Constructing a 3D Object Model from Multiple Visual Features

Jiang Yu Zheng

Faculty of Computer Science and Systems Engineering Kyushu Institute of Technology Iizuka, Fukuoka 820, Japan

Abstract

This work aims at building 3D graphics models of objects using multiple visual cues. Different materials and shapes may vield different visual cues such as corner, pattern, highlight, and contour on objects. These cues may have similar appearance in the images, extracted as edge, peak and shading. To obtain a correct shape, different shape recovery methods must be applied to those features separately. We rotate an object and take a dense image sequence. In each Epipolar Plane Image of rotation plane, we classify visual cues according to their image characteristics and motion behavior, and integrate shapes from different algorithms. Features are classified when they pass through a series of modules which first identify distinct features and then reasoning ambiguous features in reference to the identified ones. A carefully considered strategy is proposed and experiments on real objects .

1. Introduction

Recent progresses on multimedia, information network, and virtual reality raise a new requirement of modeling 3D objects. Such objects can be commercial products, industrial parts, folk arts, sculptures, environment, antiques and archaeological finds, etc. New display tools, for example VRML, have also been developed for viewing objects via network. How to input 3D models of objects conveniently and stably thus becomes a challenging problem. Although various laser range finders have been developed for measuring 3D shape, they are still too expensive. We select video camera as the input device for 3D modeling. Our objective is as follows: receiving a standard image sequence through network from a user or an image sequence recorded in a video tape, we construct 3D models and return the data to the user.

To make reconstruction process robust, we fix the way of taking images rather than just waving the camera in the space and solving a general problem. We rotate an object around an axis and take an image sequence in the direction orthogonal to the axis. The rotation angle is controllable or measured in the images. This camera setting is easy to be achieved. It not only simplifies the algorithms of shape recovery, but also creates an intrinsic circumstance for investigating natures of different visual cues. We have studied shape recovery on each rotation plane using contour [1], highlight [2,3], edge [4,5], and shading [13] according to their motions in the corresponding epipolar plane image. We will discuss how to combine their results for a complete model; the key problem here is the classification of features when different visual cues are mixed in the images. Fusion of visual cues is far more difficult than fusing different sensor data which do not interfere each other. Before quantitatively fusing the measured data to get an accurate result, qualitative classification must be done first to avoid a module picking up wrong points to compute.

In order to separate multiple features, we analyze their image behaviors and examine their image properties such as intensity, contrast, image velocity, and coherence of measured positions from consecutive images. The relation of different cues also provides spatial constraints to each other when a feature is unable to be classified according to its own attributes. We propose an approach to do this classification and look visible features in detail to obtain more surface data.

In the following, we will introduce our camera and light setting, and briefly discuss feature behaviors and several shape from X methods for our system. By compare characteristics of features and their relation, we design approaches to identify different cues. Finally, integration of results is done and experiments are shown.

2. Multiple Visual Cues

2.1 Physical Conditions and Visual Cues

Visual features taken into an image are related to object properties including shape, materiel, color and environment attributes including illumination, view point background, and relative motion between a camera and objects. Multiple visual cues appear when both object and environment are given; silhouettes (contour), corners, patterns, texture, highlights, shadows, and shading may be visible on objects (table 1). Unfortunately, it is almost impossible to determine both environment and object only from images. Many research works have restricted type of visual cues to make motion free in shape recovery. In this work, however, we fixed the environment to achieve flexibility in modeling different objects, since our objective is to establish various 3D models. An observed feature can be categorized as either *fixed feature* or moving feature, depending on if it is static on the surface when the object moves. To apply shape from X methods to corresponding cues, we need to qualitatively separate image features into different cues.

2.2 Imaging System and EPIs

We rotate an object around an axis, A camera takes its continuous images in a short time interval (or a small rotation angle). The camera axis is orthogonal to the rotation axis as shown in Fig. 1. A vertical linear light is set at the camera side in the plane containing both rotation axis and the image origin. There is hence no shadow visible on the object. We assume orthogonal projection from the scene to the camera.

Through a simple calibration in the image, we can figure out the projection of rotation axis. We put the camera centered coordinate system C-xyz so that the x and y axes are aligned with EPI and the rotation axis in the image, respectively. The object centered coordinate system O-XYZ is located at the rotation axis and the Z and z axes in the two systems are collinear when the rotation begins. The rotation is clock-wise and the rotation angle is denoted by θ . The angle can either be controlled or can be measured in the image sequence.

object			environment						visual cue		image		shape from X
shape	material	color	light		camera		ra	motion	visual ede		feature		shape nom X
plane surface	diffused		Ţ	C	1	ection)	amera	rotation translation	corner pattern	fixed	edge	ocity	shading
rough smooth convex	specular		direction	radiatio	direction	ance (proj	umber of c	or both	contour highlight shadow	moving	or peak	image velo	contour highlight
concave					dist		Ũ		shading		intensity		

Table 1. Different objects and environments cause multiple visual cues observable in the images.



Fig. 1 Taking images of a rotating object and determining Epipolar Plane Images parallel to rotation planes.

For each rotation plane, we can determine an image line that is the projection of the plane. An Epipolar-plane image (EPI) is collected from the dynamic changing date on the line [1,12]. We will compute shape on each rotation plane using generated Epipolar Plane Image and then build the model by connecting shapes at all rotation planes.

2.3 Motion of Surface Points

As the object rotates, a surface point **P** moves on a circle in the rotation plane and its projection is retained in the corresponding EPI. The point **P**(X, Y, Z) described in the object coordinate system O-XYZ is continuously projected as $p(x, y, \theta)$ in the spatial temporal volume. According to the camera geometry in Fig. 2, the viewing direction is **V**(-sin θ , 0, $\cos\theta$) in the system O-XYZ. The image position of the point viewed at the rotation angle θ can be written as

$$\mathbf{x}(\theta) = \mathbf{P} \cdot \mathbf{x} = \mathbf{X}(\theta) \cos\theta + \mathbf{Z}(\theta) \sin\theta$$

where **x** is the unit vector of the horizontal image axis. Thus, the trajectory of the point in the EPI is a sinusoidal function of the rotation angle with the period of 2π , even the point can not be distinguished as a visual cue.

3. Shape from Xs for a Rotating Object

In this section, we explain individual shape from Xs applied to the rotating objects. We will list conditions for using different visual cues, their motion behaviors in the EPI, algorithms for computing their positions, and accuracy of algorithms. Figure 3 gives an example of EPI showing moving traces of different visual points. Table 2 gives a summary of following explained methods concerning with their principle, result, algorithm complexity, approximation, and source of error.

3.1 Fixed Points and Position Estimation

For a fixed feature such as a corner and a pattern edge on the rotational plane, its position will not change on the object during the rotation; the lines of sight looking at it from different angles always pass the same 3D point that is its 3D location. In the EPI, a fixed point draws moving trajectory as a sinusoidal curve of 2π period. Each point on the trajectory gives the horizontal image position of the point in the corresponding image. Through such an image position, a line of sight can be stretched on the rotation plane, which constrains 3D position of the fixed point on the line. Multiple lines of sight through projections on the trajectory must cross at the fixed point, if there is no pixel error in locating the trajectory. In the real situation, we can use the least squared error method to obtain an accurate measure of the crossing point.



Fig. 2 Camera and object relation in a rotation plane.

0

(1)



Fig. 3 An epipolar plane image in which moving features are marked. The object shape on the rotation plane is like a dumbbell which contains both convex and concave segments. H: highlight, F: fixed point, C: contour

Tracking trajectory of a fixed point gives multiple projections, as well as the same number of the linear equations (Eq. 1). The coordinates of \mathbf{P} from the least squared error method are

Visual cues	Extracted from	Image point constrain what	Recovered primitives	Computational complexity	Measured data versus output	Approximation
Fixed Points	patterns, texture, corners	3D position	point	linear equation	reduntant data single output	improved by averaging multiple output
Contour	static background	surface tangent	envolope, concave part missing	1st order derivative	single measure single output	fitting the derivative using discrete points
Highlight	specular reflection	surface normal	surface curve	1st order differential equation	output using measured data before	integral approximated by sumation

Table 2 Shape recovery methods for different visual cues and their algorithms using approximation.

$$\begin{bmatrix} X \\ Z \end{bmatrix} = \begin{pmatrix} \theta_n & \sin^2 \theta \sum_{\theta=\theta_1}^{\theta_n} \cos^2 \theta - \begin{pmatrix} \theta_n \\ \sum_{\theta=\theta_1}^{\theta_n} \cos \theta \sin \theta \end{pmatrix} \end{pmatrix}^{-2} \\ \begin{bmatrix} \sum_{\theta=\theta_1}^{\theta_n} x(\theta) \cos \theta \sum_{\theta=\theta_1}^{\theta_n} \sin^2 \theta - \sum_{\theta=\theta_1}^{\theta_n} x(\theta) \sin \theta \sum_{\theta=\theta_1}^{\theta_n} \sin \theta \cos \theta \\ \\ \begin{bmatrix} \theta_n \\ \theta_n \\ \sum_{\theta=\theta_1}^{\theta_n} x(\theta) \sin \theta \sum_{\theta=\theta_1}^{\theta_n} \cos^2 \theta - \sum_{\theta=\theta_1}^{\theta_n} x(\theta) \cos \theta \sum_{\theta=\theta_1}^{\theta_n} \cos \theta \sin \theta \end{bmatrix} \end{bmatrix}^{-2}$$
(2)

where $x(\theta)$ is the x coordinate of a trajectory at rotation angle θ , and θ_1 and θ_2 are the first and last angles the fixed point is tracked. The detection of a fixed point is filtering EPIs with a horizontal edge detector.

3.2 Contour and Surface Estimation

The trace of a moving feature in the EPI is not projected from the same surface point. Moving features marked in an EPI traverse on object surface during the rotation. Under orthogonal projection, such traversing for a contour point is on the same rotation plane. At an instance, we can still describe the relation between the contour in the image and the rim on the surface by Eq. 1. Further, the line of sight is the surface tangent on the rim. We can compute derivative of Eq. 1 with respect to rotation angle θ to obtain tangent of the surface (X_{θ}', Z_{θ}') in the rotation plane. By confining the tangent orthogonal to the image plane, we can obtain the position of the surface point as

$$\begin{bmatrix} X \\ Z \end{bmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \end{pmatrix} \begin{pmatrix} x \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \partial x \\ \partial \partial \theta \end{pmatrix}$$

from Eq. 1 and its derivative.

A contour trace can be extracted from a static homogeneous background. In the solution, the derivative of contour position respecting to angle θ is computed from tangent of the contour trace in the EPI using nearby points. The shape in a rotation plane is estimated as the tracking of contour trace proceeds.

(3),

A degenerate case of the method is for nonconvex shape in the rotation plane. A contour trace in the EPI is no longer smooth and the shape unexposed as contour can not be recovered. We have developed an approach to locate such unexposed regions on objects by detecting nonsmooth points on the contour traces in the EPIs [1].

3.3 Highlight on Specular Surfaces

If an object surface has a strong specular reflection, there may exist a highlight on the surface. For objects with fewer corners and patterns, such a highlight provides information for shape recovery. When a surface point becomes a highlight, the angle of incidence is equal to the angle of reflection. Although surface points on a rotation plane may have different Y components of surface normal, they will be highlight under the long linear light when their normal components in the rotation plane rotate to the camera direction (same as the light direction). We assume here the illuminating rays are in the plane containing the rotation axis and the light. If more lights can be put around the object in the same plane, more surface points whose normals are close to the rotation axis can be highlighted.

The highlight shifts over surface points during the rotation, though these points may actually catch rays from different parts of light depending on their normals. For a highlight point, we also have the line of sight as described by Eq. 1. Differentiating Eq. 1 respect to θ , we get equation containing surface tangent $(X_{\theta}^{'}, Z_{\theta}^{'})$. Using the constraint that tangent of a highlight is orthogonal to the line of sight in our camera and light setting, we obtain a 1st order differential equation in the domain $\theta \neq 0$, π as

$$\frac{\partial X}{\partial \theta} - Xctg\,\theta = \frac{\partial x}{\partial \theta}\cos\theta - xctg\,\theta\cos\theta$$

(4)

(5)

(7)

and equation in the domains $\theta \neq \pi/2$, $3\pi/2$ as

$$\frac{\partial Z}{\partial \theta} + Z \tan \theta = \frac{\partial x}{\partial \theta} \sin \theta + x \tan \theta \sin \theta$$

From which the solution of surface points are computed by

$$X(\theta) = \left(\int_{\theta_i}^{\theta} \left(\frac{\partial}{\partial \theta} \cos \theta - x c t g \theta \cos \theta \right) e^{-\int_{\theta_i}^{\theta} c t g \, \theta d \theta} d\theta + X_{\theta_i} \right) e^{\int_{\theta_i}^{\theta} c t g \, \theta d \theta}_{\theta_i}$$

$$Z(\theta) = (x(\theta) - X(\theta)\cos\theta) / \sin\theta$$
(6)

in the θ domains $[\pi/4, 3\pi/4]$ and $[\pi+\pi/4, \pi+3\pi/4]$, and $X(\theta) = (x(\theta) - Z(\theta) \sin\theta) / \cos\theta$

$$Z(\theta) = \left(\int_{\theta_i}^{\theta} \left(\frac{\partial x}{\partial \theta}\sin\theta + x\sin\theta\tan\theta\right) \mathcal{C}\int_{\theta_i}^{\theta} \tan\theta d\theta d\theta + Z_{\theta_i} \mathcal{C}_{\theta_i}^{\theta} - \int_{\theta}^{\theta} \tan\theta d\theta d\theta \right)$$

in the θ domains $[-\pi/4, \pi/4]$ and $[\pi-\pi/4, \pi+\pi/4]$. In the solutions, $[X_{\theta_i}, Z_{\theta_i}]$ is a known point highlighted at angle θ_i , which works as initial condition of the equation.

We track highlight trace from $\theta_{\theta i}$ in the EPI, and estimate positions of highlight passed points by computing the summations for the integral terms in the solutions. It results a curve that is the shape in the rotation plane. This computation may include accumulated error if a highlight trace is not located precisely in the EPI.

The moving trace of a highlight has some interesting characteristics. A linear shape which has zero curvature on the rotation plane will generate a horizontal highlight stripe in the EPI, because the points on a plane have the same normals and become highlights at the same rotation angle. At convex and concave boundaries separated by a zero curvature point or segment, highlights may appear at the same time (or viewed from the same rotation angle). They will either split or merge at the zero curvature points [2,3,10].

A highlight trace can be located in the EPI by picking up

intensity peaks. For a strong highlight, two edges bounding a trace of peak can also be located in the EPI.

4. Relations between Different Visual Cues

This section discusses the relationship between different visual cues presented in intensities, traces in EPI and estimated locations.

4.1 Relation on Motions of Visual Cues

We look at the motion of contour, highlight, and fixed point in the EPI. Figure 4 displays three shapes on the rotation plane and traces of multi-features in the EPIs. The trace of a fixed point from either corner or pattern edge diverges from a trace of contour and then merges into the other trace of contour after half period of rotation. A trace of fixed point may also be cut off by a contour trace in the case of occlusion.

A trace of highlight moves over traces of surface points. If a surface shape in the rotation plane has a high curvature, it becomes a corner either convex or concave. Highlight thus keeps staying there and the trace of highlight is then close to the sinusoidal trace of the corner. If the corner has ideal infinite curvature, the highlight has no width so that its trace disappears. The trace of corner only connects traces of highlight.

On the other hand, if a fixed point comes from an edge of pattern on the smooth surface, its trace may have an intersection with a highlight trace. Depending on the sign of curvature of the surface, a trace of highlight intersects trace of surface points as well as pattern edges in different way. For a convex surface in the rotation plane, the highlight moves slower than surface points so that its trace is more vertical than those of pattern edges in the EPI. For a concave surface, however, the highlight moves fast than surface points so that its trace is more horizontal. A crossing point of highlight and fixed point traces provides the initial position for estimating other points along the highlight trace.

In addition, a highlight trace may have a wave at a concave surface generated region in the EPI (see second row of Fig. 4). It is in fact three segments of traces. A highlight splits into two at a zero curvature point. One of which moves on the concave surface and then merges with another highlight at another zero curvature point.

As to the relation between highlight and contour, there is no intersection between their traces in the EPI, because the light is with the camera. However, if a highlight trace has no intersection with fixed points, we can draw a trace of surface point from contour to intersect it for acquiring the initial condition. Figure 5 summarizes the above discussed relations.

4.2 Intensities of Different Visual Cues

The basic approaches to detect fixed points, contours and highlights are filtering EPI by an edge detector and locating peaks, respectively. By using a background with different color from the object, contour is easy to be located if we search edges from both sides towards the center.

The contrast of a corner point is reversed in sign one time on its trace. At a particular angle where the corner rotates toward the light and camera, the intensities on its both sides have no difference so that it can not be extracted as edge. For a pattern edge, however, its contrast becomes maximum when the normal there directs to the light and camera.

The intensity of a surface point captured by a camera is the sum of diffused component and specular component of reflection. Diffused reflectance changes at pattern edges. If a surface has a strong specular component of reflection, a highlight can be located at a sharp peak of intensity in the EPI without influenced by underlying pattern changes. If only weak specular component exists, bright pixels in the EPI may contain both highlights and bright patterns. The intensity on the highlight trace may become low when the trace moves over a dark region on the surface. This causes difficulty in locating highlight traces.



Fig. 4 Shapes with different cues and their corresponding traces in the EPIs. Left: shapes, Middle: EPIs, Right: traces.



Fig. 5 Relation in traces of visual cues.

5. Classification and Fusion of Visual Cues 5.1 Separating Different Traces in EPIs

Based on above analyses, we will design a strategy to separate traces of different visual cues for final fusion. It contains three phases based on image property, spatial distribution of measured data, and motion characteristics. Figure 6 gives detailed module and data flow.

We first extract intensity peaks higher than a given threshold in the EPI. The result may include strong highlights and bright patterns on convex surfaces. At the same time, edges are detected using ordinary edge detector. It results traces of fixed points and bounding curves at highlight traces. Searching edges from both sides, we obtain contour traces. Deleting pixels belong to contour and strong highlight, the remaining edges are fixed points and weak highlights.

The second check is to see whether the locally measured positions along a trace of edge are centralized at a point. If the condition is verified, it is a trace of fixed point; otherwise, it is a trace of moving point. For those short traces not providing enough evidence in test, we leave them as candidates of fixed points since the possible computation applied to them is local and their positions will not spread away.

The third phase is to look at image velocities of highlight and fixed point. We focus on intersections of two traces in the EPI. An intersection of traces means that two visual cues are visible at the same surface point. If two conflict measures are obtained from them, only one result is true. Once their traces have an intersection, the intersecting type tells highlight and fixed point, if we get curvature sign from contour.



Figure 6. Modules for classification and data flow of different visual cues.

5.2. Integrating Results from Visual Cues

The integration of measured data from different algorithms is mainly qualitative here. Fixed points have highest accuracy. The second accurate cue is contour, then followed by highlight. We choose the result from a cue with less error, if multiple measured data are available at a surface point. At a point with one or without measured data, we just use it or connect nearby points using linear graphics patches.

In more detail cases, if we obtain data from both contour and fixed point, we choose fixed point. A long trace of fixed point can provide multiple measure of the position. By extending such a long trace to contour trace using the measured position, we can locate which contour point is overlapped with the fixed point. On the other hand, highlight measure can not be done without initial condition. Whenever a highlight trace intersects a trace of fixed point, we start estimation of highlight from the crossing point; the highlight position integrated from far away along its trace may have accumulated error.

6. Experiments

We have done experiments on real objects. A video camera is put 3-4m away from the objects. The object is put on a turn table. Camera lens is adjusted to long focal length to produce close condition of orthogonal projection. Figure 7 and Figure 8 give two examples of modeling objects. Each shows the recovered model and shape estimated at one EPI, which is combined from results of different shape from X's. EPIs are taken at each degree of rotation.

7. Conclusion

In this paper, we proposed a strategy to classify multiple visual cues and integrate measured data from their separated 3D recovery modules. Objects are rotated and we extract contour, corner, pattern, and highlight, as well as their motion information for visual data fusion. The analysis and processing are given at epipolar plane image parallel to the rotation plane. By carefully comparing characteristics of different visual cues and recovery algorithms, we are able to design a qualitative scheme to do the fusion. Dealing multi-features makes the 3D modelling process more robust; a system handling only one visual cue will fail to give correct results if other visual data are mixed in due to different object conditions. It also produces more data than a single visual cue does, which is important in generating graphics models for multimedia use.

References

- J.Y. Zheng, "Acquiring 3D models from sequences of contours", IEEE PAMI, Vol.16, No.2, Feb. pp.163-178, 1994.
- [2] J. Y. Zheng, Y. Fukagawa, T. Ohtsuka, N. Abe, "Acquiring 3D models from rotation and a highlight", 12th ICPR, Vol. 1, pp. 331-336, 1994.
- [3] J. Y. Zheng, Y. Fukagawa, N. Abe, Shape and Model from Specular Motion, 5th ICCV, pp.92-97, 1995.
- [4] J. Y. Zheng, and F. Kishino, "Verifying and combining different visual cues into a complete 3D model", CVPR92, pp. 777-780, 1992.
- [5] J. Y. Zheng, H. Kakinoki, K. Tanaka, and N. Abe, "Acquiring 3D models from fixed point during rotation", 3th ICARCV, Vol. 1, pp.459-463, 1994.
- [6] K. P. Horn and M. J. Brooks, "Shape from shading", MIT Press, Cambridge, MA, 1989.
- [7] A. Pentland, "Local shading analysis", IEEE PAMI, 6:170-187, 1984.
- [8] R. J. Woodham, "Gradient and curvature from photometric stereo including local confidence estimation", J. Opt. Soc. Amer., Vol. 11, no. 11, 3050-3068, 1994.
- [9] J. Aloimonos and A. Bandyopadhyay, "Active vision", 1st ICCV, pp.35-55, 1987.
- [10] J. Y. Zheng, Y. Fukagawa, and N. Abe, "3D shape estimation and model construction from specular motion in the images" submitted to IEEE PAMI, 1995.
- [11] R. Vaillant and O. D. Faugeras, "Using extremal boundaries for 3D object modeling", IEEE PAMI, Vol. 14, No. 2, pp. 157-173. Feb. 1992.

- [12] H. Baker, and R. Bolles, "Generalizing epipolar-plane image analysis on the spatiotemporal surface", CVPR-88, pp.2-9, 1988.
- [13] J. Y. Zheng, H. Kakinoki, K. Tanaka and N. Abe "Computing 3D Models of Rotating Objects from Moving Shading", ICPR96.
- [14] J. Lu and J. Little, "Reflectance function estimation and shape recovery from image sequence of a rotating object", 5th ICCV, pp. 80-86, 1995.





(k)





(i) Shape from contour. (j) Positions of fixed points.



(k) Partial shape from highlight. (l) integrated result of shapes from several cues.

Figure 8. Recovery of a plane model from contours, fixed points and highlights. d-l shows detailed processing at an EPI. The \times signs in the shapes indicate the positions of the rotation axis.