A Method of 3D Object Reconstruction by Fusing Vision with Touch Using Internal Models with Global and Local Deformations

著者	石黒 章夫
journal or publication title	IEEE International Conference on Robotics and Automation, 1993. Proceedings
volume	1993
number	2
page range	782-787
year	1993
URL	http://hdl.handle.net/10097/46681

doi: 10.1109/ROBOT.1993.291939

A Method of 3D Object Reconstruction by Fusing Vision with Touch Using Internal Models with Global and Local Deformations

Yasuharu YAMADA, Akio ISHIGURO, Yoshiki UCHIKAWA

Dept. of Electronic-Mechanical Engineering., School of Engineering.
Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-01, JAPAN

Abstract

Recently, the necessity of developing sensor fusion systems has become much higher. In this paper, a method for sensor fusion of vision and touch using an internal model with both global and local deformation is introduced. We utilize superquadrics with local deformations as internal models. Our proposed method consists of two phases. At the first phase, we recover the object shape parametrically by changing the parameters of superquadrics using visual data. The internal model constructed in this phase is, of course, a rough representation of object shape, since visual data are given merely from the visible portion of the object, and parametric models such as superquadrics have inevitable limitation in shape representation. But thanks to this parametric model, we can easily extract regions which are invisible and/or have large errors. Thus, at the second phase, we make the tactile sensor explore to get information of the above-mentioned regions, and deform the internal model locally based on the defined energy functions. The feasibility of the proposed method is confirmed by simulations.

1. Introduction

We can easily recognize the objects in the outer world by fusing many different kinds of senses. Engineering realization of such function is called a sensor fusion, and recently its necessity has become much higher for autonomous robots and so on [1]. With regard to measurements and recognition of the objects, it is not too much to say that most of them are performed using vision and touch. Thus it is very useful to consider how to fuse visual and tactile information. Several methods have been already reported on object recognition using vision and touch, but their methods have problems since they need tremendously large database on objects and so on [2]-[4].

As for sensor fusion, the importance of constructing internal models which reflect the outer world has been pointed out [1][5]. Considering the case of human's ability to recognize objects, it is found that we firstly obtain global but approximate information on objects by vision, and then get local but exact information by touch. By using these characteristics of senses, we can exactly recognize the surrounding environment. From these considerations, we propose a method of fusing vision and touch by introducing energy functions based on the above-mentioned characteristics of sensors. To realize our proposed method, we utilize superquadrics with local deformation as the internal models. Moreover, we show this internal model is useful for recognition.

2. Internal models

It is very important to determine how to construct internal models. We must construct them reflecting the precise knowledge of the observed object, and control numerous sensors by operating this model. There are various methods for representation of internal models, but the following characteristics are needed:

- Internal models should be represented by easy descriptions.
- Internal models can represent various shapes.
- Each sensor can efficiently extract the information of internal models.
- Recognition of the shape of an object is also possible with use of the obtained internal models.

The conventional methods to represent 3D objects can be generally classified into two types. One is a parametric model such as *generalized cylinders*, and the other is a nonparametric model such as *snakes* which are deformed according to the specified energy function[10].

When we use a parametric model as an internal model, the knowledge about the represented object can be extracted easily with a few parameters. But this model has a serious problem due to its limitation of shape representation.

1050-4729/93 $3.00 © 1993 IEEE

Contrary to the above model, a nonparametric model can represent various shapes in detail [6]. However, it is difficult to capture the feature of the represented shapes. To satisfy these conflicting requirements, *i.e.* shape representation and recognition, we solve this problem by deforming a parametric model nonparametrically. The details of this method will be discussed later.

3. Proposed method for fusing vision with touch

3.1 Definition of sensor fusion

We define the sensor fusion as minimization of the total sum of each sensor's deforming energy of the internal model. The energy function is defined as:

$$\boldsymbol{E} = k_1 \boldsymbol{E}_{vision} + k_2 \boldsymbol{E}_{touch} \ , \tag{1}$$

where $\boldsymbol{E}$ represents the total deforming energy of the internal model. $\boldsymbol{E}_{vision}$ and $\boldsymbol{E}_{touch}$ denote the deforming energies caused by visual and tactile information, respectively. k_1 and k_2 are coefficients. As mentioned before, we define $\boldsymbol{E}_{vision}$ and $\boldsymbol{E}_{touch}$ based on the characteristic of each sensor. Namely, $\boldsymbol{E}_{vision}$ works to deform the internal model globally and parametrically, whereas $\boldsymbol{E}_{touch}$ works locally and non-parametrically.

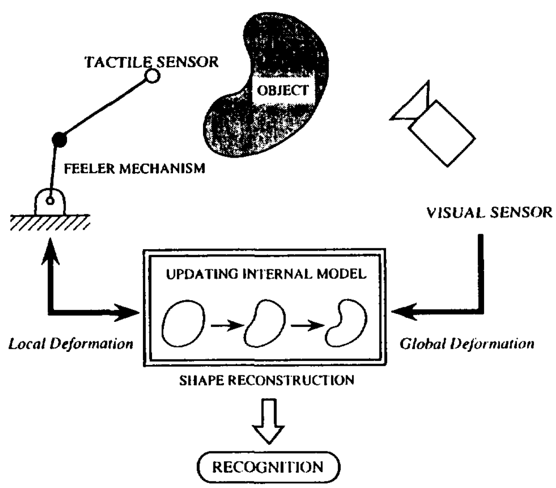


Fig.1: System configuration.

Fig.1 shows the system configuration, schematically. From the above definition, we can construct the internal model efficiently. Moreover, we can decide the superiority of the sensors by changing the coefficients k_1 and k_2. As a rudimentary stage of investigation, we make the sensing process perform under visual superiority. Namely, we simply divide the sensing process into two phases. At the first phase, the sensing process is proceeded under the condition of $k_1 = 1, k_2 = 0$. This means that the internal model is exclusively deformed by the obtained visual information. Then, at the second phase, the internal model is updated locally using the tactile information under the condition of $k_1 = 0, k_2 = 1$.

3.2 The first phase of the sensing process

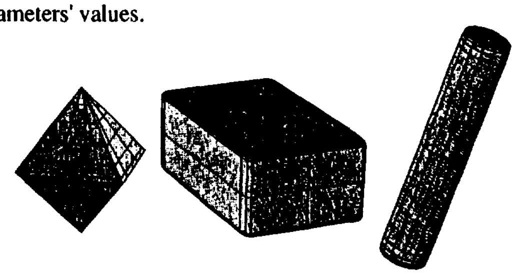
3.2.1 Shape recovery of superquadrics

As stated above, we use a parametric model as a basic internal model before local deformation. Therefore we must construct the internal model by using only the visible portion of the object at this phase. Superquadrics are well-known parametric models, and have been used for shape representation in the field of computer graphics. Points on superquadric surfaces are defined by the following simple equations:

$$\begin{aligned} x &= a_1 \cos^{\varepsilon_1}(\eta)\cos^{\varepsilon_2}(\omega) \\ y &= a_2 \cos^{\varepsilon_1}(\eta)\sin^{\varepsilon_2}(\omega) \\ z &= a_3 \sin^{\varepsilon_1}(\eta) \ , \end{aligned} \tag{2}$$

where the angles η and ω represent the latitude and longitude angles in a spherical coordinate system, respectively. a_1, a_2, a_3 denote the superquadrics size along x, y and z axes, respectively. ε_1 and ε_2 represent the squareness parameters in the latitude and longitude plane, respectively. Superquadrics can represent various shapes by changing these parameters. Fig.2 shows the examples of shape representations. And more complex shapes can be realized by introducing deformed functions[8]. Fig.3 shows the deformed superquadrics along the z axis as examples.

By using superquadrics as the internal models, we can represent an approximate model of the object with a few parameters. Moreover, we can capture the feature of the object shape easily since parametric models such as superquadrics have their shape information as a set of parameters' values.



$(\varepsilon_1, \varepsilon_2) = (2.0, 2.0)$ $(\varepsilon_1, \varepsilon_2) = (0.1, 0.1)$ $(\varepsilon_1, \varepsilon_2) = (0.1, 1.0)$
$(a_1, a_2, a_3) = (10, 10, 10)$ $(a_1, a_2, a_3) = (10, 6, 4)$ $(a_1, a_2, a_3) = (2, 2, 10)$

Fig.2: Examples of superquadrics.

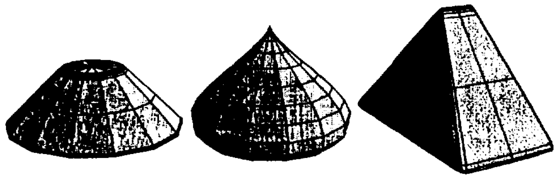


Fig.3: Examples of deformed superquadrics.
(Tapering deformation along axis z)

For shape recovery using visual data, we have to rewrite the equation (2) by eliminating the parameters η and ω as follows:

$$F\,(x,y,z) = \left[\left\{ \left(\frac{x}{a_1}\right)^{\frac{2}{\varepsilon_2}} + \left(\frac{y}{a_2}\right)^{\frac{2}{\varepsilon_2}} \right\}^{\frac{\varepsilon_2}{\varepsilon_1}} + \left(\frac{z}{a_3}\right)^{\frac{2}{\varepsilon_1}} \right]^{\varepsilon_1} = 1. \qquad (3)$$

This function is called the *inside-outside function*, and has the following characteristics:

- on the surface of the superquadric $\rightarrow F = 1$
- inside the superquadric $\rightarrow F < 1$
- outside the superquadric $\rightarrow F > 1$

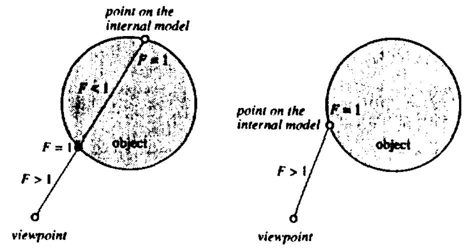
Therefore using these characteristics we can obtain a set of parameters which fairly fit the object shape by minimizing the following expression [7][8] :

$$\sum_{i=1}^{N} [1 - F\,(x_i, y_i, z_i\,;\, a_1, a_2, a_3, \varepsilon_1, \varepsilon_2)]^2, \qquad (4)$$

where N is the number of the data points obtained from the visual sensor. We should notice that the above equation corresponds to $\boldsymbol{E}_{vision}$. To minimize the above equation, several papers report the availability of the *Levenberg-Marquardt* method for nonlinear least squares minimization.

3.2.2 Extraction of the invisible portion

The internal model constructed by the above method is just what was recovered from the visible portion of the object. We have to get data of the invisible portion to obtain the internal model on which the exact shape of the object can be defined. Thus, we extract the invisible portion using a global internal model approximated by superquadrics, and then make the tactile sensor search there, if necessary. To extract the invisible portion, we use the inside-outside function expressed in equation (3). Fig.4 shows the proposed method for extracting the invisible portion. From the same figure, it is understood that the values of the inside-outside function between the viewpoint and the point on the visible portion of the internal model, *i.e.* superquadrics, are always larger than 1.0. On the other hand, comparing with the above case, the values between the viewpoint and the point on the invisible portion are different as shown in Fig.4(a). Using these characteristics of the inside-outside function, we can extract almost the whole portion of the invisible one.



(a) invisible point (b) visible point
Fig.4: A method of extracting invisible points.

3.2.3 Extraction of a portion with large errors

As mentioned before, superquadrics have a limitation in shape representation. For this reason, when we fit superquadrics to the complicated object which has an uneven surface and so on, the fitting between superquadrics and the object is not uniform. To overcome this problem, we must firstly extract the portion with large errors, and then deform it locally.

Thanks to the characteristics of the inside-outside function, we can easily extract the portion with large errors the same as the invisible portion. The extraction is carried out by comparing the values of the inside-outside function with 1.0 at every points on the internal model.

3.3 The second phase of the sensing process

3.3.1 A method of local deformation

To construct the internal model completely, it is necessary to operate local deformation to the obtained superquadrics at the first phase to reflect the details of the object on the internal model. Thus, we define the following local deforming energy for the points on the internal model:

$$\boldsymbol{E}_p = k_1' \boldsymbol{E}_{int} + k_2' \boldsymbol{E}_{ext} \;, \qquad (5)$$

where E_{int} represents the internal energy of the surface which gives physical constraints. Although there are various constraints, we use the following second-order continuity; curvature, for simplicity, based on the assumption that the surface of the object varies smoothly along the latitudinal and longitudinal directions s:

$$E_{int} = \left| \frac{d^2 v_{ij}}{ds^2} \right|^2 \approx | v_{i-1j} - 2v_{ij} + v_{i+1j} |^2 + | v_{ij-1} - 2v_{ij} + v_{ij+1} |^2, \tag{6}$$

where v_{ij} is the 3D vector which represents the coordinates of the points obtained by dividing the superquadric surface along the latitudinal and longitudinal directions (see Fig.5). As a rudimentary stage of the investigation, we assume that the data obtained by the tactile sensor is the coordinate of the touched point, and the coordinates of visual and tactile data can be converted into the same reference frame.

E_{ext} denotes the external energy which minimize the distance between the obtained tactile data and the corresponding points on the internal model, and is expressed as:

$$E_{ext} = | v_{ij} - data_{ij} | . \tag{7}$$

By minimizing these energy functions, we can deform the surface of the internal model previously constructed by the visual data as shown in Fig.5. To minimize the energy function efficiently, we use the *greedy algorithm*, evaluating the values of the energy function at the nodal points around the observed data points [9].

3.3.2 Correspondence between the obtained tactile data and points on the internal model

When the local deformation is performed, it is necessary to consider how to correspond the obtained data points to the nodal points on the surface of the internal model. We should notice that the global internal model has already been constructed parametrically using the visual data. Therefore, to realize the local deformation efficiently, we do not have to deform every points on the internal model, since most of the data points are close to the surface of the internal model.

From the above consideration, the correspondence between these points is proceeded only for the points of the invisible portion and/or with large errors. Fig.6 shows the proposed method of correspondence, schematically. This procedure is done by evaluating the angles between the vectors $\boldsymbol{p}$ and $\boldsymbol{d}$. When the angle is less than the prespecified threshold, we select this point as the one which corresponds to the data. But using this method, we often face the situation where one data point corresponds to numerous nodal points. In this case, to obtain smooth deformation, we evaluate the energy function expressed in equation (5) after multiplying the external energy E_{ext} by the weights which are inversely proportional to the distance between these nodal and data points.

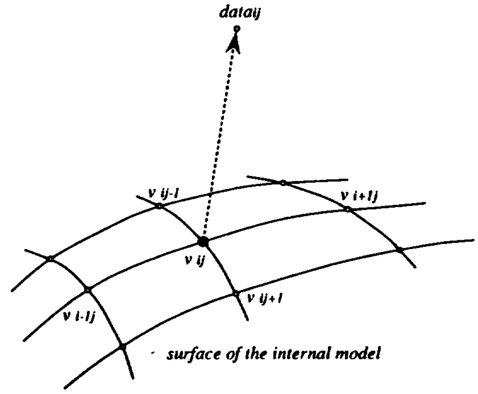


Fig.5: Local deformation.

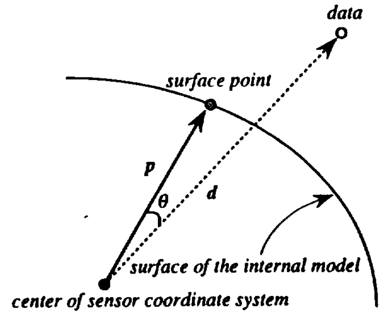
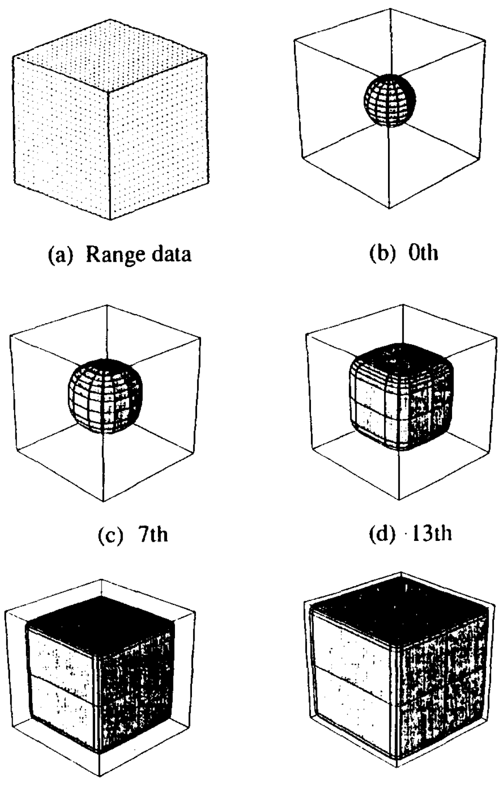


Fig.6: Correspondence between the data and the point on the internal model.

4. Simulation results

To confirm the ability of the global deformation of the internal model at the first phase, we fit superquadrics to cubes as an example using a visual sensor. We assume a range finder as a visual sensor which is located to obtain data of the three sides of the cube. Fig.7 shows the process of convergence of the internal model. From the figures, we can see that the internal model globally recovers the shape of the object using only the visible portion.

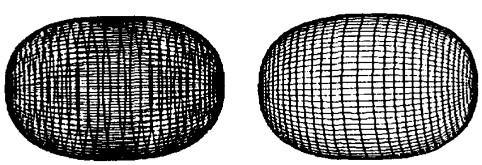


(a) Range data (b) 0th

(c) 7th (d) 13th

(e) 25th (f) 37th

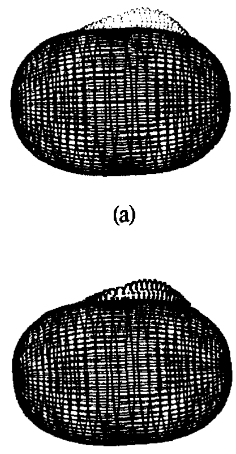
Fig.7: Shape recovery of a cube.



(a) whole part (b) invisible part

Fig.8: Result of extracting the invisible part.

Fig.8 shows the simulation results of extraction of an invisible portion by using the method mentioned in 3.2.2. The viewpoint is located at the front of this figure, and the surface is represented by latitudinal and longitudinal lines with five-degree-interval. Comparing (a) with (b), we can see that the invisible portion is extracted completely.



(a)

(b)

Fig.9: Result of extracting points with large errors.

Next, we investigate the ability of extracting the points with large errors using the uneven object as shown in Fig.9(a). The extracted points are denoted by solid discs as shown in Fig.9(b). From the figure, it can be seen that the extraction is carried out exactly.

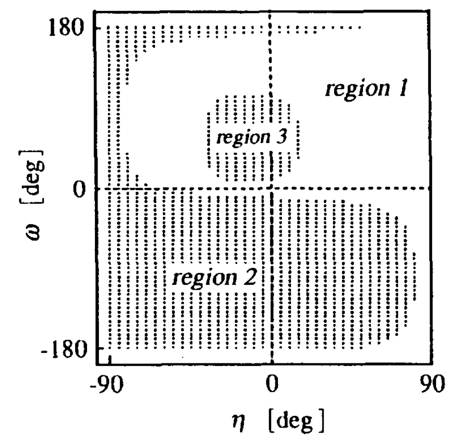
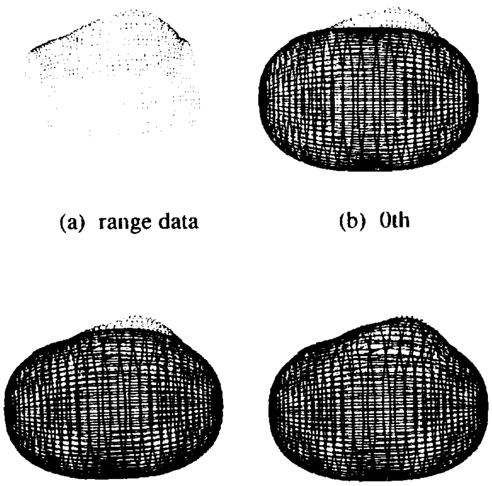


Fig.10: A map of points that are invisible or have large errors.

Fig.10 shows the map where the portions extracted in Fig.8 and 9 are located in η -ω plane. In the same figure, the region 1 represents the visible portion whose points are fitted to data. The regions 2 and 3 denote the portions extracted in Fig.8 and 9, respectively. Since this map is derived parametrically between the fusion of vision and touch, we can easily make demands for tactile sensing by referring to this map.



(a) range data (b) 0th

(c) 10th (d) 45th

Fig.11: Result of local deformation.

Fig 11 shows the fitting process of the local deformation with use of the same object as shown in Fig.9. In the figure, the numbers below the figures denote the iteration number of greedy algorithm. In this case, we use a simple shaped object, but this method is also available for more complicated objects.

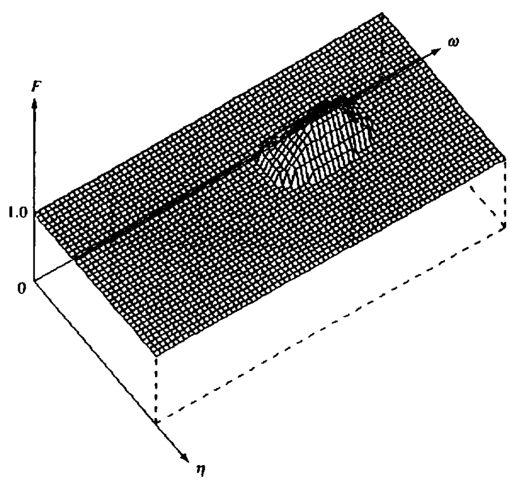


Fig.12: A graph of the relation among F, η and ω.

Fig.12 shows the relationship between the inside outside function and η-ω plane after fusing visual and tactile information. Comparing this relationship with the obtained parameters of superquadrics, we can capture the feature of the object shape.

5.Conclusions

In this paper, a method of fusing vision and touch by introducing energy function is investigated. By using the deformation of an internal model based on the characteristics of senses, we show that the internal model is constructed efficiently. Moreover, we show that this model is also available for recognition. And the feasibility of the proposed method is confirmed by simulations. Further research will consider the cooperative motion between vision and touch based on this internal model and the conversion from visual and tactile reference frame to the same frame in detail.

References

[1] R.C.Luo, et al.,"Multisensor Integration and Fusion in Intelligent Systems", IEEE Trans. on S.M.C., Vol.SMC-19. No.5, pp.901-931, 1989

[2] P.K.Allen,"Robotic Object Recognition Using Vision and Touch", Kluwer, 1987

[3] P.K.Allen,"Integrating Vision and Touch for Object Recognition Tasks", Int. J. Robotics Res., Vol.7, No.6, pp.15-33, 1988

[4] S.A.Stansfield,"A Robotic Perceptual System Utilizing Passive Vision and Active Touch", Int. J. Robotics Res., Vol.7, No.6, pp.138-161, 1988

[5] J.S.Albus,"Brains, Behavior, and Robotics", McGraw-Hill, 1981

[6] D.Terzopoulos, et al.,"Symmetry-seeking models for 3D object reconstruction", Int. J. Comput. Vision, Vol.1, No.3, 1988

[7] T.E.Boult, et al.,"Recovery of superquadrics from depth information", in Proc. Spatial Reasoning and Multi-Sensor Fusion Workshop, St.Charles, IL, pp.128-137, 1987

[8] F.Solina, et al.,"Recovery of Parametric Models from Range Images:The Case for Superquadrics with Global Deformations", IEEE Trans. on P.A.M.I., Vol.PAMI-12. No.2, pp.131-147, 1990

[9] D.J.Williams, et al.,"A Fast Algorithm for Active Contours and Curvature Estimation", CVGIP:IMAGE UNDERSTANDING, Vol.55, No.1, pp.14-26, 1992

[10] M.Kass, et al.,"Snakes:Active Contour Models", International Journal of Computer Vision, Vol.1, No.4, pp.321-331, 1980