國立交通大學

多媒體工程研究所

碩 士 論 文

單一視角影片中折疊表面之深度重建

Foldable 3D Surface Reconstruction from Single-view Video

研 究 生：蔡明翰
指導教授：林奕成　教授

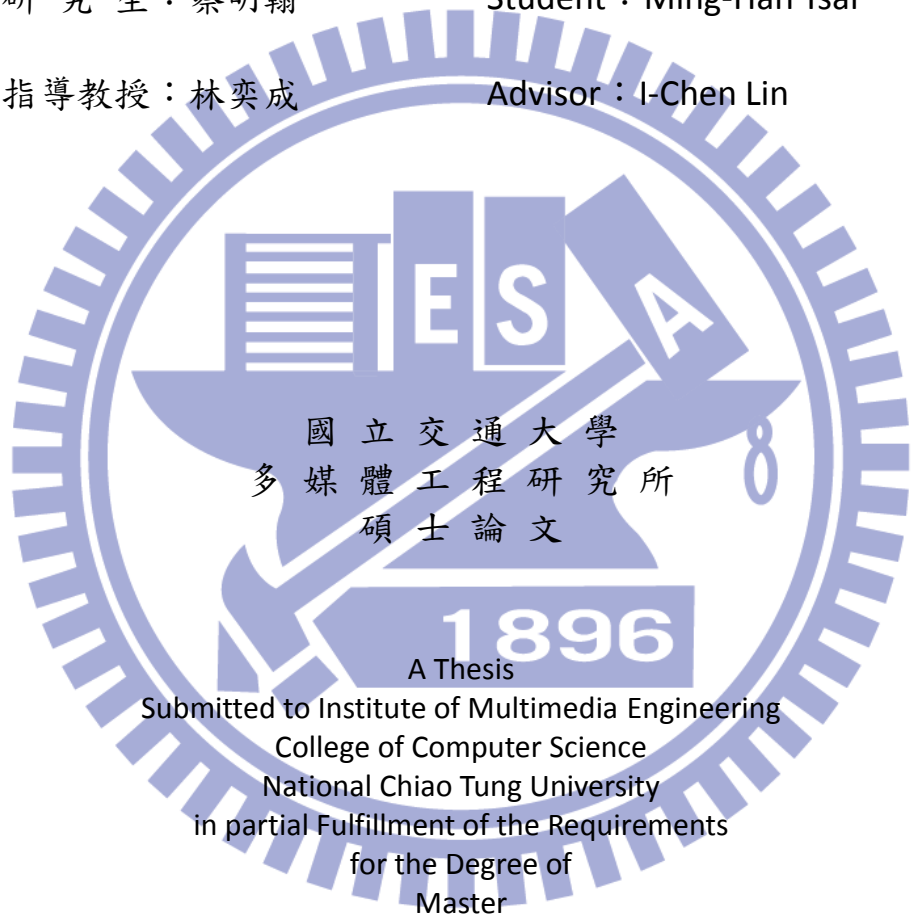
中 華 民 國　九 十 八　年 七 月

單一視角影片中折疊表面之深度重建

Foldable 3D Surface Reconstruction from Single-view Video

研 究 生：蔡明翰　　Student：Ming-Han Tsai

指導教授：林奕成　　Advisor：I-Chen Lin



國 立 交 通 大 學
多 媒 體 工 程 研 究 所
碩 士 論 文

A Thesis
Submitted to Institute of Multimedia Engineering
College of Computer Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master
in

Computer Science

July 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年七月

單一視角影片中折疊表面之深度重建

研究生：蔡明翰　指導教授：林奕成　助理教授

國立交通大學

多媒體工程研究所

摘要

這篇論文提供了新的方法還原影片中有皺折的表面的深度值，像是衣服或飄動的旗子，並使其結果可應用在立體顯示器上. 我們的方法結合了 Shape-from-shading 和基於主軸分析的子空間近似法，使結果能保留表面的高低起伏變化，同時不會受影片中的雜訊影響.

關鍵字: 深度還原，由明暗還原形狀，影片處理.

Foldable 3D Surface Reconstruction from Single-view Video

Student: Ming-Han Tsai Advisor: Dr. I-Chen Lin

Institute of Multimedia Engineering

National Chiao Tung University

Abstract

In this thesis, we propose a novel method to reconstruct fluttering surface such as flags and cloth in the video sequence which can be used for the 3D display. Shape recovery of real object from a video sequence is a difficult subject. Here, we focus only on fluttering surface which can possibly be folded. While most shape-from-shading can only deal with single-material smooth objects, we propose using shape-from-shading and decoloring techniques to reconstruct more detailed surfaces under a single directional lighting condition, the surface can be multi-material with folding. To alleviate the noise and ill-pose problem, we take shape-from-shading as initial-guess, and further use Principle Component Analysis (PCA)-based subspace approximation to recover full video sequence.

With the proposed method, users only have to designate the flag by graph-cut-based tool. We can then automatically recover waving flag's 3D geometry and change its texture. Our results demonstrate that our system work satisfactorily even under a noisy situation, and provide a reasonable solution for free-view point content generation.

Keyword: Depth Recovery, Shape-From-shading, Video Editing.

致謝

在二年的碩士學習生涯中，受到了許多師長及朋友們的協助，能夠順利的完成論文，我由衷地感謝他們：

1. 首先得感謝我的母親和家人，他們始終給予最大的支持與鼓勵，給予我最大的學習自由，是我最強力的精神後盾及心靈的避風港。
2. 指導教授 林奕成博士，是碩士生涯中的良師益友，總是能提出精湛的見解，指引我研究的方向。在兩年的碩士中，不僅教導我研究的專業知識，也讓我學習許多待人處事的道理。
3. 實驗室的各位同學，因為有他們的陪伴，讓我的二年生活過的多采多姿。無論學業上的切磋砥礪、課餘的休閒活動，都是令我難忘的回憶。

Content

1. Introduction

1.1 Motivation

Televisions are important in every family. They bring not only information but also entertainment to us. In recent years, televisions have dramatic improvements, especially in size and resolution, and one remarkable technology of them is the free viewpoint 3D display. In the near future, one can respect to see 3D movie at home.

To show a video without view limitation, we display not only the color frame channels but also the depth of the scene. New 3D content videos are recorded by a special camera with two view points for depth estimation. In computer graphics-based video, since the scenes were rendered according to polygonal models, depths were retrieved inherently. Unfortunately, conventional video doesn't have depth channel so it cannot directly be displayed over 3D display. That is the goal of this thesis.

1.2 Background

There are many methods proposed to generate depth maps, and it's a prevailing approach by using multiple views which requires synchronously captured images and accurate pixel correspondence in each view. Due to the scene ambiguity (not-rigid body、similar color object、occlusion), it is a challenge to full automatically recover depth map of a general scene by the multiple-view technique. Even though, with manual assistants, multiple views can get a more effective result. However, traditional popularly–used videos, like existing DVD movies, were not captured by multiple views. For these reason, we want to find out an efficient and practical approach such

that we can generate sequence of depth map toward the target object from the common video sequence.

The challenge is : we have no any clues except image color in the video. The surface is undulation so it cannot be regarded as rigid body; the video was taken by a single-view point so we cannot get synchronized potion parallax as these in binocular views. Besides, the low-resolution and noise of video make it more difficult.

Our goal is to recover the depth maps of target surface from a general single-view video sequence like movies or family video illuminated at a simple lighting condition. Because the target surface in the video sequence can be folded, of multi-material, or with non-rigid-body properties, we combine the SFS and the PC-basis Morphable model with space-time optimization to approximate its surface as real as possible.

1.3 Framework

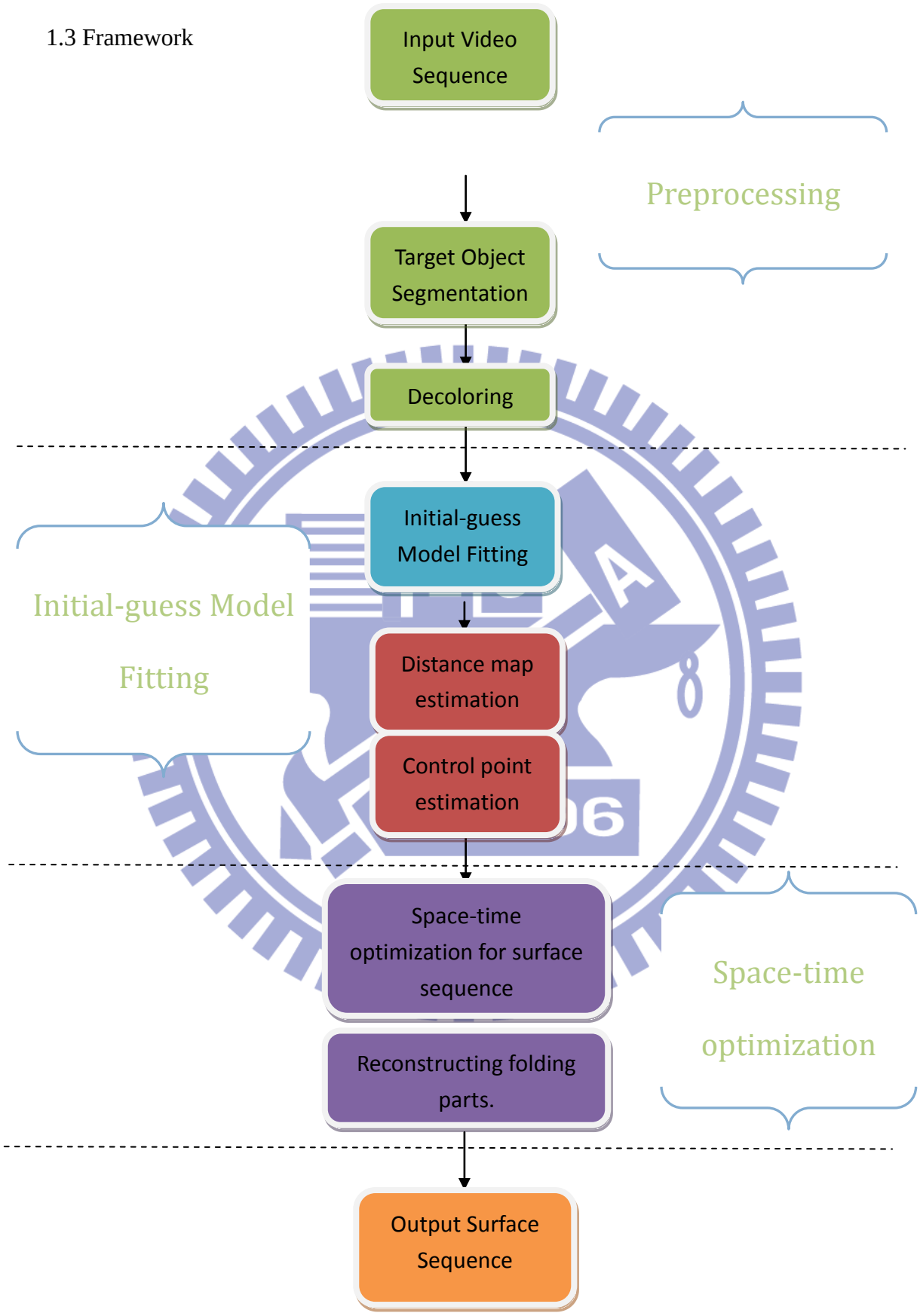
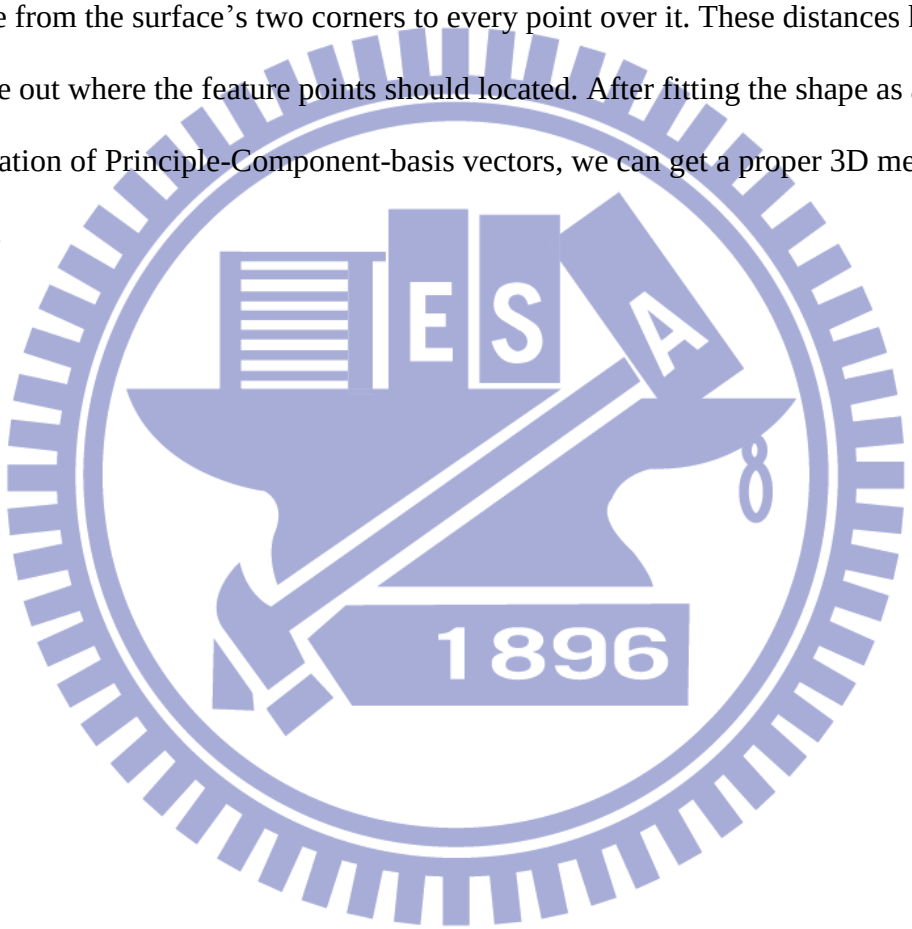


Fig1.Framework of our system

In this paper, we focus on fluttering surface such as flags and clothes. Our system has three parts: In pre-processing step, we take one video sequence as input and segment the target object frame by frame, and then remove the color component of each frame but keep the shading.

After decoloring, we apply SFS on each frame. Although the result is still imperfect, it provides a reasonable initial-guess depth map. Then we calculate all distance from the surface's two corners to every point over it. These distances help us to figure out where the feature points should located. After fitting the shape as a linear combination of Principle-Component-basis vectors, we can get a proper 3D mesh surface.



2. Related work

This section has three parts: First of all, we introduce geometry-based 3D construction method like Blender3D and Maya. Then we show image-based reconstruction techniques such as Structured-line, Multi-view reconstruction, Photometric stereo, and SFS. Finally, we describe some applications of reconstruction.

2.1 Geometry-based Method

The first approach is interactive modeling with manual assistance, like Blender3D、Maya. Through these tools we can build a model manually, but it is labor-intensive to create a photorealistic result. Several intelligent or hybrid modeling system were proposed to reduce manual intervention. [Hengel et al. 2007.] proposed building a realistic 3D models from video by point clouds with a small number of simple 2D sketches as constraints.

[Debevec et al. 1996] provided another method to construct 3D model from video. Users needed only draw some structure lines, and the system then built 3D buildings and retargeted the texture over it.

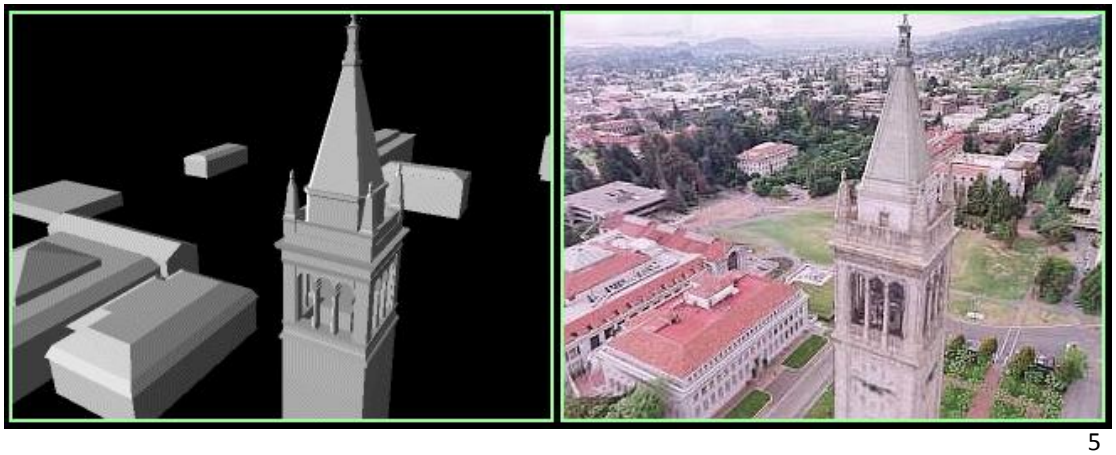


Fig2.P. Devebec et al's 3D reconstrunction from video.

2.2 Active Structured-line System

By contrast, the second kind of approach using active structured light system is faster and more convenient. It is also the main stream of the high-accuracy 3D recovery. In the early years, 3D scanning technique was only suitable for static objects, and it needed more scanning time. [S.Rusinkiewicz et al. 2002.] developed the system based on the structured-light system and a real-time variant of ICP(iterative closest points) to align the shapes acquired from multiple views. It made the significant impact of rapid 3D recovery. [L.Zhang et al. 2004.] used the consistent space-time stereo technique to enhance the reliability of acquired 3D data. By usage of the structure-light system, we can precisely estimate the shape of the target object, but it still has several defects limiting its usability. The target object is limited to be a nearly-lambertian object and not suitable for the one too big or the scene outdoors.

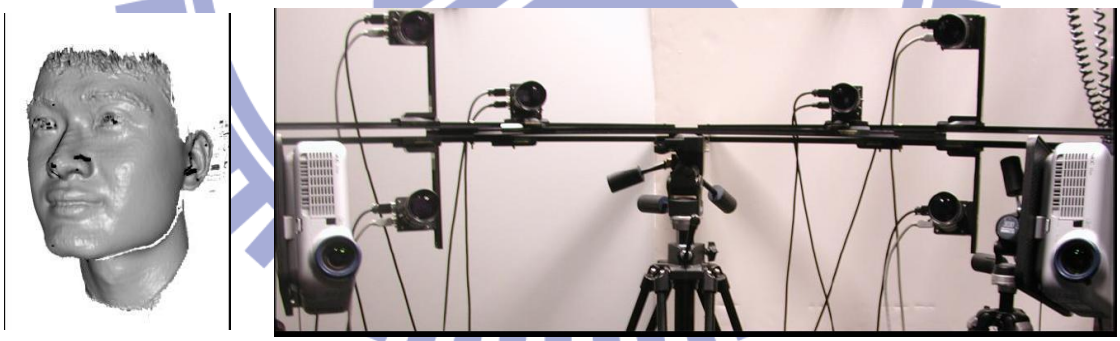


Fig3. L.Zhang et al.’s facial detail reconstruction system and depth data

2.3Multiple views

Multiple view technique plays an important role in reconstruction. With well calibration and correspondence matching, we can reconstruct scene’s surface. But it’s cost-high to find the reliable correspondence matching at untextural or highly

repeated regions and the occlusion of correspondences are also the problem. Even though, it's still widely used as the constrain to other technique or a coarse shape recovery.

[Vogiatzis et al. 2005.] proposed a novel technique combined multi-view stereo with graph-cut based optimization for detailed surface reconstruction. They used the visual hull as the initial shape, and then defined a continuous photo-consistency function as a flow graph to minimize the detailed surface.

Structure form motion is the same technique but use only one single camera instead and suitable for the moving rigid or static object. With such a uncalibrated property, it is more suitable for the common video sequence.

[Pollefeys et al. 2006.] used the corner detection to find out the feature points, and then found out the correspondences by use of the epipolar geometric properties. The affine transformations between multiple-views were therefore acquired.

Fig4. Pollefeys et al.'s Reconstructed model and the view points

If there were few correspondences, only sparse 3D points can be estimated. [M. Lhuillier et al. 2005.] proposed an approach to generate quasi-dense 3D points toward

the surface with fewer feature points. They produced a dense disparity map and used it to improve numbers and qualities of the feature correspondences matching by the correlation method. Moreover, they proposed a fast gauge-fee algorithm to estimate the accuracy of the recovered 3D depth.

For the non-rigid body, [Torresami et al. 2003.] proposed a method combined with structure from motion to recover the target shape from the video. They defined a non-rigid body as a rigid transformation combined with a non-rigid deformation in the time frames. Under the assumption that the object shape at each time frame was organized from a Gaussian distribution, they simultaneously estimated 3D shapes in each time frame, learned the parameters of the Gaussian, and also recovered the missing data points. Finally, they implemented the space-time constrain to the object shape for the better consistent result.

2.4 Photometric stereo

Photometric stereo estimates local surface orientation by using several images of a surface taken from the same viewpoint but under illumination from different directions.

Fig5. The left two images are the reconstructed surface by the M. seitz et al's method. The right four images are the reference and target object used for

[M. seitz et al. 2004.] proposed the example-based photometric stereo method. They introduced orientation-consistency concept to reconstruct the surface normal from the reference images where the reference objects with identical materials were also taken. Combined with traditional photometric stereo, a more detailed surface can be recovered. The technique is reliable to be applied to a broader class of objects than previous photometric stereo technique.

.

[Carlos et al. 2008.] used the silhouettes in multiple views to recover camera motion and then got a coarse shape of the object by the visual hull. Besides, they proposed a robust technique to estimate light directions and introduced a novel formulation to combine photometric stereo and 3D points from visual hull.

2.5 Shape-from-shading (SFS)

Shape-form-shading recovers the shape from the gradual variation of shading of one single image. However, it has several limitations. For example, it is sensitive to the noise of intensity, and the light condition is limited to simple lighting conditions. SFS techniques only work for single material object by its principle. Most important of all, SFS techniques can only recover continuous surface, so it cannot deal with folding. Even though, it's single-view requirement is a benefit for image-based modeling. We need only one single shot and without the correspondences matching compared to multiple views technique.

Due to its intrinsic ill-pose problem, [Zeng et al. 2005.] proposed a user-assistant solution of continuous surface. Users input surface normal on specific feature points and the system refined the surface variations to the whole face. This method applied a

Fast Marching Method speed up the computation. After optimizing the energy function combining with each local surface, it can evaluate a global solution toward synthetic and real-world data.

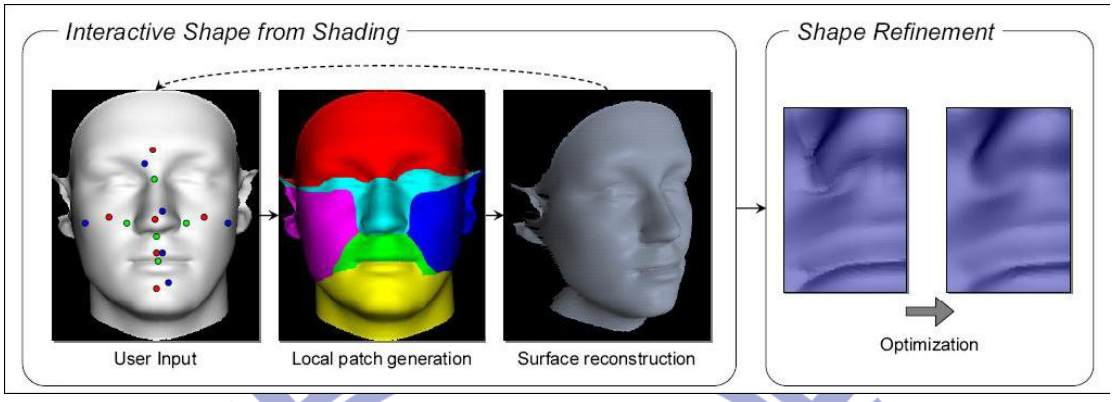


Fig6.Interactive Shape-from-shading

[Tai-Pang et al. 2008.] made a extension of the above one. Toward the biases of the light direction, they reformulated SFS and produced good initial normals for a large region to leave most noticeable errors mainly in the smooth part. They also developed an easily used 2D user-interface to edit and correct the normal map.

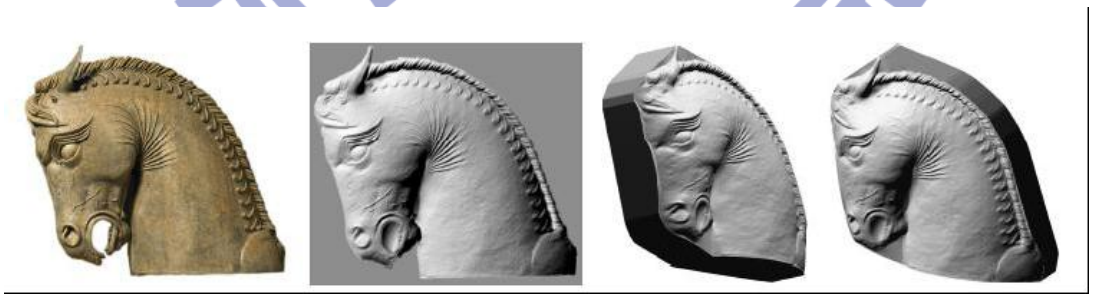


Fig7. Tai-Pang et al’s reconstructed surface

2.6 Others

[Fang et al. 2006.] combined the Shape-from-shading and texture synthesis to re-texture the target object in the photograph and video. They used optical flow keeping the texture coordinate in each frame. However, this approach is error-prone due to the Lambertian surface assumption and simple lighting conditions. It was only suitable for some simple objects, like t-shirt or sculptures, and needed manual rectification. Furthermore, it can only recover normal vectors.

Fig8. Fang et al.'s method pastes an image on a surface in video.

[Lin et al. 2004] analyzed the texture's type of geometry. They viewed near-regular textures as statistical departures from a regular texture along different dimensions. So they used shape-from-texture to recover geometry. This method worked mainly on surface with regular/near-regular texture and it also only recovered normal vectors. The two methods don't really recover surface geometry, but they synthesized realistic results by texture synthesis. It motivates us that we may not need to recover the exact depth map, but related depths for view changes.

Fig9.Lin et al perform texture replacement on an outdoor photo.

[Mathieu et al. 2009.] provided another optimization method to recover 3D mesh with inequality constraints. It also combined with Principle Component Analysis (PCA) so that it can folds and individual images. Nevertheless, this method needed an initial mesh on the surface, and cannot deal with self-occlusion.

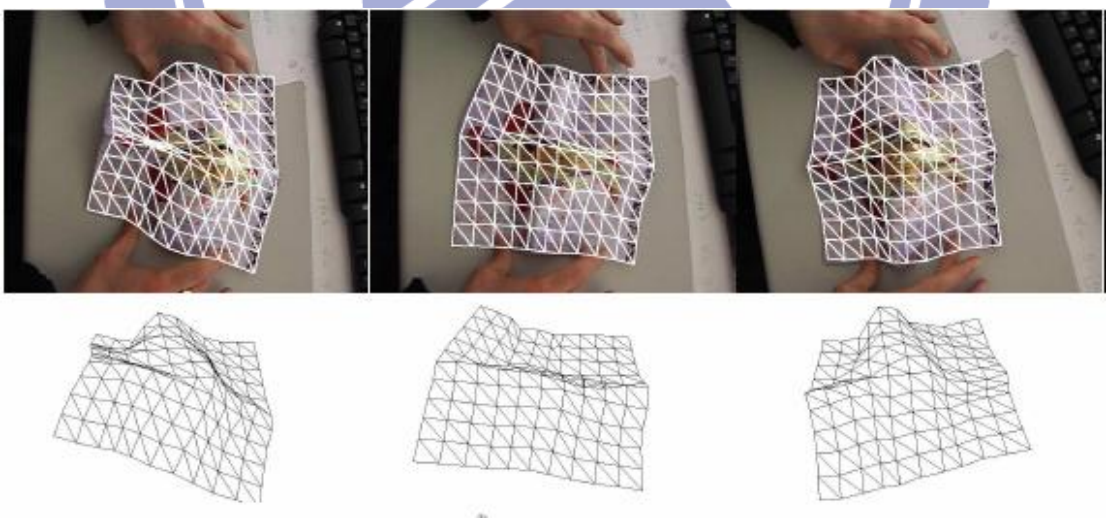


Fig10. Mathieu et al.'s method recovers 3D mesh in video.

3. Method

The input of our system is a single-view video sequence. Here we divide our method into four stages: **Preprocessing**, **Initial surface by Shape-from-shading** , **Control Point Estimation** and **PCA-based space-time optimization**. Details of these stages are then introduced in the following sections.

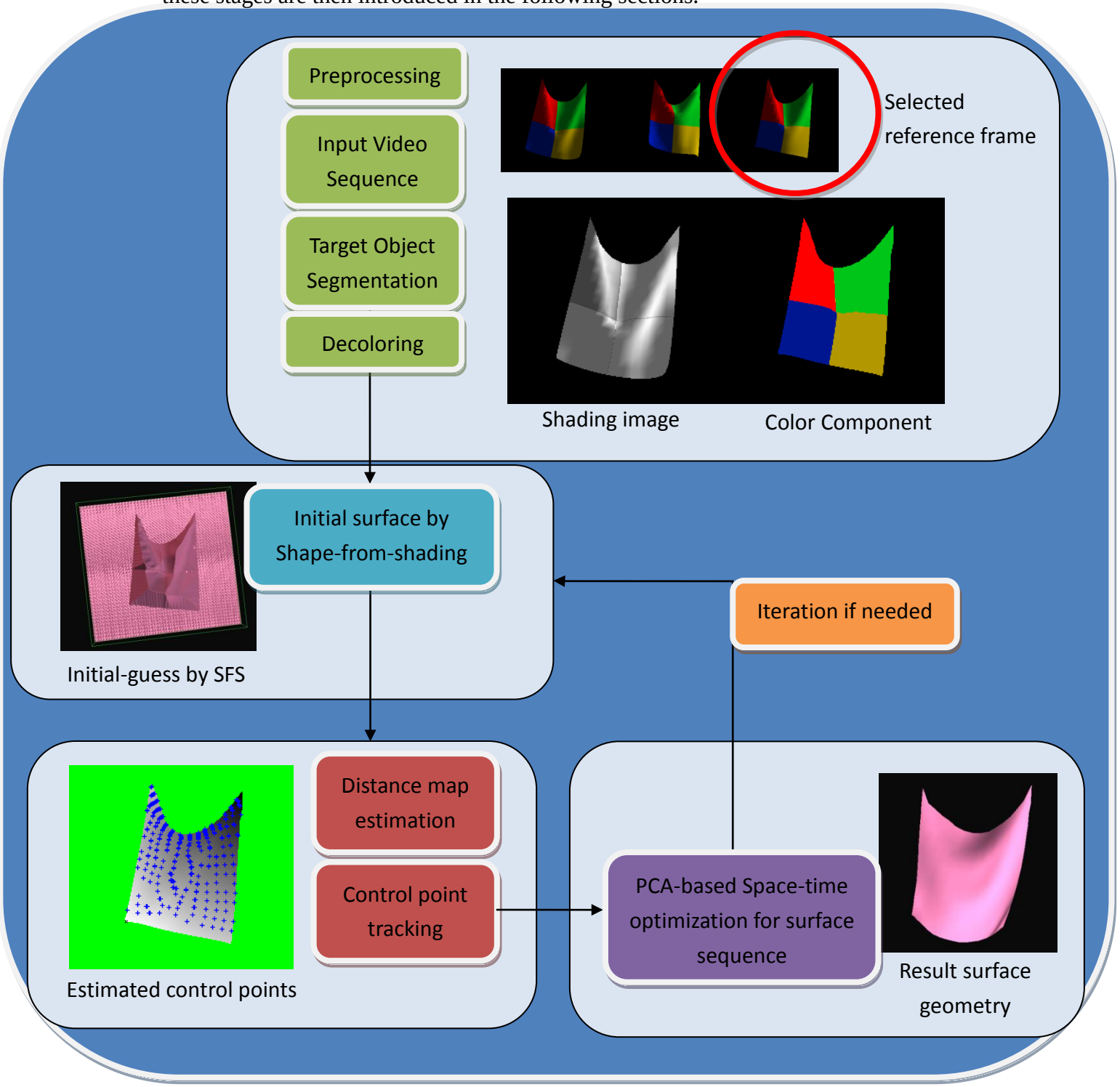


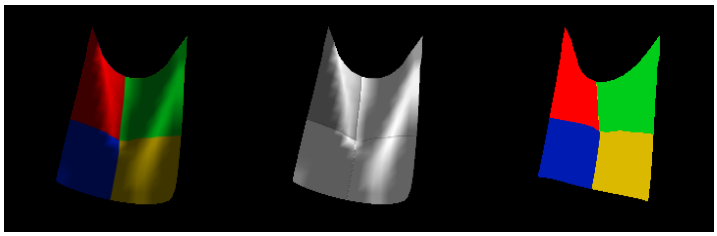
Fig11.Demostration of our system.

3.1 Preprocessing:

Given a single-view input video, our first step is to perform image/video segmentation for tar region extraction. Here, we modify [Jue Wang 2005.]'s video cut method.

For a common cloth image, its reflected intensities result from both shading from surface gradients and material reflection. Before we estimate its surface undulation, we should remove the material's effect and reduce our problem to Shape-from-shading of a single-material surface..

We combine [Marshall F. Tappen 2005.]'s methods to remove the material-reflectance component. First, we create a color histogram to store colors of extracted regions of the whole video sequence. For each pixel in one particular image, we calculate the color vector C(x,y) = $<\mathcal{R},\mathcal{G},\mathcal{B}>$ and find the same direction color vector in the histogram which has maximum intensity. Because the histogram is created from all frames, for a highly deformed surface, it's highly possible that there exists one pixel of an individual material has faced the light direction. Based on the Lambertian reflection, we assume the maximum intensity of a color vector as the individual material color. After all, we can calculate normal of each pixel (x,y) as $N(x,y) = \frac{C(x,y)-I_{Min}}{I_{Max}-I_{Min}}$.



(a) (b) (c)

Fig12.The decoloring processing. (a) Source image. (b)Recovered shading (c) Color component.

3.2 Initial surface by Shape-from-shading:

3D mesh is usually a good parametric representation of geometry reconstruction. Nevertheless, for estimating highly deformed surface from a single view, exploring the whole degrees of freedoms of meshes is too expensive in computation and is also easily trapped into local minima. Instead, we use a lower dimensional parametric space, called morphable model space(or PCA space) for more stable shape recovery.

We assume that our target surface has following properties:

(1) Foldable but with little self-occlusion.

(2) The maximum of surface undulation is less than $\frac{1}{4}$ of an edge.

(3) The four boundary vertices are nearly laid on one plane.

Surface with these properties can provide more reasonable Shape-from-Shading result. Furthermore, to track the motion markless cloth, we select the surface with the biggest area as our "reference frame" from the video sequence.

The following step is to generate an initial-guess depth map. Our input is only one single image and has no other viewpoints. Shape-from-Shading (SFS) is the few solution that can deal with such limited information. Here, we adapt Pentland's linear Shape-from-shading method.

Shape-from-Shading recovers depth from normal vector. Due to noise and insufficient scene information, it cannot tell us the surface's really height. In other word, Shape-from-Shading only recovers "relative" height.

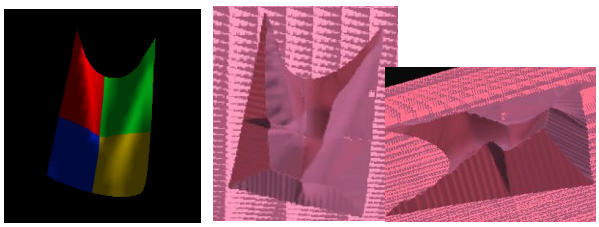
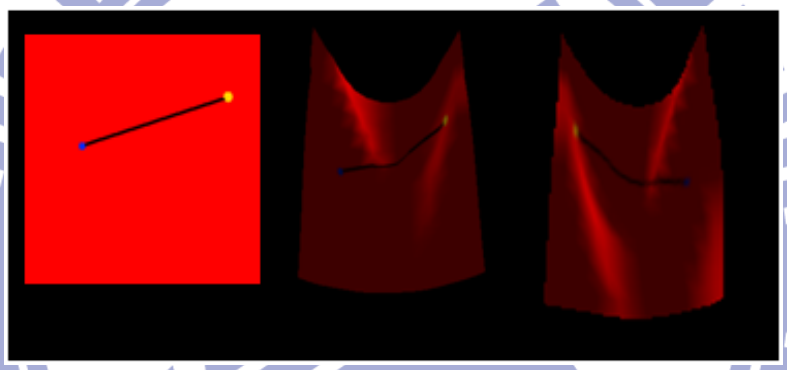


Fig13.SFS result in different viewpoint.

3.3 Control Point Estimation

To deal with unstable and highly deformed clothes or flags, we propose using morphable model for shape recovery. Nevertheless, mapping between morphable grids to input image is not straightforward. Here we consider the geodesic distances on the surface we recovered in SFS stage. Given point X and Y on the surface, if the surface is flat, the straight line across X and Y have minima distance. When the surface is undulating, the "line" with minimum geodesic distance may not look "straight" at the camera viewpoint as shown in Fig.10. No matter the line looks like, the geodesic distance is the same location.



(a) (b) (c)

Fig14. The line across X and Y. (a) Source texture with straight line. (b) (c) The same line on not-flat geometry.

Without loss of generality, we consider the distance from the upper two vertices to all other points over the surface. Our generic mesh can be seen as lattices with the same size. Assume the width and height of the square flags is L and break the flag into n*n grid. Every grid' size is $\frac{L}{n} * \frac{L}{n}$. We name every vertex $V(u, v)$ of mesh as Fig.15.

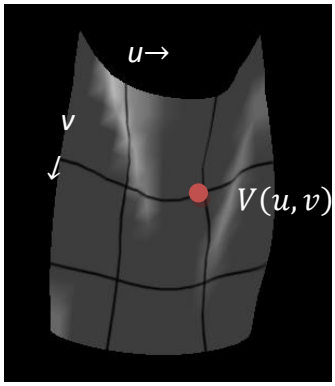


Fig15.An example of 3*3 mesh.

First we consider the *v-axis* .To the point over the *v=k* axis, we can write:

$$\sqrt{dist(V_{0,0},V_{u,k})^2-h^2}+\sqrt{dist(V_{n,0},V_{u,k})^2-h^2}=L$$

$$where \quad h=L*\frac{k}{n} \quad .$$

Then about the *u-axis*. To all points over the *i=k* axis we can write:

$$dist(V_{0,0},V_{k,v})^2-D^2=dist(V_{n,0},V_{k,v})^2-(L-D)^2$$

$$where \quad D=L*\frac{k}{n} \quad .$$

$$\begin{aligned} dist(V_{0,0},V_{k,v})^2-dist(V_{n,0},V_{k,v})^2 &= D^2-(L-D)^2 \\ &= (D+(L+D))(D-(L-D)) \\ &= L(2D-L) \end{aligned}$$

With the above two equation, we can locate control points over the 3D geometry of the surface.

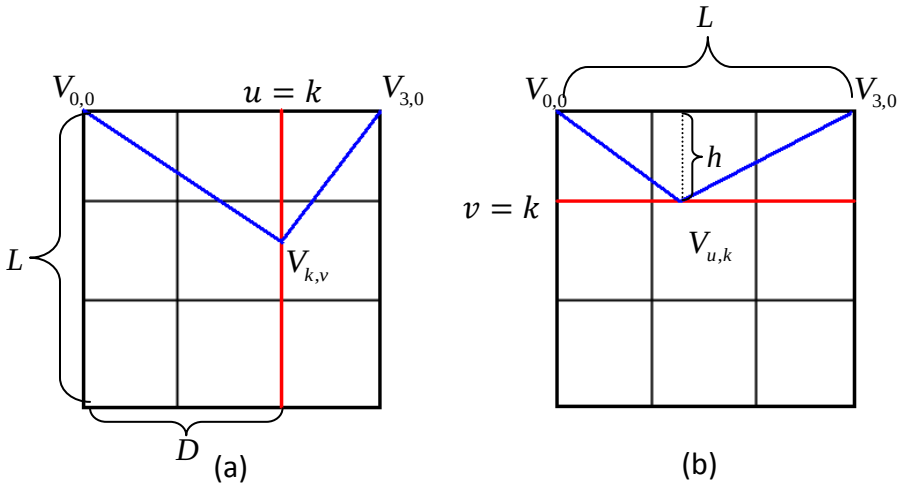


Fig16.An example of distance between $V_{u,v}$ and $V_{0,0}, V_{3,0}$.

(a) $dist(V_{0,0},V_{k,v})^2-dist(V_{n,0},V_{k,v})^2=D^2-(L-D)^2$

(b) $\sqrt{dist(V_{0,0},V_{u,k})^2-h^2}+\sqrt{dist(V_{n,0},V_{u,k})^2-h^2}=L$

After we define the above two equations, the next step is how to calculate the distance between $V_{u,v}$ and $V_{0,0}$ over the surface. Of course we cannot directly calculate the L2 norm between the two vertexes. The solution is to use Dijkstra algorithm.

Dijkstra algorithm bases on the following concept: If there exists a path P from $V_{0,0}$ to $V_{x,y}$ which has minima geodesic length, the path P' from $V_{0,0}$ to some point just in front of $V_{x,y}$ is also a minima path. In our experiment, the basic dynamic programming function is:

$$dist(V_{0,0},V_{x,y}) = \min \begin{cases} dist(V_{0,0},V_{x-1,y}) + norm(V_{x-1,y},V_{x,y}) \\ dist(V_{0,0},V_{x,y-1}) + norm(V_{x,y-1},V_{x,y}) \\ dist(V_{0,0},V_{x-1,y-1}) + norm(V_{x-1,y-1},V_{x,y}) \end{cases}$$

We calculate the distance function all over the surface, therefore, we get a table saved all vertex's minima distance to $V_{0,0}$.

Generally, the above equation cannot work well because of our decoloring method. The decoloring method is not perfect so there exists gaps between color regions. Surface's Normals near the gap sharp thus effect shape-from-shading's result and so that the $dist(V_{0,0},V_{i,j})$ will become larger if the path P cross the gap. So we adjust the dist function:

$$dist(V_{0,0},V_{x,y}) = \min \begin{cases} dist(V_{0,0},V_{x-1,y}) + \min(T_d, norm(V_{x-1,y},V_{x,y})) \\ dist(V_{0,0},V_{x,y-1}) + \min(T_d, norm(V_{x,y-1},V_{x,y})) \\ dist(V_{0,0},V_{x-1,y-1}) + \min(T_d, norm(V_{x-1,y-1},V_{x,y})) \end{cases}$$

Here T_d acts like an upper bound of distance of any two close vertices. So the color's effect is diminished. Not only we calculate all the distance from $V_{0,0}$ to all

other vertex, we also take $V_{n,0}$ into account. After the two dynamic programming processes, we get two distance map from the two upper vertexes.

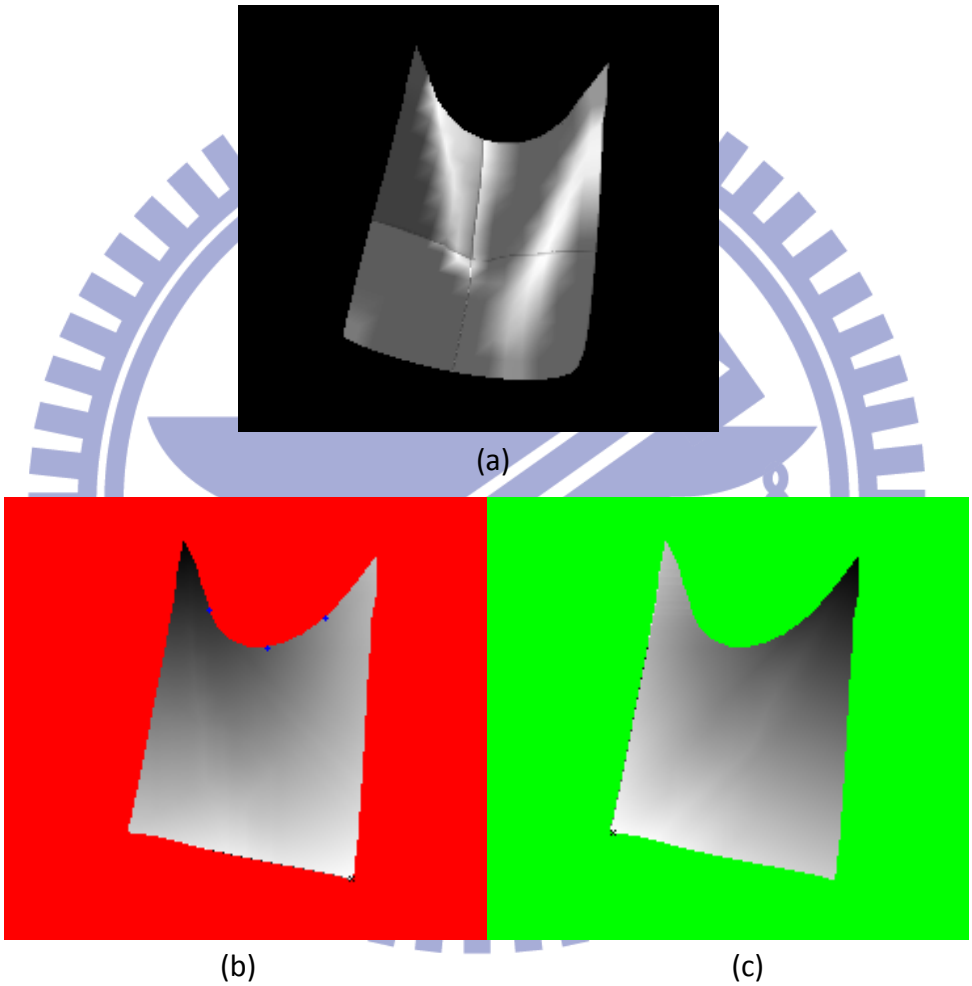


Fig17. The distance map.(a) The recovered shading image. (b) From the left-up vertex (c) From the right-up vertex. The Darker is closer.

Finally we check all the pixels to find someone located at right distance from two upper vertices:

$V_{0,0} = P_{0,0}$, $V_{n,0} = P_{w,0}$ //Assign the two upper point's position

For u = 1 to W

For v = 1 to H

For $1 \le k_r, k_c \le$ n

If($\| \sqrt{dist(V_{0,0}, u, v_r)^2 - (L\frac{k_r}{n})^2} + \sqrt{dist(V_{n,0}, P_{u,v_r})^2 - (L\frac{k_r}{n})^2} - L \| < T_{row}$)

And($\| dist(V_{0,0}, P_{u,v})^2 - dist(V_{n,0}, P_{u,v})^2 - L(2L\frac{k_c}{n} - L) \| < T_{col}$)

$V_{k_c,k_r} = P_{u,v_r}$

End k_r, k_c

End j

End i

Fig18. The control points algorithm. T_{col} and T_{row} are thresholds.

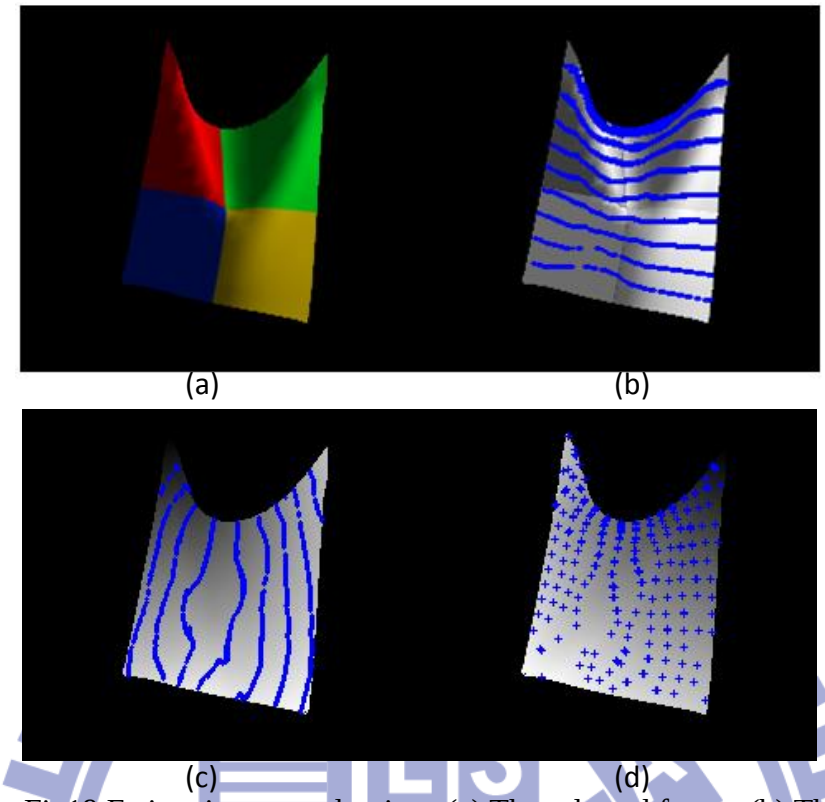


Fig19.Estimating control points. (a) The selected frame. (b) The columns when n=10. (c) The columns when n=10. Notice that the two row/column of k=0 and 10 is not appear. (d) The control point we really recovered. The mesh is 16*15.

The above mentioned algorithm has a problem: because of error propogation, the estimated lengh and position are not precise in lower region (far from the origin vertices). From Fig19, we can see that control points are compact in upper region, but are distorted in lower part.

3.4 Space Optimization of Mesh Recovering

It is difficult and imprecise to treat the above result as the really mesh point's position.The z-axis of each points is from shape-from-shading result, which is just a "related" depth. Worst of all, the control points miss much and we easily see that those points didn't locate at what they should be as i and j increased.

Because of the above reason, we need a stable method to recover the geometry and take its characteristic into account. Meanwhile, the method should restrict the improved control points' location near our estimation. One appropriate technique is

using Principle Component Analysis(PCA). First we generate training data by cloth simulation, then find out their statistics characteristics, called principle componenes. Then we use the PCs to recover the mesh by subspace optimization.

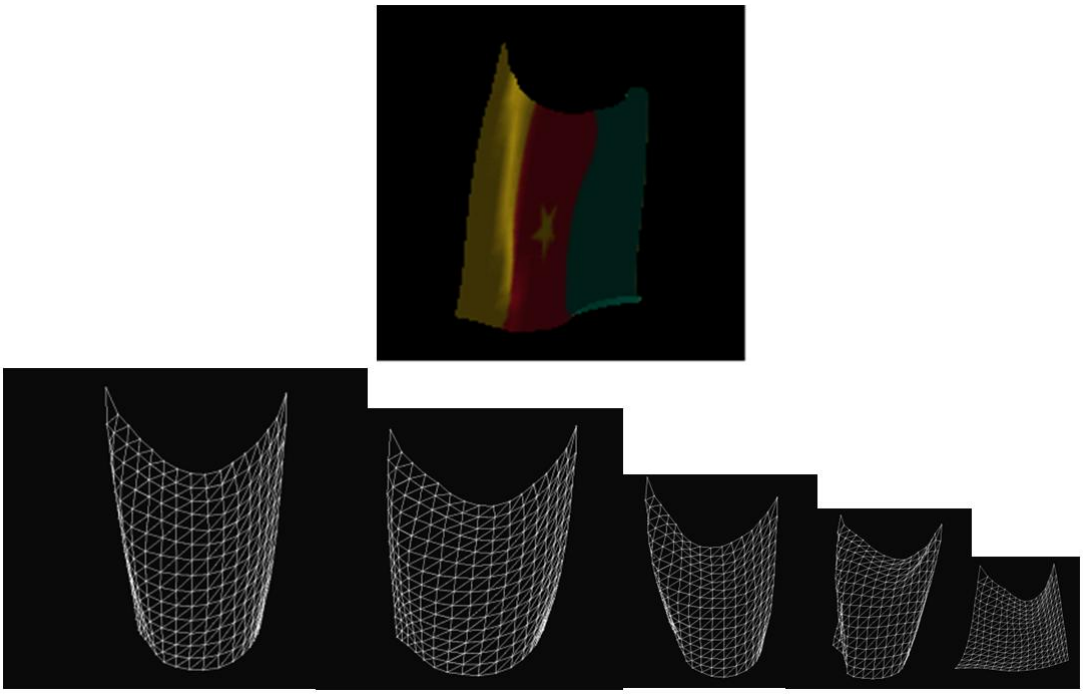


Fig20. The PCs of training data.

The relation among input estimation control points (X , Y , Z) generated at the 3.3, PCs and PC projected coefficient w_k is:

$$STR_\rho R_\theta \begin{bmatrix} & & & & \\ PC_1 & PC_2 & PC_3 \cdots & PC_k & \\ & & & & \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_k \end{bmatrix} = \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ \vdots \\ Z_n \end{bmatrix}$$

where θ and ρ are two rotate angles, T is tranition and S is scalar matrix.

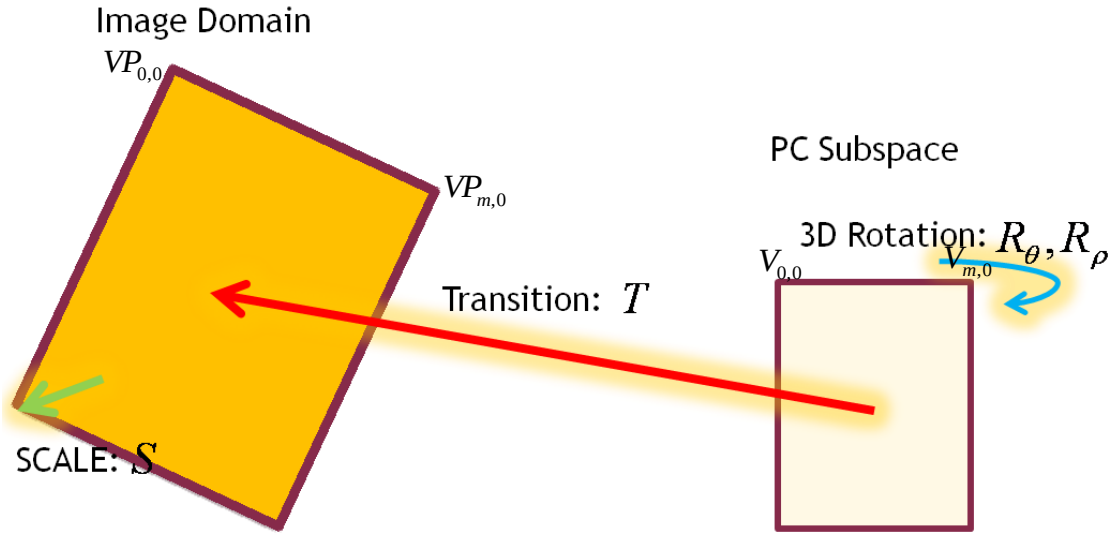


Fig21. The relation between PC subspace and input image domain.

To reduce the computing complexity, we change the matrix order as follows:

$$R_\rho R_\theta \begin{bmatrix} PC_1 & PC_2 & PC_3 \cdots & PC_k \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_k \end{bmatrix} = T^{-1} S^{-1} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ \vdots \\ Z_n \end{bmatrix}$$

where $T^{-1}S^{-1}$ are inverse of T and S.

Because the two upper vertices should be translated to the same position, we can easily define the scalar matrix S' by scale the distance between $V_{0,0}$ and $V_{i,j}$ to $VP_{0,0}$ and $VP_{i,j}$. At the mean while, the translation vector T' can be defined as the vector $\overrightarrow{VP_{0,0} - V_{0,0}}$. Finally, we consider the following equation:

For each $V_{i,j} = (X'_{i,j}, Y'_{i,j}, Z'_{i,j})$,

$$\Re \begin{bmatrix} PC_1 & PC_2 & PC_3 \cdots & PC_k \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix} = \begin{bmatrix} X'_{i,j} \\ Y'_{i,j} \\ Z'_{i,j} \end{bmatrix}$$

where $\Re = \begin{bmatrix} r_a & r_b & r_c \\ r_d & r_e & r_f \\ r_g & r_h & r_i \end{bmatrix}$ is $R_\rho * R_\theta$.

We therefore recover the shape by solving the optimization problem:

$$\min_{\Re, w_p, p=1 to k} \sum_{u,v} \left\{ \Re \begin{bmatrix} PC_1 & PC_2 & PC_3 \cdots & PC_k \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix} - \begin{bmatrix} X'_{u,v} \\ Y'_{u,v} \\ Z'_{u,v} \end{bmatrix} \right\}$$

As we mentioned before, our estimation control points' correctness is inversely proportional to the distance from upper vertex. So we modified the optimization equation with a weight $w_{ctrl} = \frac{K_c}{(u - \frac{m}{2})^2 + v^2}$ .As uv go near the lower corner, w_{ctrl} decreases. Here we set K_c = 1.0.

$$\min_{\Re, w_p, p=1 to k} \sum_{u,v} \left\{ \frac{K_c}{(u - m/2)^2 + v^2} \Re \begin{bmatrix} PC_1 & PC_2 & PC_3 \cdots & PC_k \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix} - \begin{bmatrix} X'_{u,v} \\ Y'_{u,v} \\ Z'_{u,v} \end{bmatrix} \right\}$$

Besides, the grid's size will keeps no matter how the flag waves. So to each $V_{i,j}$,the distance of $\overrightarrow{V_{i,j} - V_{i+1,j}}$ and $\overrightarrow{V_{i,j} - V_{i,j+1}}$ should be limited. The problem is: as [Mathieu Salzmann 2009] mentioned, when deformed, while the geodesic distance between the two points is preserved, the projected one decreases. Here we don't directly deal with this problem, but we just limit the distance in a reasonable range. While we implement it as penalty according to length variation, it can be regarded as a spring constraint between vertices. The following energy function will be added in our minimize function:

$$E_d = \sum_{\forall V_{u,v} \in mesh} \{K_d(\|V_{u,v} - V_{u,v+1}\| - L) + K_d(\|V_{u,v} - V_{u+1,v}\| - L)\}$$

where L is the grid's length and E_c is our previous optimization equation.

We call the rear part E_d as distance constraint.

Last but not least, we have to add one more constraint in the optimization function. Because of the w_{ctrl}, the farther control points may move to unpredicted location as follows:

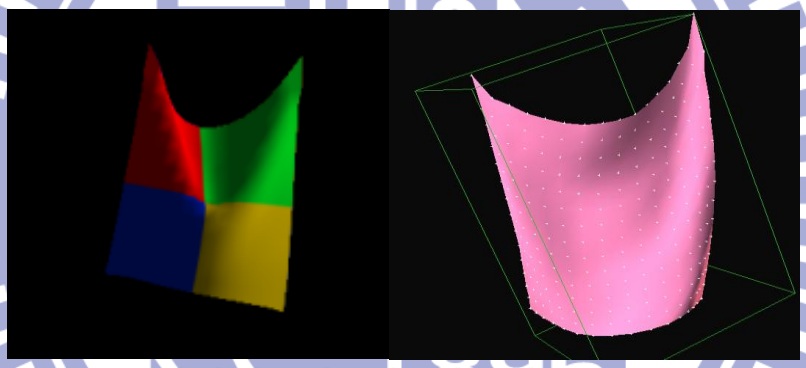


Fig22. The bad result without boundary constraint.

To get a better result, we give the boundary point a bigger weight. Althrough our minima distance method may not works well over the boundary, the boundary vertex's location is easy to directly estimate from our SFS result. We define the boundary cost function E_b :

$$E_b = \sum_{i,j \in \Omega} \left\| K_b \Re \begin{bmatrix} PC_1 & PC_2 & PC_3 \cdots & PC_k \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix} - \begin{bmatrix} X'_{i,j} \\ Y'_{i,j} \\ Z'_{i,j} \end{bmatrix} \right\|$$

where $K_b > K_c$ constraints the boundary's 2D position.

The final optimization function we used is:

$$\min_{\Re, w_p, p=1 \text{to} k} \sum_{u,v} \left\{ \frac{K_c}{(u - m/2)^2 + v^2} \Re \left[PC_1 \quad PC_2 \quad PC_3 \cdots \quad PC_k \right] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix} - \begin{bmatrix} X'_{u,v} \\ Y'_{u,v} \\ Z'_{u,v} \end{bmatrix} \right\} + E_d + E_b$$

$$\text{where} \quad E_b = \sum_{P_{u,v} \in \Omega} \left\| K_b \Re \left[PC_1 \quad PC_2 \quad PC_3 \cdots \quad PC_k \right] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix} - \begin{bmatrix} X'_{u,v} \\ Y'_{u,v} \\ Z'_{u,v} \end{bmatrix} \right\|$$

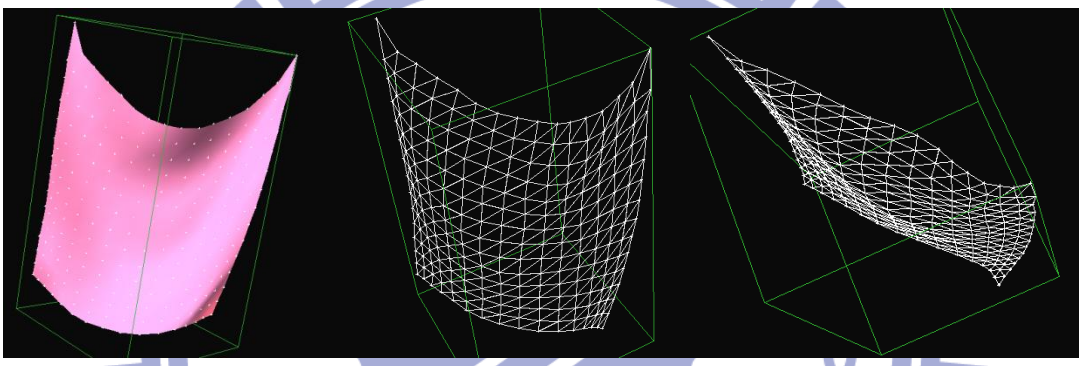


Fig23. The bettwe result and its wireframes.

3.5 Time Optimization

Finally we have to map the mesh on all the frames in the video. Here we use the method in [Mathieu Salzmann,2009]. This paper introduce a optimal solution with a given 3D mesh and mapping the mesh to all frames. The optimal equation is:

$$\max_{X^{T-1}, X^T, X^{T+1}} \sum_{\varsigma=-1}^{1} E_t(X^{T+\varsigma}) - w_m E_m(X^{T-1}, X^T, X^{T+1})$$

Where E_t is it's objective function for a single frame,

and $E_m(X^{T-1}, X^T, X^{T+1}) = \| X^{T-1} - 2X^T + X^{T+1} \|$.

We replace the E_t component with our objective function and apply to all frames. Thus we get a continuous video sequence.

4. Experimental Results

Our experiment has three steps: First we apply our method to images captured from synthetic video sequence. Then we add in time constraint and recover the whole surface sequence. In the end, we try to use our method in real world data, shows the limitation of our presented method. Furthermore, we also include heavy noise in synthetic image to verify the stability of the proposed method.

We run our experiment in a Core 2 Duo PC with 3G RAM. Our input image's size is less than 300*300 pixels. The whole process of each frame is finished in less than 5 seconds.

4.1 Synthetic Data(single)

We generate some frames from our cloth simulation program, and then apply our method on each one without time constraints in Fig23&24 below. Clearly see that our method can keep the surface's geometry, and the recovered surface is smooth, which means it is not affected by material color and noise.

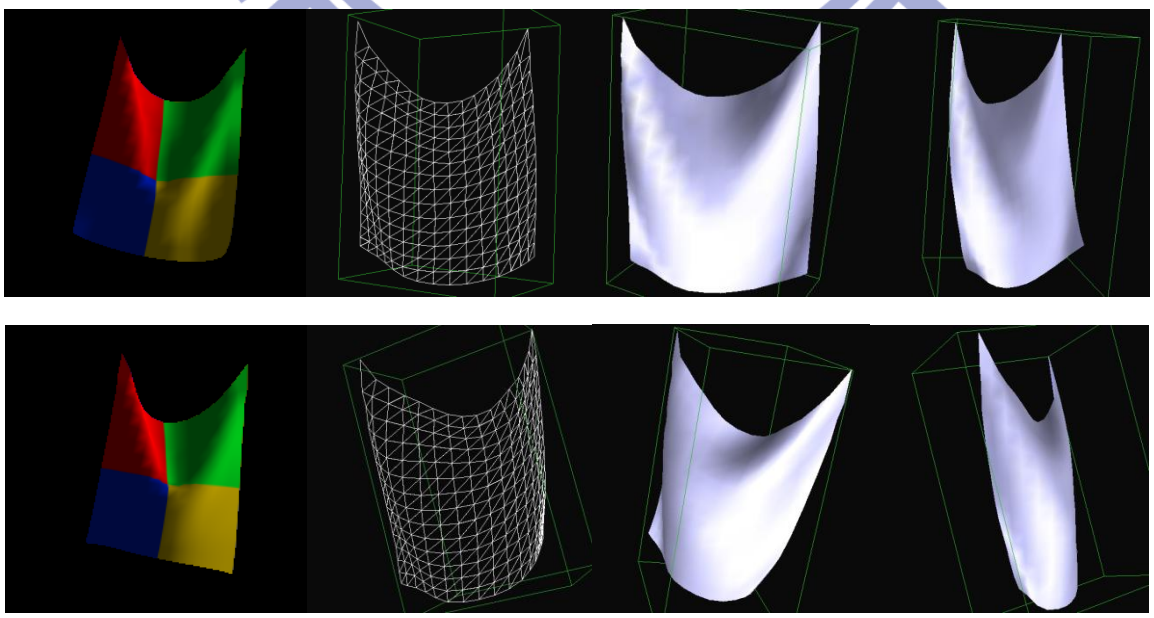


Fig24.Two synthetic frames from our cloth simulation and the reconstruction geometries. As we mentioned, Our method perform well in upper region but in lower region, the boundary is still inconsistent.

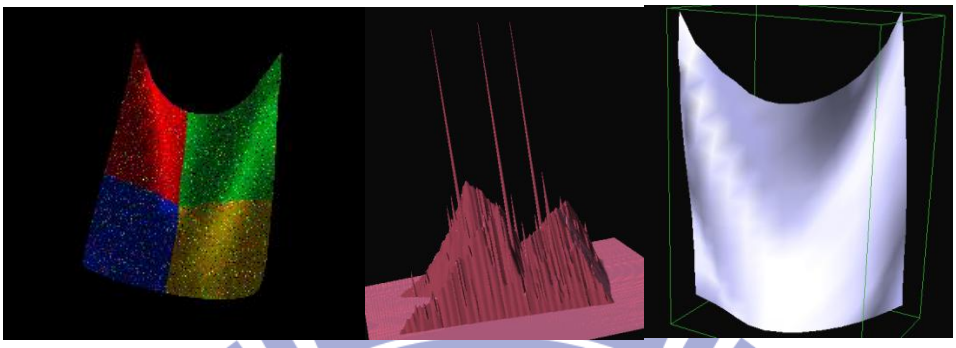


Fig25.Synthetic data with noise. Easily see the recovered surface is noise-proof.

4.2 Synthetic data(sequential)

Here we generate a sequence or frames and apply our method with time constraints. The motion is smooth but the boundary errors become unpredictable. That's because we take the frame with biggest area as reference frame, and then all other boundary points will be contraction because of interpolation.

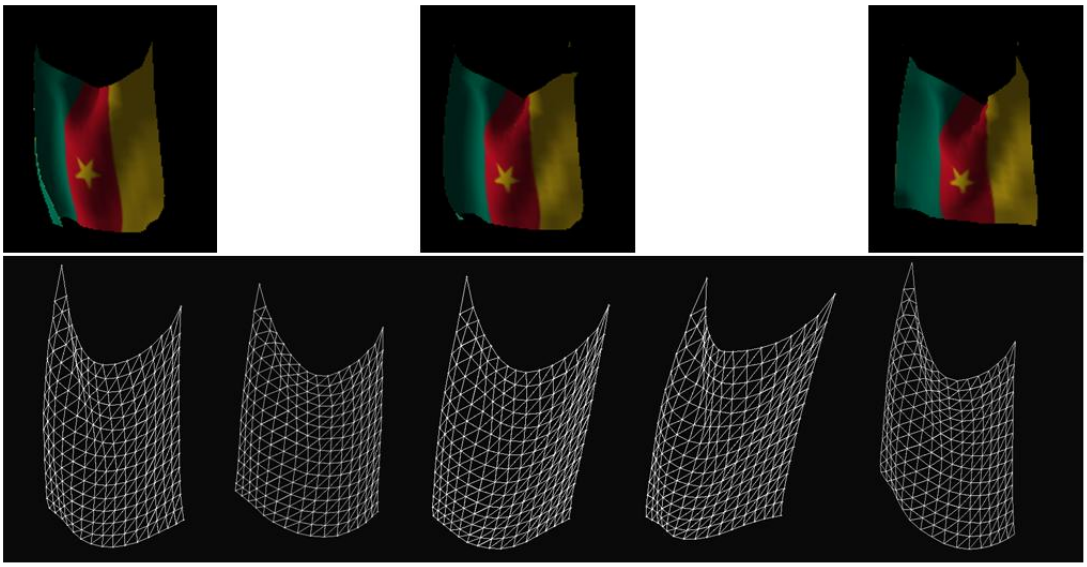


Fig26.Synthetic frames sequence. First row is synthetic frames at t=1, t=5 and t=9, and the second row is the recovered geometry. The boundary is not consistent because of the interpolation in optimization processing.

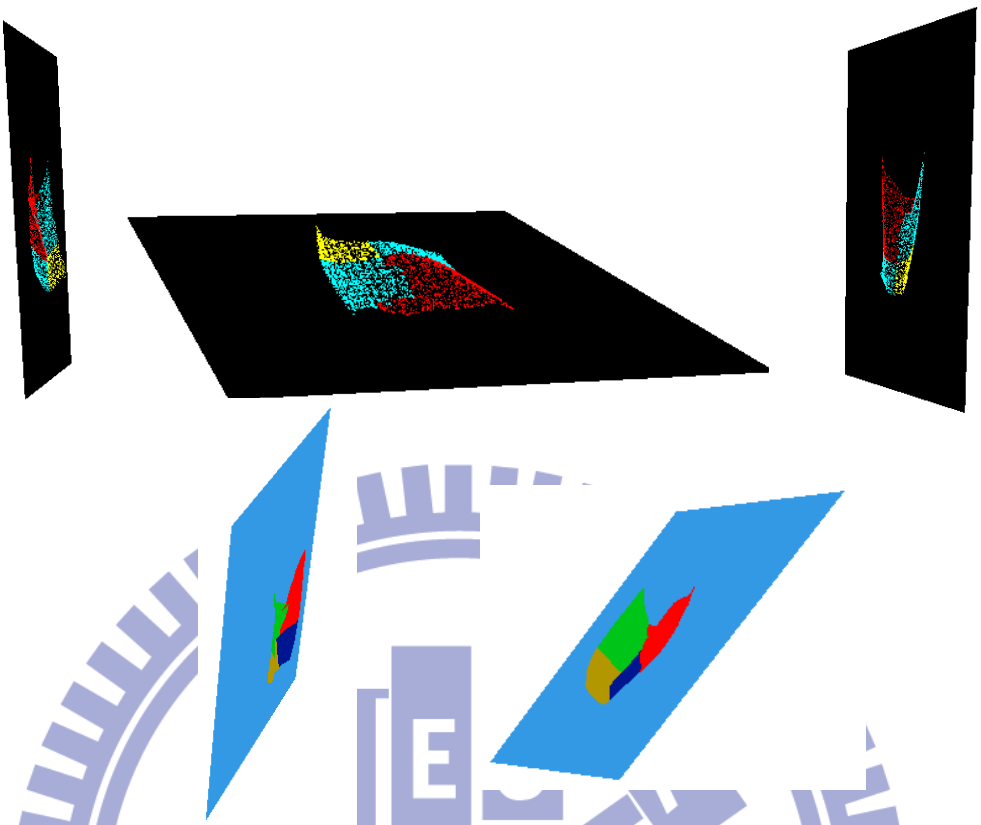


Fig27. Recovered surface displays in 3DTV.

4.3 Real data surface

While applying our method in real data, we found that our mesh is not delicate enough to capture the details in real data surface. Our method designed to be not affected by small variation like noise, but at the meanwhile, it cannot capture the undulation in a small grid, which will be captured by shape-from-shading. That is the problem we are going to handle.

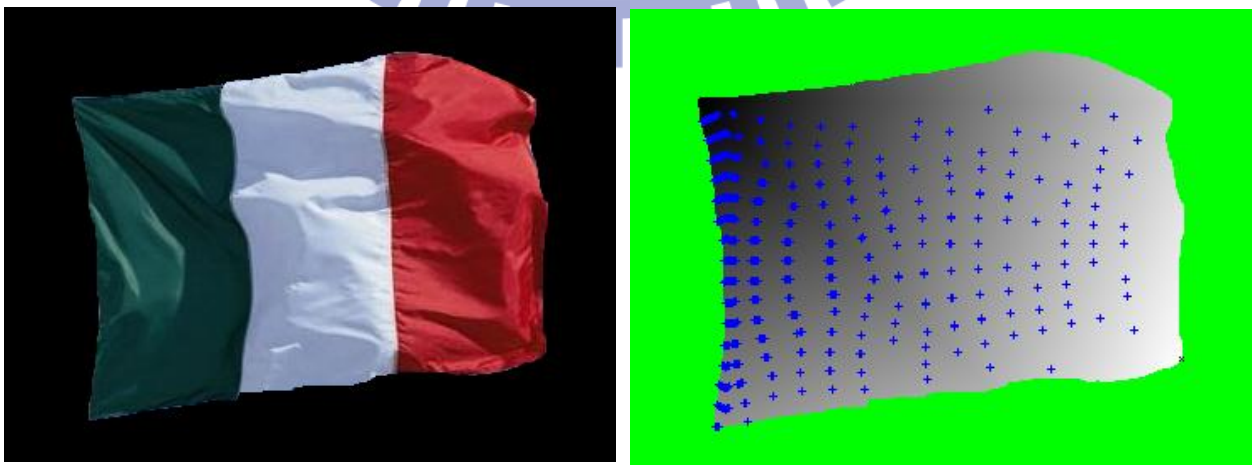


Fig28.Our test real data and it's estimated control points position.

5. Conclusion

In this paper, we presented a novel approach to recover a surface's geometry with PC-bases. Because shape-from-shading result is not precise and most reconstruction method cannot deal with self-occlusion, we define a lower dimensional PCA space for more stable shape recovery. Our method can automatically recover surface geometry, without any user interaction. This method is also less effected by material, and keeps the surface's physical characteristics. The result can be used in 3DTV and apply to other applications.

In future work, we will seek to recover the small variations over surface. More specifically, we try to combine SFS results with our mesh. This will helps our results more realistic.

Basically, one purpose of our method is to recover self-occlusion surface. Simple SFS cannot recover the covered parts, so we introduce PC-bases subspace. Currently, our method can only recover small self-occlusion surface by time coherence. In the near future, we plan to include color constraint and short-term structure from motion for more accurate tracking.

6. Reference

Aaron Hertzmann and Steven M. Seitz, “Example-Based Photometric Stereo ： Shape Reconstruction with General, Varying BRDFs”. In IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)’05 vol.27 no. 8 pp. 1254-1264.

Anton van den Hengel, Anthony Dick, Thorsten Thorm¨ahlen, Ben Ward, Philip H.S. Torr,” VideoTrace: Rapid interactive scene modeling from video”. In ACM SIGGRAPH’07.

Carlos Hernandez Esteban, George Vogiatzis, Roberto Cipolla, ”Multi-view photometric stereo”. In IEEE Trans. on Pattern Analysis and Machine Intelligence(PAMI)’08 vol.30 no. 3 pp. 548-554.

G. Vogiatzis, P.H.S. Torr, R. Cipolla, ”Multi-view Stereo via Volumetric Graph-cuts”. In Computer Vision and Pattern Recognition (CVPR)’ 05 vol.2 pp. 391- 398.

Gang Zeng, Yasuyuki Matsushita, Long Quan, Heung-Yeung Shum, ” Interactive Shape from shading”. In Computer Vision and Pattern Recognition (CVPR)’ 05 vol.1 pp. 343- 350.

Hui Fang, John C. Hart,”Roto Texture: Automated Tools for Texturing Raw Video”. In IEEE Trans. on Visualization and Computer Graphics (TVCG)’06 vol.12 pp. 1580-1589.

Hui Fang, John C. Hart," Textureshop: Texture Synthesis as a Photograph Editing Tool". In ACM SIGGRAPH '04

Jue Wang, Pravin Bhat, R. Alex Colburn, Maneesh Agrawala, Michael F. Cohen."Interactive video cutout". In ACM SIGGRAPH '05.

L.Zhang, N. Snavely, B. Curless, S.M. Seitz, "Spacetime Faces: High Resolution Capture for Modeling and Animation". In ACM SIGGRAPH'04.

Lorenzo Torresani, Aaron Hertzmann,Chris Bregler,"Learning Non-Rigid 3D Shape from 2D Motion". In Neural Information Processing Systems (NIPS)' 03

M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, R. Koch," Visual modeling with a hand-held camera". In International Journal of Computer Vision (IJCV)'04 vol. 59 no. 3 pp. 207-232.

Marshall F Tappen, William T Freeman, Edward H Adelson, "Recovering Intrinsic Images from a Single Image". In IEEE Trans. on Pattern Analysis and Machine Intelligence(PAMI)'05 vol. 27 no. 9 pp. 1459-1472.

Mathieu Salzmann, Pascal Fua, "Reconstructing Sharply Folding Surfaces: A Convex Formulation", In Computer Vision and Pattern Recognition (CVPR)' 09.

Maxime Lhuillier, Long Quan, "A quasi-dense approach to surface reconstruction from uncalibrated images", In IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)'05 vol.27 no. 3 pp. 418-433.

Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. “Modeling and Rending Architecture from Photographs: A Hybrid Geometry- and Image-Based Approach”. In ACM SIGGRAPH ‘96.

Ping Tan, Stephen Lin, Long Quan, “Separation of Highlight Reflections on Textured Surfaces”. In Computer Vision and Pattern Recognition (CVPR) ’06 vol.2 pp. 1855-1860.

Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer and Mubarak Shah, “Shape from shading: A Survey”. In IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)’99 vol.21 no.8 pp. 690-705.

Szymon Rusinkiewicz Olaf Hall-Holt Marc Levoy, ”Real-Time 3D Model Acquisition”. In ACM Transaction On Graphics(TOG)’02 vol.21 issue.3 pp.438-446. .

Tai-Pang Wu, Jian Sun, Chi-Keung Tang, Heung-Yeung Shum,”Interactive Normal Reconstruction from a Single Image”. In ACM SIGGRAPH’08.

Yanxi Liu, Wen-Chieh Lin, James Hays, “Near-Regular Texture Analysis and Manipulation”. In ACM SIGGRAPH ’04.

Zheng Qin Michael D. McCool Craig S. Kaplan, “Real-Time Texture-Mapped Vector Glyphs”. In Interactive 3D graphics and games (I3D)’06 pp.125-132.