# Reconstructing a Dynamic Surface from Video Sequences Using Graph Cuts in 4D Space-Time*

Tianli Yu, Ning Xu and Narendra Ahuja
*University of Illinois at Urbana-Champaign*
*{tianli, ningxu, ahuja}@vision.ai.uiuc.edu*

## Abstract

*This paper is concerned with the problem of dynamically reconstructing the 3D surface of an object undergoing non-rigid motion. The problem is cast as reconstructing a continuous optimal 3D hyper-surface in 4D space-time from a set of calibrated video sequences. The imaging model of video cameras in 4D space-time is derived and a photo-inconsistency cost function is defined for a hyper-surface in the 4D space-time. We use a 4D node-cut algorithm to find a global minimum of the cost function and obtain the corresponding optimal hyper-surface. Experimental results show that the proposed algorithm is effective in recovering continuously changing shapes and exhibits good noise resistance.*

## 1. Introduction

Volumetric model representation has become a popular choice recently for 3D reconstruction from multiple view images [1]. This is because it greatly simplifies the correspondence problem which is otherwise quite complicated in the multiple view case. The volumetric representation also lends itself to novel optimization methods such as graph cut based methods [2, 3]. These methods can give a global optimal reconstruction in polynomial time. In this paper, we propose an extension of a graph cut based 3D reconstruction algorithm [3] to 4D space-time domain, i.e., to reconstruct a continuously changing 3D surface (not necessarily in rigid motion) from a set of video sequences captured by several fixed cameras. There are two major reasons that motivate this extension: 1. Applications such as facial motion analysis or medical imaging of live organs prefer that the 3D reconstruction be performed continuously over a period of time. 2. Reconstruction in 4D space-time can make use of the temporal coherence to help better regularize the solution and overcome noise.

* The support of National Science Foundation under grant ECS 02-25523 is gratefully acknowledged.

Most of the multiple view 3D reconstruction algorithms deal with static scenes. Directly applying these algorithms to each frame of the video sequences is not always appropriate since the temporal constraints are neglected, and consequently, the reconstructed shape may not be continuous in time. One remedy to this is to treat the space and time domain uniformly and define a 4D reconstruction problem. Previous work has shown that joint use of both space and time offers noticeable advantages. Through analogy to the 2D planar motion case, Hall-Holt and Rusinkiewicz [4] extend structured light scanning to 4D space-time, and design a space-time code based on the rules used in 3D scanning. The method can extract the structure of a moving object in real time. Matheny and Goldgof [5] extend the spherical harmonic representation of object shape to 4D space-time and use it to represent object shape undergoing non-rigid motion. They also show that 4D spherical harmonics provide an improved model for the motion of the left ventricle of the heart.

The 4D space-time reconstruction problem has received more attention in the field of medical image processing, where volume sequences (data already in 4D form) such as continuous-time MRI or SPECT are available. Deformable models [6, 7, 8] are often used to segment and track the non-rigid motion of the object of interest. In [7], a mesh model is constructed at each time step and the corresponding mesh points are connected to enforce the temporal constraints. One of the problems in these methods is that the results are usually obtained by local optimization methods which require a good initial estimate. Another issue is that the connected mesh points do not necessarily correspond to any fixed points on the object and the temporal constraints are only of mathematical meaning.

In this paper, we formulate the dynamic 3D surface reconstruction problem from calibrated video cameras as a hyper-surface reconstruction problem in the 4D space-time from video sequences (which can be viewed as 3D image data). This formulation naturally enforces the time continuity constraint. We use 4D node capacitated graph cut, which is an extension of the method used in [3] to find a globally optimal hyper-surface in 4D space-time that best explains the captured video sequences.

The paper is organized as follows: Section 2 presents our formulation that converts the dynamic 3D surface reconstruction problem into a hyper-surface reconstruction problem in 4D space-time. Section 3 describes the 4D node capacitated graph cut algorithm. Section 4 presents some experimental results. Section 5 gives the conclusion.

## 2. Problem Formulation

In this section, we will formulate the dynamic 3D reconstruction problem as that of finding a hyper-surface in 4D space-time that is consistent with all the input data. As a simpler case, we first consider a dynamic shape reconstruction problem in 2D space.

### 2.1. Dynamic 2D shape reconstruction

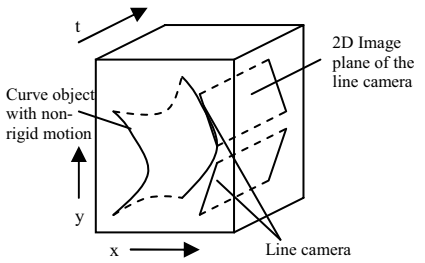



**Fig. 1 The imaging of a line object moving in 3D space-time domain**

Suppose a moving curve (line) object is confined to a 2D plane X-Y, as shown in Fig. 1. The object is imaged by fixed line cameras which are the 2D version of the usual pin-hole cameras in 3D space. The image plane of a line camera is a line in the plane, and every point in 2D plane can be mapped on to the line using projective transform. With several line cameras at different locations, we can reconstruct the 2D shape of the object using triangulation.

Now let us consider the problem in 3D space-time domain. If we stack together the line images captured by the line cameras at each time step, we can form a 2D video. The 2D video can be thought of as captured by a 2D image plane parallel to the time axis (Fig.1). The moving object also forms a continuous surface in the 3D space time. Instead of reconstructing the line shape in 2D at each frame, we can view the problem as reconstructing a continuous surface in 3D space-time using a set of 2D images.

### 2.2 Projective Transform in 4D space-time

Analogous to the 3D space-time, a moving surface of a 3D object forms a 3D hyper-surface embedded in 4D space-time domain, and should be continuous for a physically plausible object. The video cameras continuously capture the light rays reaching their image planes for a period of time. Stacking these images (frames) together yields a 3D volume that can be used for 4D reconstruction. We denote the 3D volume by $I(x',y',t')$, where $(x',y')$ are the coordinates on the image plane and $t'$ is the local time of camera. The projective relationship between a point $(x,y,z,t)$ in 4D and the pixel $(x',y',t')$ on the 3D image volume can be modeled as the extrinsic and intrinsic transform. The extrinsic transform is:

$$\begin{pmatrix} \vec{u}_c \\ t_c \end{pmatrix} = \begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \vec{u} \\ t \end{pmatrix} + \begin{pmatrix} T \\ t_0 \end{pmatrix} \quad (1)$$

where $(\vec{u}^T\ t)^T = (x\ y\ z\ t)^T$ denotes the world coordinates of the 4D point, and $(\vec{u}_c^T\ t_c)^T = (x_c\ y_c\ z_c\ t_c)^T$ denotes the point in the camera coordinates. $R$ is the 3x3 rotation matrix and $T$ is the 3x1 translation vector, and $t_0$ is the shift between the local camera time and the world time. The intrinsic projective transformation can be written as:

$$\begin{pmatrix} x' \\ y' \\ t' \end{pmatrix} = \begin{pmatrix} fx_c / z_c \\ fy_c / z_c \\ t_c \end{pmatrix} \quad (2)$$

where $f$ is the focal length of the camera.

The camera paremeters $R$, $T$ can be obtained through calibration. For the simplicity of following analysis we also assume the cameras are synchronized and $t_0 = 0$ for all the cameras.

### 2.3. Optimal condition for the reconstruction

We formulate the 4D space-time reconstruction problem as finding a continuous hyper-surface that optimally explains our observations. We use the same photo inconsistency value as proposed in [3]. For a point $p$ in 4D space-time, the photo inconsistency value $C(p)$ is defined as:

$$C(p) = std(S) + c_0, \text{ where}$$
$$S = \{I_j(q_j) \mid \pi_j(p) = q_j,\ j = 1...n\} \quad (3)$$

and $\pi_j$ is the projective transform from 4D space-time to the j[th] observed image volume, $c_0$ is a small positive constant value (to avoid zero capacitated nodes in the construction of graph cut algorithm given in section 3) and $std(\bullet)$ is the standard deviation of a set of intensity values. We assume the object has a lambertian surface. If there is no occlusion, the photo inconsistency value reaches minimal when a point is on the true hyper-surface. An optimal hyper-surface $\Gamma$ should therefore minimize the sum of $C(p)$ on the

surface, i.e., minimize the cost function:

$$\Gamma = \arg\min \int_{p \in \Gamma} C(p) dS \quad (4)$$

## 3. 4D Node Capacitated Graph Cut

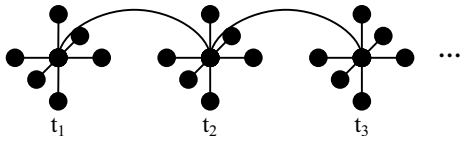



**Fig. 2 4D connectivity of the grid points**

To solve the minimization problem in (4), we first discretize the 4D space-time using a 4D grid. We propose to use 4D node capacitated graph cut (4D node-cut), which is an extension of the 3D node-cut proposed in [3], to find the global minimum. A node-capacitated graph is constructed whose nodes are grid points in the 4D space-time. Each node connects to its neighboring points via 8 edges (Fig. 2). The weight of each node is its photo inconsistency value $C(p)$.

We assume that the object shape in 3D space can be represented by a depth image, where z is the depth direction. The object shape in 4D can be expressed as:

$$z = D(x, y, t) \quad (5)$$

Suppose the object depth remains confined to a given range, i.e., $z \in [z_1, z_2]$. We can then treat those nodes at $z = z_1$ as forming the S (source) plane, and the nodes at $z = z_2$ as forming the T (target) plane (Fig. 3), and use S-T minimum cut algorithm to find a sub-graph with minimum sum of weights such that by removing this sub-graph, the original graph will be divided into two disconnected parts, each of which contains either S-plane or T-plane.

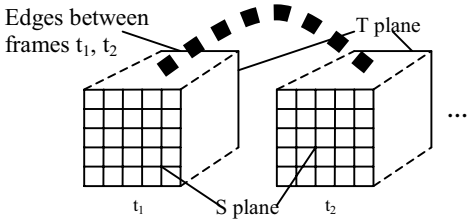



**Fig. 3 The S and T planes in the 4D graph**

This S-T node capacitated graph cut problem can be converted to edge capacitated graph cut problem and solved efficiently [3]. In order to disconnect S-plane from T-plane, each point $(x, y, t)$ in the depth image will have at least one corresponding node in the separating sub-graph. The z value of the corresponding node is the estimated depth for the point. If there are more than one corresponding nodes, we use the average z value of these nodes.

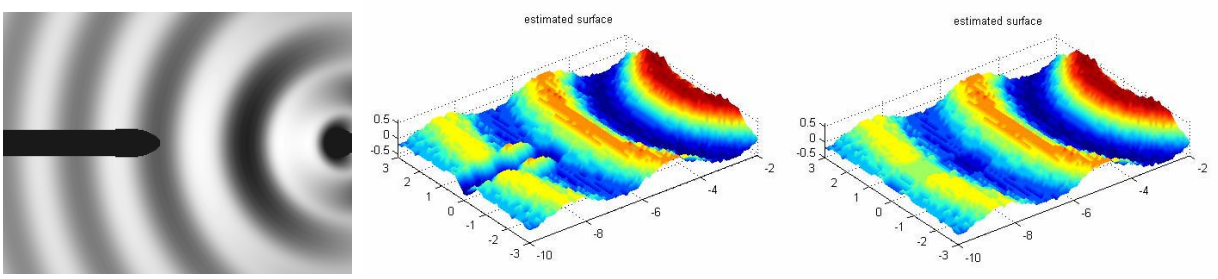
## 4. Experimental Results

We tested our 4D node-cut algorithm on two synthetic data sets. Synthetic data with known ground truth was used to allow quantitative evaluation of the reconstruction. The Wave data set consists of four image sequences (each containing 10 frames) of a wave surface viewed from four different cameras positioned as a 2x2 rig. The wave surface is propagating outward and there is a moving shadow caused by a rod moving across the light source. The 6th frame of the top left camera is shown in Fig. 4 (a). The corresponding reconstruction at the time of frame 6 using 3D node-cut (Fig. 4(b)) and using 4D node-cut on the entire sequence (Fig. 4(c)) are also shown. The Root Mean Square Error (RMSE) of the reconstructed depth at each frame is shown in Fig. 6(a). The error of the quantized version of the ground truth, which is the best possible depth representation using the same volume grid, is also shown for comparison. In the sequence, the moving shadow appears from frame 4 to frame 8. The dark shadow in the image creates a featureless area and thus a low $C(p)$ (ambiguous) neighborhood in the 3D volume. The normal 3D node-cut algorithm fails in this case. However, by taking the time continuity into account, 4D node-cut algorithm finds a global optimal sequence of surfaces directly in the 4D space, thus alleviating the ambiguity problem.

The Morphing Head data set consists of three image sequences (each containing 8 frames) of a human head model morphing from one expression to another. A flat gray moving patch is added to occlude the 3D model in frame 4 and 5 (Fig. 5(a) shows two of the three views at frame 5). Since the patch is outside the 3D reconstruction volume, it should be considered as noise and removed. The reconstruction results for both 4D node-cut on these sequences and 3D node-cut on each frame set are shown in Fig. 5(b). 3D node-cut creates a noticeable artifact on the nose of the face, while 4D node-cut gives a much better result. In Fig. 6(b), the RMSE of the estimated depth also shows that 4D node-cut gives a lower error in the presence of occlusion.

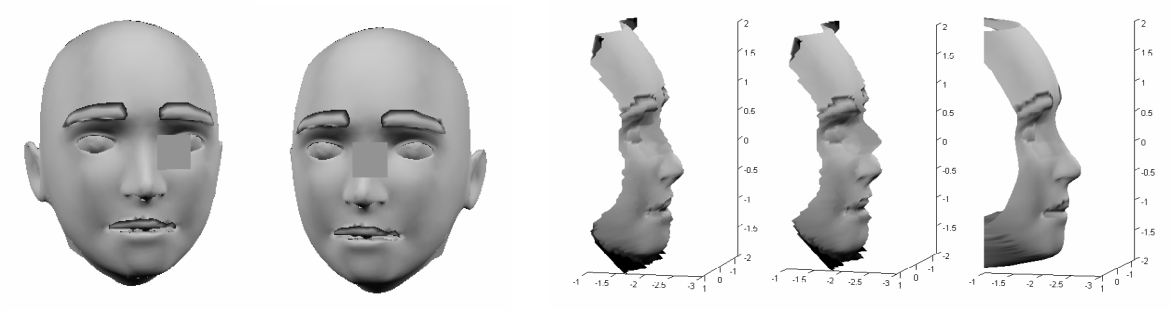
## 5. Conclusion

We have formulated the dynamic 3D shape reconstruction problem as hyper-surface reconstruction in 4D space-time and proposed to use 4D node-cut algorithm to find the optimal solution. The experimental results show that 4D space-time node-cut algorithm exhibits better noise resistance than reconstruction methods that work solely in space domain.



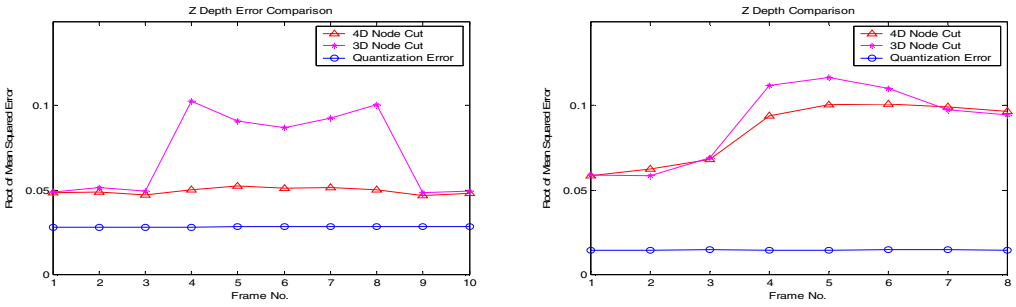

(a) (b) (c)

**Fig. 4 (a) Frame 6 of one input sequence in the Wave data set. (b) The reconstruction by 3D node-cut at frame 6. (c) The reconstruction for frame 6 by 4D node-cut over the entire data set.**



(a) (b)

**Fig. 5 (a) Two images corresponding to frame 5 in Morphing Head data set. (b) The reconstructed surface by 4D node-cut (left) and 3D node-cut (middle) compared with ground truth (right)**




(a) (b)

**Fig. 6 RMSE of the estimated depth for Wave data set (a) and Morphing Head data set (b)**

## 6. References


[1] S. M. Seitz and C. R. Dyer. Photorealistic scene reconstruction by voxel coloring. In Proc. of Conf. on CVPR, pages 1067–1073, 1997.

[2] V. Kolmogorov and R. Zabih. Multi-camera Scene Reconstruction via Graph Cuts. In Proc. of European Conference on Computer Vision, pp 82–96, 2002.

[3] Ning Xu, Tianli Yu and Narendra Ahuja. Shape from color consistency using node cut. In Proc. of Asian Conference on Computer Vision, Jan. 2004

[4] O. Hall-Holt, S. Rusinkiewicz. Stripe Boundary Codes for Real-Time Structured-Light Range Scanning of Moving Objects. Eighth IEEE International Conference on Computer Vision, vol. II, pp.359-366, 2001.

[5] A. Matheny and D. B. Goldgof. The Use of Three- and Four-Dimensional Surface Harmonics for Rigid and Nonrigid Shape Recovery and Representation. IEEE Trans. PAMI, Vol. 17, No. 10, pp967-981, Oct. 1995

[6] L. D. Cohen. Deformable surfaces and parametric models to fit and track 3D data. Systems, Man, and Cybernetics, 1996., IEEE International Conf. on , Volume: 4 , Page(s): 2451 -2456 vol.4, 1996

[7] J. Montagnat and H. Delingette, Space and Time Shape Constrained Deformable Surfaces for 4D Medical Image Segmentation. vol. 1935 of Lectures Notes in Computer Science, Pittsburgh, USA, pp 196-205, Oct. 2000. Springer

[8] J.G. Tamez-Pena, K.J. Parker, and S. Totterman. The Integration of Automatic Segmentation and Motion Tracking for 4D Reconstruction and Visualization of Musculoskeletal Structures. Biomedical Image Analysis, 1998. Proceedings. Workshop on , 26-27 June 1998 pp 154 -163