Stereovision by Active Surface Model

M. Yokomichi*, H. Sugiyama*, and M. Kono*

*Department of Computer Science and Systems Engineering, University of Miyazaki, Miyazaki 889-2155, Japan
(Tel: +81-985-58-7420; Email:yokomich@cs.miyazaki-u.ac.jp)

Abstract: Stereovision is known to be one of the most important tools for robot vision systems. Previously, 2D active contour model has been applied to stereovision by defining the contour on the 3D space instead of image plane. However, the proposed model is still that of curve so that some complex shapes such as surfaces with high curvature can not be properly estimated because of occlusion phenomena. In this paper, the authors extend the curve model to the surface model. The surface is approximated by polygons and new energy function and its optimization method for surface estimation is proposed. Its effectiveness is examined by experiments with real stereo images.

Keywords: stereovision, active contour, active surface, polygon

1. Introduction

In the fields of robot vision, stereo vision is known to be one of the main tools for passive sensors. There has been proposed many approaches for stereovision algorithms such as region based approaches (block matching, etc.), feature based approaches, neural networks approaches, and so on. Many of these approaches uses stereo matching, that is, for every point in the left (right) image, finding the corresponding point in the right (left) image leads to the three dimensional position of the point. However, finding the corresponding point itself is not easy task, because there are many possible false matches due to photometric distortion, projective distortion, lack of texture, noise, and occlusion, etc.

Active Contour Model (SNAKES), proposed by Kass et al. [6] was originally applied to the contour tracking problem for single image (sequence) and then extended to the stereovision. Terzopoulos et al. [7] use multiple 2D contours in each of stereo image plane and the contours are coupled by a stereo disparity smoothness term in their potential functionals. A deformable B-spline model for stereo matching in 3D space was proposed by Bascle and Deriche [1], in which an additive potential function based on the projection of the curve to each image was used. In Cham and Cipolla [3], the authors propose an affine epipolar geometry scheme for coupling pairs of active contours in stereo images. Deriche et al. [4] use stereo vision and geodesic active contours to trace 2D planer curves that lie at the intersection of the observed scene with a given plane in 3D space. Recently, Zaritsky et al. [8] extended the 3D curve model to the stereo image sequences by velocity-guided tracking.

However, their models are still that of curve so that they can not treat the occlusions of the object properly.

In this paper, the authors extend the results in [8] by applying the active surface model. The active (deformable) surface model has already been proposed by Terzopoulos et al. [7] and studied by many researchers, e.g., Caselles et al. [2] and Faugeras and Keriven [5] including the stereovision. However, many of them mainly analyzed the theoretical aspects and practical implementations were not considered so match.

The proposed model is already discretized by means of polygons. In this case, the energy function is consist of two parts: the energy for points on the boundary of the surface and that for the polygons. The image energy in [8] is modified such that the occlusion is properly included. Furthermore, in order to decrease the computational cost, the position of the internal points on the surface, which defines the polygons are decided such that only the area of the surface is minimized.

The effectiveness of the proposed method is examined by the experiment with real stereo images.

2. Preliminaries

In this section, briefly we preview the stereovision, the 2D-active contour model, and 3D-active contour model.

2.1. Stereovision

In this paper, we consider the general stereo camera configuration described in Fig1. Let $M = (X,Y,Z)$ denote a point in the 3D space with some fixed Cartesian coordinate frame. Each camera (left and right) has own coordinate frame ($(X_l,Y_l,Z_l)^T$ and $(X_r,Y_r,Z_r)^T$, respectively) and image coordinate ($(x_l,y_l)^T$ and $(x_r,y_r)^T$, respectively). Suppose that the point M is projected to left (right) image plane at $m^l = (m^l_x, m^l_y)^T$ ($m^r = (m^r_x, m^r_y)^T$). If all internal and external parameters of each camera is available with no uncertainties, we can specify the projections

$$m^k = H^k(M), k = l, r.$$

Conversely, if m^l and m^r is obtained, we can obtain M by means of geometric relations.

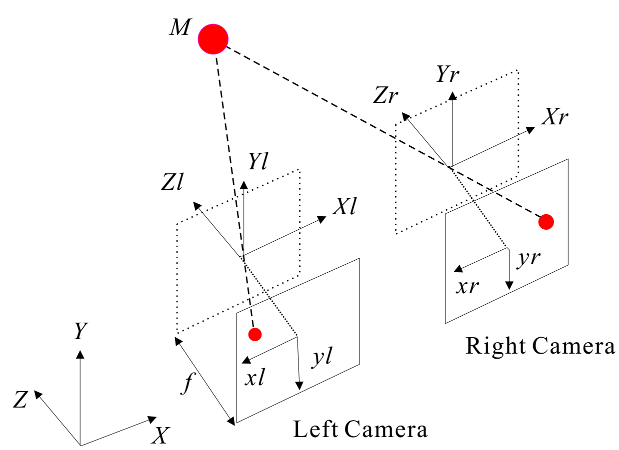

Fig. 1. The stereo camera configuration

2.2. SNAKES

The active contour model (SNAKES) was originally proposed as a contour tracking method for single (2D) image and then extended to stereovision. Let $I : [0 : a] \times [0 : b] \mapsto \Re^+$ be the intensity of the given image. In the single image problem, the contour is defined by parametrized curve $v(t)$ on the image plane. If we set (x,y) as the coordinate system on the image plane, then the contour is described as

$$v(t) = (v_x(t), v_y(t)), 0 \le t \le 1. \tag{1}$$

The initial contour is deformed such that the properly defined energy functional is minimized. In the general setting, the energy functional is consist of three parts; internal energy, image energy, and external energy (optional)

$$E(v) = \int_0^1 \alpha E_{\text{int}}(v(t)) + \beta E_{\text{image}}(v(t)) + \gamma E_{\text{ext}}(v(t))dt, \tag{2}$$

where $\alpha > 0$, $\beta > 0$, and $\gamma > 0$ are weight parameters and usually

$$E_{\text{int}}(v(t)) = \|\frac{dv}{dt}\|^2 + \lambda \|\frac{d^2v}{dt^2}\|^2, \lambda > 0$$
$$E_{\text{image}}(v(t)) = -\|\nabla I(v(t))\|^2$$

are used. The above definition means that, if the contour become smooth and it approaches the edge on the image, then the energy decreases. In the real implementation, the curve $v(t)$ is approximated by sequence of points $\{v_1, ..., v_n\}$ and the energy function is minimized by several approaches such that

- Dynamic programming
- Greedy algorithm
- Variational method

and so on.

2.3. 3D SNAKES

In [8], 2D SNAKE is extended to the stereovision by the following manner. Firstly, the contour is defined on the real 3D space instead of the image plane. Thus, it is described as

$$v(t) = (v_x(t), v_y(t), v_z(t))^T. \tag{3}$$

If both internal and external camera parameters are known, this curve can be projected to each image plane. Let us denote the projected curves as $H^l(v(t))$ and $H^r(v(t))$, and define $I^l(v(t)) := I^l(H^l(v(t)))$ and $I^l(v(t)) := I^l(H^l(v(t)))$ as the corresponding intensities. Then, the energy functional is defined, similar to the 2D case, as

$$E(v) = \int_0^1 \alpha E_{\text{int}}(v(t)) + \beta E_{\text{image}}(v(t)) + \gamma E_{\text{ext}}(v(t))dt, \tag{4}$$

where the first and the third term are the same as the 2D case. The second term can be defined using two intensities $I^l(v(t))$ and $I^r(v(t))$ as several manners.

1. Additive: $E_{\text{image}}(v(t)) = -(\|\nabla I^l(v(t))\|^2 + \|\nabla I^r(v(t))\|^2)$
2. Multiplicative: $E_{\text{image}}(v(t)) = -\sqrt{(\|\nabla I^l(v(t))\|^2 \|\nabla I^r(v(t))\|^2)}$

The purposes of these image energies are to move the contour so that its projections are attracted to relevant image features, such as image edges.

2.4. Implementation issues

The mathematical model of contour is a continuously parametrized curve (3). In the practical use, it is discretized or approximated by several typical curves, e.g., spline curve. If the discretization by sequence of points is used, the curve is represented by

$$v_i = (x_i, y_i, z_i), i = 0, ..., N; v_0 = v_N. \tag{5}$$

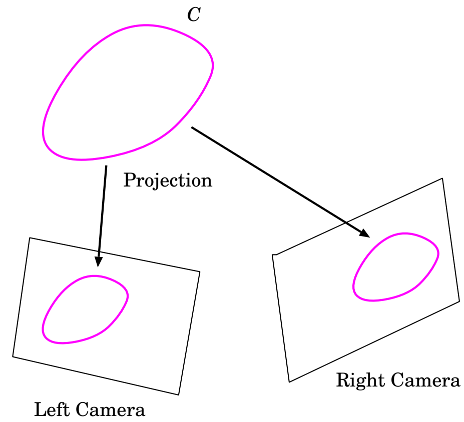

Fig. 2. 3D SNAKES

In this case, the differentials in the internal energy become finite difference and the energy functional is described by sum of energy functions

$$E(\mathbf{v}) = \sum_{i=1}^{N} \left(\alpha E_{\text{int}}(v_i) + \beta E_{\text{image}}(v_i) + \gamma E_{\text{ext}}(v_i)\right). \tag{6}$$

In order to minimize the energy function, two kinds of approaches are generally used. The one is the search-based approach, and the other is variational one.

3. Main Results

The 3D active contour model explained in the previous section is that of curves, so that it can not treat the occlusions correctly. Let us consider a situation depicted by Fig. 3. In this case, the contour

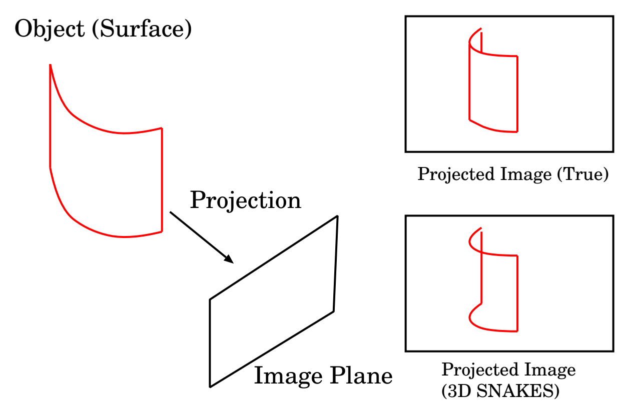

Fig. 3. Occlusion problem

is given as a boundary curve of certain curved surface with high curvature. The projected images are shown in the right side of Fig. 3. If the 3D contour model is applied, all points in the contour should be seen in the projected image. However, in the real image, some points are hidden by the surface because of the occlusion. In addition, some internal points on the surface can be seen as image edges. This leads to the fact that the image energy in the 3D active contour model does not have correct value for hidden points. To overcome this difficulty, the surface model should be introduced.

The surface can be modeled in many ways both continuous and discrete cases. In this paper, taking into practical use, an discrete model which is based on the polygon is adopted.

3.1. Polygon model

In the proposed model, a surface S is defined by two kinds of points–the points $\mathbf{v}^e = \{v_j^e, j = 0, ..., n\}$ on the edge of the surface and the internal points $\mathbf{v}^{\text{in}} = \{v_j^{\text{in}}, j = 0, ..., m\}$, and triangles $d_j, j = 0, ..., r$ defined by three points in $\mathbf{v}^e$ and $\mathbf{v}^{\text{in}}$ such that

1. $\forall i \neq j,\ d_i \cap d_j = \emptyset$.
2. $\cup_i d_i = S$.

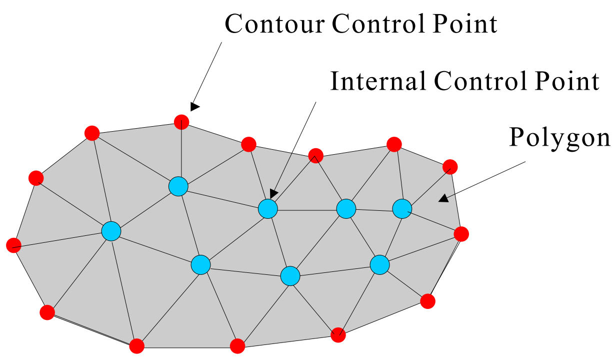

Fig. 4. Polygon model

Furthermore, for every points in $\mathbf{v}^e$ ($\mathbf{v}^{\text{in}}$), we define D_j^{in} (D_j^{e}) as the set of triangles d_i's that contains the points.

3.2. Energy functions

In the above settings, the energy functions are defined for both points on the edge and the polygons.

For the polygons, the energy is defined such that they spans the surface uniformly and the area become smaller. Here, suppose that $\Delta(d_j)$ denotes the area of a polygon d_j and d_j consist of three points v_{j1}, v_{j2}, and v_{j3}. If

$$\begin{aligned}\delta(d_j) := &|(\|v_{j1}-v_{j2}\| - \|v_{j1}-v_{j3}\|)|/\Delta(d_j)\\ &+|(\|v_{j2}-v_{j3}\| - \|v_{j2}-v_{j1}\|)|/\Delta(d_j)\\ &+|(\|v_{j3}-v_{j1}\| - \|v_{j3}-v_{j2}\|)/\Delta(d_j)|\end{aligned}$$

is small, then, d_j becomes close to equilateral triangle. This discussion leads to the following candidate of the energy function:

$$E_{\text{polygon}} := \sum_{j=1}^{r} \alpha_1 \Delta(d_j) + \alpha_2 \delta(d_j). \tag{7}$$

For the points on the edge, the energy function is similar to that of the 3D active contour model (4). However, the image energy differs because of the occlusions. The occlusion is checked by the following manner and is depicted in Fig 5. In Fig 5, v, d, and f are the point and the polygon to be checked, and the focal point, respectively. If the line segment between v and f crosses d, then occlusion occurs. In this paper, we pose the assumption below:

Assumption: *For any points on the edge of the contour, the occlusion occurs at most one camera.*

Based on this discussion, the energy function for $\mathbf{v}^e$ is modified as follows:

$$E_{\text{edge}} = \alpha E_{\text{int}} + \beta E_{\text{image}} + \gamma E_{\text{ext}}, \tag{8}$$

where, E_{int} and E_{ext} is the same as before, and E_{image} is modified as (additive case)

$$E_{\text{image}}(v_j^e) = \begin{cases} -(\|\nabla I^l(v_j^e)\|^2 + \|\nabla I^r(v_j^e)\|^2) : (\text{no occlusion}) \\ -2\|\nabla I^l(v_j^e)\|^2 : (\text{Right camera has occlusion}) \\ -2\|\nabla I^r(v_j^e)\|^2 : (\text{Left camera has occlusion}). \end{cases} \tag{9}$$

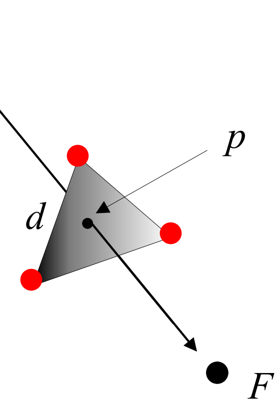

Fig. 5. Occlusion check

Note that this energy function depends on the position of the internal points.

3.3. Optimization Algorithm

According to the extension described above, the total energy function is given as

$$E(\mathbf{v}^e, \mathbf{v}^{\text{in}}) = E_{\text{polygon}}(\mathbf{v}^e, \mathbf{v}^{\text{in}}) + E_{\text{edge}}(\mathbf{v}^e, \mathbf{v}^{\text{in}}), \tag{10}$$

where each term in the right hand side of (10) is defined by (7),(8), and (9).

It is clear that the energy function (10) depends on both the internal points and the points on the edge. If we use an optimization method such as greedy algorithm, DP, and so on, for both points equally, it may be expensive computationally for real time operation because the number of the points becomes very large. To overcome this difficulty, in this paper, a two-step optimization algorithm is adopted. That is, the following minimization problem is considered.

At first, for given points on the edge $\mathbf{v}^e$, suppose that $\bar{\mathbf{e}}^{\text{in}}(\mathbf{v}^e)$ denotes one set of the internal points which minimize E_{polygon}, that is,

$$E_{\text{polygon}}(\mathbf{v}^e, \bar{\mathbf{v}}^{\text{in}}(\mathbf{v}^e)) = \min_{\mathbf{v}^e} E_{\text{polygon}}(\mathbf{v}^e, \mathbf{v}^{\text{in}}).$$

Then, the minimization problem we should solve is described as follows:

Problem:

Find $\mathbf{v}^e$ such that

$$E_{\text{edge}}(\mathbf{v}^e, \bar{\mathbf{v}}^{\text{in}}), + E_{\text{polygon}}(\mathbf{v}^e, \bar{\mathbf{v}}^{\text{in}}) \tag{11}$$

is minimized.

In the case where greedy method is used, the algorithm is summarized as follows:

1. For all $j \in [0 : N]$:
   (a) Move v_j^e in the search region.
   i. Move all internal points in each search region such that E_{polygon} is minimized.
   ii. Compute E_{edge} and E.
   (b) Set the position of v_j^e which minimize E as new v_j^e.

4. Experiments

In this section, some experimental results are presented. All images (of size 320x240, color depth RGB24) are obtained by a pair of CCD cameras (Creative WebCam NX Pro). The configuration of the cameras and the object is shown in Fig. 6.

Fig. 6. Configuration of object and cameras

The internal and external parameters of each camera is calibrated by the method proposed by Zhang [9]. The stereo images is shown in Fig. 7. We can see that the left hand side of the object is occluded in the right image.

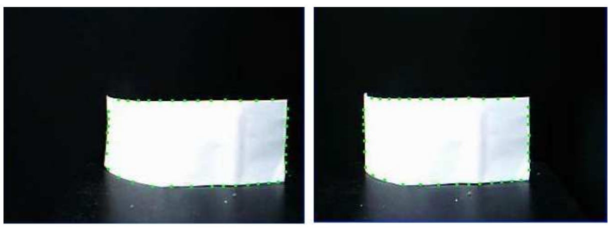

Fig. 7. stereo images (left: left image, right: right image)

In this experiment, three types of methods are examined.

1. Original 3D active contour.
2. 3D active surface without considering the occlusion.
3. 3D active surface with considering the occlusion.

In each method, the contour is consist of 40 points. In the active surface model, 78 internal points and 195 polygons are used. The results are presented in Fig. 8, Fig. 9, and Fig. 10.

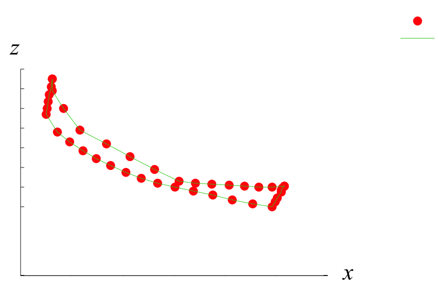

Fig. 8. Extracted contour by 3D SNAKES

We can see that the active surface model which allows the occlusion extracts the shape of the object correctly.

5. Conclusions

In this paper, an active surface model for stereovision is proposed. From the experiments, it is shown that the proposed model can extract 3D shapes correctly. In this paper, search-based optimization is adopted. Applying the other methods such as geodesic method will also be examined in the future works.

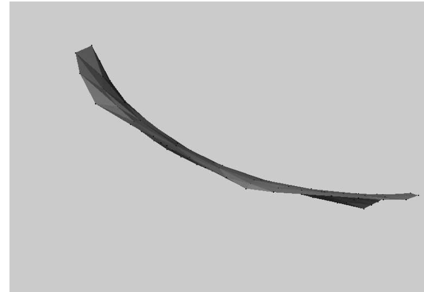

Fig. 9. Extracted surface (unconsidered occlusion)

Fig. 10. Extracted surface (improved)

References

[1] B. Bascle and R. Deriche, “Stereo matching, reconstruction and refinement of 3D curves using deformable contours”’ *IEEE Int. Conf. of Computer Vision*,Berlin, pp. 421–430, 1993.

[2] V. Caselles, R. Kimmel, and G. Sapiro, “Minimal surface based object segmentation,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, pp. 394–398, 1997.

[3] T. Cham and R. Cipolla, “Stereo coupled active contours,” *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Puerto Rico, pp. 1094–1099, 1997.

[4] R. Deriche, S. Bouvin, and O. Faugeras, “Front propagation and level-set approach for geodesic active stereovision,” *Third Asian Conference on Computer Vision*, Hong Kong, 1998.

[5] O. Faugeras and R. Kimmel, “Variational principles, surface evolution, PDE’s revel set methods and the stereo vision,” *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 336–344, 1999.

[6] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” *Int. J. Computer Vision*, vol. 1, no. 4, pp. 321–331, 1987.

[7] D. Terzopoulos, A. Witkin, and M. Kass, “Constraints on deformable models: recovering 3D shape and nonrigid motion,” *Artificial Intelligence*, vol. 36, pp. 91–123, 1988.

[8] R. Zaritsky, N. Peterfreund, and N. Shimkin, “Velocity-guided tracking of deformable contours in three dimensional space,” *Int. J. of Computer Vision*, vol. 51, no. 3, pp. 219–238, 2003.

[9] Z.Zhang, “A flexible new technique for camera calibration,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.