the smoothness of 3D folds.

Let $f_i, i = 1, 2, \dots, N_f$ be all the folds in G and $v_{ij}, j = 1, 2, \dots, |f_i|$ be all the vertices on fold f_i . The smoothness prior model for fold f_i can be represented as,

$$p(f_i) = p(v_{i1}, v_{i2})p(v_{i3}|v_{i1}, v_{i2}) \prod_{j=4}^{n_i} p(v_{ij}|v_{i,j-1}, v_{i,j-2}, v_{i,j-3})$$

The probability $p(v_{i1}, v_{i2})$ is assumed to be uniform, $p(v_{i3}|v_{i1}, v_{i2})$ is a two gram represented by a 2-way joint histogram and $p(v_{ij}|v_{i,j-1}, v_{i,j-2}, v_{i,j-3})$ is a trigram representation by three way joint histogram. The first histogram is learned from some 2D curves of folds, while the second histogram is learned from some manually obtained 3D curves of folds by computing three variables:

1. the angle between $(v_{i,j-1}, v_{i,j-2})$ and $(v_{i,j-2}, v_{i,j-3})$,
2. the angle between $(v_{i,j-1}, v_{i,j-2})$ and $(v_{i,j-1}, v_{i,j})$,
3. the distance from $v_{i,j}$ to the plane fitting through $v_{i,j-1}, v_{i,j-2}$ and $v_{i,j-3}$.

So the spatial regularity prior model for folds is,

$$p(G) = \prod_{i=1}^{N_f} p(f_i).$$

4 Inference

The two-layer representation based SFS model for cloth reconstruction may need MCMC method for global inference. Here we propose a two step greedy method: First, we run a process called “cloth sketching” to find all the folds and recover their 3D shape at the same time using the 3D fold primitives in learned dictionary Δ . Second, after obtaining the 3D shapes of fold areas, we infer the 3D shape of non-fold areas by using these 3D folds as boundary condition.

4.1 “Cloth sketching” process

In this process, we try to find all the folds in the given image I and recover their 3D shape simultaneously. The process is explained as below.

1. run a ridge detection algorithm [6] on the image I .
2. Initialize the attribute graph G of S_{fd} to \emptyset and S_{nfd} to be a constant plane.
3. Find the highest ridge strength position (x_0, y_0) , which is not covered by G and is not marked as visited. At this position, mark it as visited and fit the three types of fold primitives with different scale, rotation, and deformation to get the one with largest log-posterior ratio. If this largest log-posterior ratio is larger than a threshold, then represent this scaled, rotated and deformed fold primitive by a vertex and insert it to G and go to step 4; otherwise stop.

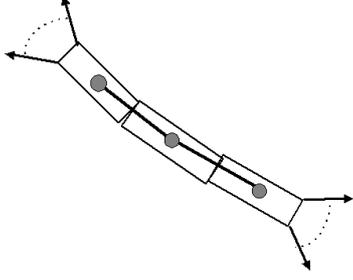


Figure 9. To grow the newly inserted vertex in fold graph G , we test the areas as illustrated in the figure.

4. Try to grow the newly inserted vertex from both ends, which is shown in Figure 9. Do the log-posterior ratio test as in step 3. If the largest log-posterior ratio is larger than a threshold, insert a new vertex and continue to grow until the grow operation is rejected for both ends.
5. Repeat step 3 and 4 until all the positions are either visited or covered by G .

After this process, we have the attribute graph G for \mathbf{S}_{fld} . From this, we can synthesize the image in fold part and get the 3D shape of folds, which are shown in Figure 11 (c) and (d) respectively.

4.2 Infer the shape of non-fold parts

After finding all the folds and obtaining their 3D shape, we infer the shape of rest parts by SFS using these folds shape as boundary condition. Since we don't know the relative positions among all the folds yet, we try to recover the normal of non-folds areas first. Then we recover the depth from the obtained normals. Since the recovered 3D folds can not only give us a good initialization for the 3D shape of non-fold areas, but also act as extra constraints to dramatically constraint the solution space, this part can be done by a lot of existing SFS algorithms. Considering both speed and accuracy, we choose a latest one based on energy minimization in [3] with some modifications.

Denoting the normals for \mathbf{S}_{nfd} in (p, q) format, the energy to be minimized in [3] is modified as:

$$E = \zeta^2 \sum_{(x,y) \in \Lambda_{\text{nfd}}} \delta(\mathbf{I}(x, y) > \tau) (\mathbf{I}(x, y) - R(x, y))^2 + \lambda_{\text{int}} \sum_{(x,y) \in \Lambda_{\text{nfd}}} (p_y(x, y) - q_x(x, y))^2$$

$$+ \lambda_{\text{smo}} \sum_{(x,y) \in \Lambda_{\text{nfd}}} (w_1(x, y) p_x(x, y)^2 + w_2(x, y) p_y(x, y)^2 + w_2(x, y) q_x(x, y)^2 + w_1(x, y) q_y(x, y)^2),$$

where $p_x(x, y) = p(x + 1, y) - p(x, y)$, $p_y(x, y) = p(x, y + 1) - p(x, y)$, $q_x(x, y) = q(x + 1, y) - q(x, y)$, $q_y(x, y) = q(x, y + 1) - q(x, y)$, ζ is the distance between two neighboring pixels, λ_{int} and λ_{smo} are two positive constants named ‘‘integrability factor’’ and ‘‘smoothing factor’’ respectively, while $\delta(\cdot)$ and w will be introduced next.

The non-fold areas are more noisy than the fold areas since the former always has occlusions and sharp valleys, where the assumed Lambertian reflectance model doesn't hold anymore. Therefore, these noisy areas should not be counted in the data term and their shape be recovered by the prior model. Since these noisy areas are always very dark, we filter them out by a threshold τ with a delta function (e.g. $\delta(\mathbf{I}(x, y) > \tau) = 1$).

In addition, we weight the smoothness prior by looking at the data as in [12] with $\{w_j(x, y), j = 1, 2, 3\}$ being local smoothness weights:

$$w_1(x, y) = (1 - |\mathbf{I}_x(x, y)|)^2$$

$$w_2(x, y) = (1 - \frac{\sqrt{2}}{2} |\mathbf{I}_x(x, y) + \mathbf{I}_y(x, y)|)^2$$

$$w_3(x, y) = (1 - |\mathbf{I}_y(x, y)|)^2$$

which are chosen to be inversely proportional to the intensity gradient along the x , diagonal, y directions respectively. This choice is intuitive and confirms well with the fact that smoother images should be produced by smoother surfaces in usual cases.

After obtaining the normals for non-folds areas, we can compute the whole 3D cloth shape as in [3] by minimizing an energy function. (Refer to [3] for details.)

5 Experiments

We test our whole algorithm on four cloth images as shown in Figure 11. The first three are big cloths hanged on wall, while the last one is a patch extracted from a clothing on a person. In the experimental results, the first row are input images, second row are the sketches of folds in the input images, third row are the synthesises based on the generative sketch model for the fold areas, third row are the 3D reconstruction results for the fold areas, while fourth and fifth rows are the final reconstruction results of the whole cloth shown in two different views.

For comparison with other SFS algorithms, we run two minimization approaches in [13] on the same testing images used for our algorithm since minimization approaches are more robust and accurate even though much slower than other approaches. The first approach is from [14], while the

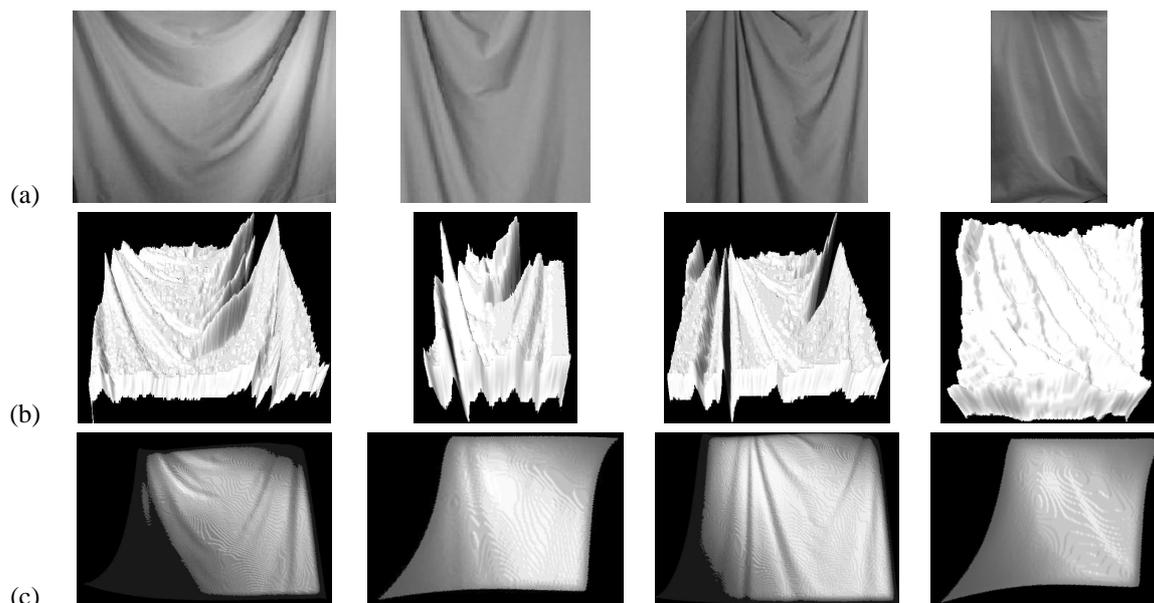


Figure 10. (a). input cloth images. (b). cloth reconstruction results by approach in [14]. (c). cloth reconstruction results by approach in [9].

second one is from [9]. The results for these two approaches are shown in first row and second respectively in Figure 10.

6 Summary and Future Work

In this paper, we present a two-layer generative model for representing both the 2D cloth image and the 3D cloth surface. The first layer represents all the folds on cloth with some low dimension 3D fold primitives, while the second layer represents the non-fold part. Based on this model, the 3D shape of folds are recovered by a process called “cloth sketching” first and then the shape of non-fold areas are recovered by using these fold shapes as boundary condition. In the future work, we will learn more 3D shape primitives and extend the “cloth sketching” process to recover more structured parts. In this way, we can use it for other objects than cloth.

Acknowledgements

This work was supported in part by National Science Foundation grants IIS-0222967 and IIS-0244763.

References

- [1] P. Belhumeur and D. Kriegman. What is the set of images of an object under all possible illumination conditions? *IJCV*, 1998.
- [2] K. Bhat, C. D. Twigg, J. K. Hodgins, P. K. Khosla, Z. Popvic, and S. M. Seitz. Estimating cloth simulation parameters from video. *Proc. Symposium on Computer Animation*, 2003.
- [3] A. Crouzil, X. Descombes, and J.-D. Durou. A multiresolution approach for shape from shading coupling deterministic and stochastic optimization. *TPAMI*, 2003.
- [4] J. Haddon and D. Forsyth. Shading primitives: Finding folds and shallow grooves. *ICCV*, 1998.
- [5] J. Haddon and D. Forsyth. Shape representations from shading primitives. *ECCV*, 1998.
- [6] R. Haralick. Ridges and valleys on digital images. *CVGIP*, 1983.
- [7] D. H. House and D. E. Breen. *Cloth Modeling and Animation*. A.K. Peters, Ltd., 2000.
- [8] P. S. Huggins, H. F. Chen, P. N. Belhumeur, and S. W. Zucker. Finding folds: On the appearance and identification of occlusion. *CVPR*, 2001.
- [9] K. Lee and C. Kuo. Shape from shading with a linear triangular element surface model. *TPAMI*, 1993.
- [10] A. Sashua. Geometry and photometry in 3d visual recognition. *Ph.D Thesis, MIT*, 1992.
- [11] P.-S. Tsai and M. Shah. Shape from shading using linear approximation. *Image and Vision Computing*, 1994.
- [12] G.-Q. Wei and G. Hirzinger. Parametric shape-from-shading by radial basis functions. *TPAMI*, 1997.
- [13] R. Zhang, P.-S. Tsai, J. Cryer, and M. Shah. Shape from shading: A survey. *TPAMI*, 1999.
- [14] Q. Zheng and R. Chellappa. Estimation of illumination direction, albedo, and shape from shading. *TPAMI*, 1991.

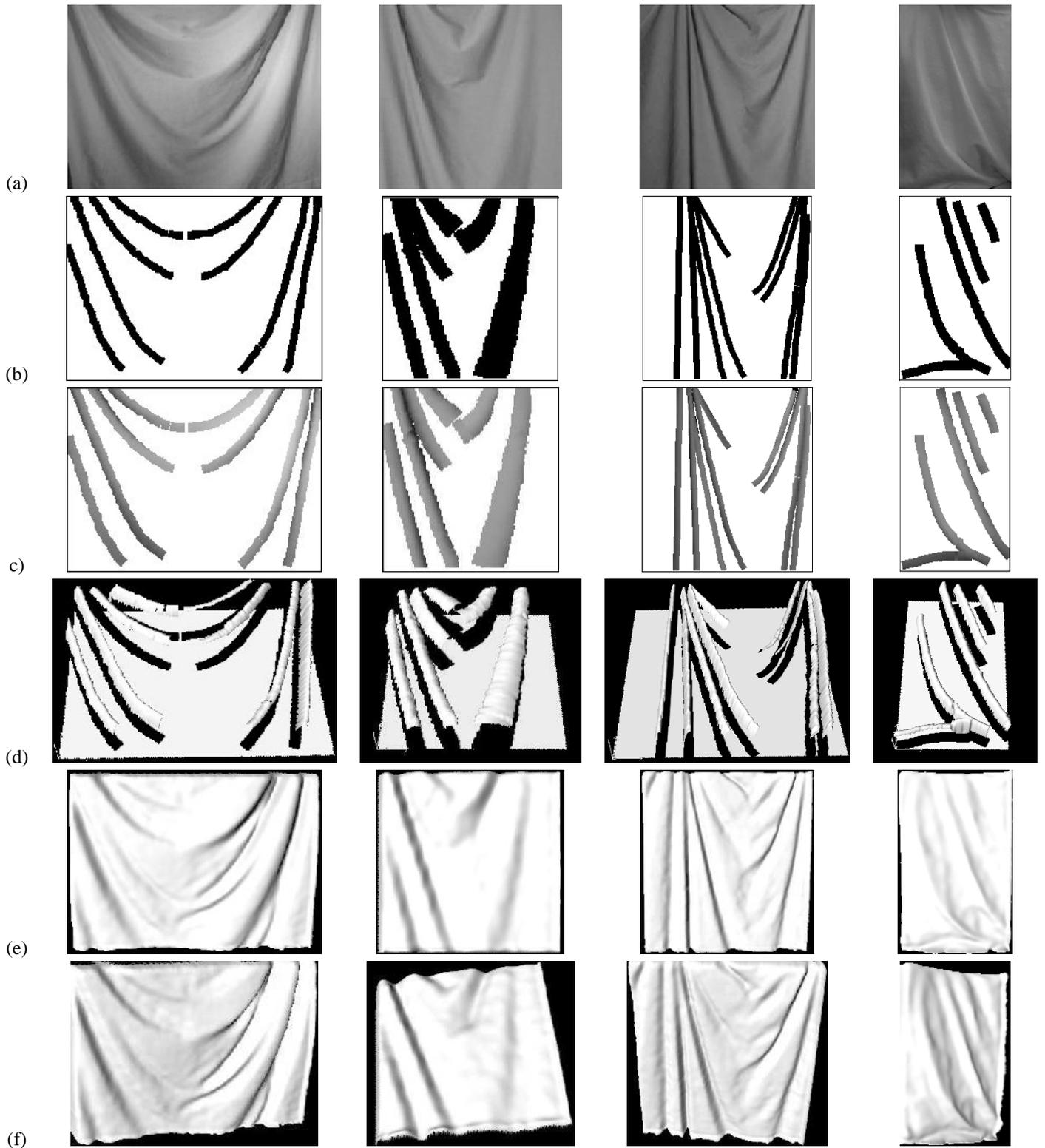


Figure 11. (a). input cloth image. (b). 2d fold sketches. (c). synthesis for 2D fold sketches. (d). 3D reconstruction results for fold areas. (e). final reconstruction results for the whole cloth. (f). final reconstruction results for the whole cloth shown in a novel view.