REPRESENT AND ACQUIRE KNOWLEDGE FOR THE DEVELOPMENT OF AUTONOMOUS VISION SYSTEM *

Xiaobu Yuan

Dept. of Computer Science, Memorial University
St. John's, Newfoundland, Canada A1C 5S7

ABSTRACT

The visual perceptional ability of computer systems to understand the environment is desirable in engineering design and manufacturing where automation is anticipated. In order to empower an engineering system with the heuristic capabilities, adequate knowledge representation techniques must be employed to resolve ambiguity in representation and uncertainty in decision making. Discussed in this paper is how to represent environmental knowledge as object models so that they can be utilized in spatial reasoning, and how to check knowledge completeness so that autonomous vision systems can be developed with self-regulated knowledge acquiring mechanism.

1 INTRODUCTION

Since the invention of the first computer, continuous effort has been devoted from both the academic and industrial communities to improve computer's capability. One of the long dreamed goals is the ability of computer to understand the 3D environment by means of sensory devices[1], i.e, visual perception that simulates human vision[2]. When furnished with the capability, manufacturing and engineering systems could obtain the knowledge of the environment in which they are working, and further use the knowledge in decision-making activities, such as automated machinery assembly and other factory automation applications.

Besides the diligent work of obtaining 3D information from 2D images or directly from sensory devices, an autonomous vision system demands for: *(a)* powerful processing techniques to extract spatial features from images in practical environment[3]; *(b)* extensive graphics methods, including modeling methods to provide transformation invariant representation and interfacing facilities[4]; *(c)* efficient algorithms of object recognition that are free of mismatching and work well even in ambiguous situations[5]; and finally *(d)* artificial intelligence to automatically acquire environmental knowledge and to use the knowledge in various spatial reasoning activities[6, 7].

This paper discusses how to represent 3D objects as spatial domain knowledge, and how to automatically acquire the knowledge and further utilize it to make decisions. View sensitivity of spatial knowledge representation (or better known in AI as *heuristic inadequacy*) is resolved by modeling objects with *localized surface parameters*. Automatic knowledge acquisition is then achieved by checking the *mass vector chain* of a model to determine if the knowledge is complete or alternatively to estimate the direction from which missing information can be obtained.

2 MOTIVATION

Knowledge representation is a descriptive notation that follows certain syntactic and semantic rules and describes those that an intelligent system needs to know. While the computational tractability requires the representation to incorporate and manipulate efficiently within a computer system, the requirements of metaphysical and epistemic adequacy demand for ambiguity-free, clear,

*This work is supported by the National Science and Engineering Council of Canada under Grant OGP0155411.

uniform, convenient, domain-relevant, and declarative representations[8].

In spatial domain, various methods have been used to represent solid objects[4]. One of the most attractive methods is the boundary representation of objects. In the method, spatial features are modeled by a function $f(\mathbf{x}, \mathbf{p}) = \mathbf{0}$, where $\mathbf{x} \in \Re^3$ is a vector of coordinates, and $\mathbf{p} \in \Re^n$ is a parameter vector[9]. When $f(\mathbf{x}, \mathbf{p}) = \mathbf{0}$ is a family of surfaces, a surface in space is then a specific instance of $f(\mathbf{x}, \mathbf{p}) = \mathbf{0}$ whose parameters take a particular value $\mathbf{p}_i$. Especially, if $f(\mathbf{x}, \mathbf{p}) = \mathbf{0}$ stands for the family of all quadratic surfaces as in the application of surface approximation[10], $\mathbf{p}$ is the vector,

$$\langle a_{11}\ a_{12}\ a_{13}\ a_{21}\ a_{22}\ a_{33}\ a_{14}\ a_{24}\ a_{34}\ a_{44} \rangle \quad (1)$$

whose elements are the coefficients of the following equation.

$$a_{11}x^2 + 2a_{12}xy + 2a_{13}xz + a_{21}y^2 + 2a_{22}yz + 2a_{33}z^2 + 2a_{14}x + 2a_{24}y + 2a_{34}z + a_{44} = 0$$

An individual surface is therefore a vector $\mathbf{p}_i$ with a particular value, e.g., the parameter vector for a planar circle $x^2 + y^2 = 4$ is $\mathbf{p}_i = \langle 1\,1\,0\,0\,0\,0\,0\,0\,0\,4 \rangle$.

As a result, the spatial domain knowledge is the description of 3D objects. It contains a group of symbolically represented object models. Suppose there are totally N objects in the space, and an object O_i has m_i pieces of boundary surfaces $S_i(j)$, $0 \leq i \leq N-1$, $0 \leq j \leq m_i - 1$. With each surface $S_i(j)$ described as a parameter vector, the knowledge of an object O_i is a model represented by a list of such variables. Spatial domain knowledge represented in such a method is metaphysically and epistemically adequate as well as computationally tractable when only static representation is concerned.

Nevertheless, the representation is not heuristic adequate[11] as it does not express the reasoning that goes through in solving a problem. It is because the parameter vector of a spatial feature $f(\mathbf{x}, \mathbf{p}) = \mathbf{0}$ depends on the coordinate system under which coordinates are defined. When a transformation with six-degree freedom takes place, as in the application of manufacturing, the description of boundary surfaces and objects changes correspondingly. Therefore, such models can not be used as system's knowledge of the environment since a same object may have different forms of representation, a fatal ambiguity of knowledge representation which is regarded as view sensitivity in computer vision. For instance, the parameter vector of the planar circle changes to another vector $\langle 0.89\ 0.25\ -0.33\ 0.89\ -0.33\ 0.13\ 0.0\ 0.0\ 0.0\ 4.0 \rangle$ if it is described under a new coordinate system after a rotation for $\pi/6$ around a vector $\langle 0.707\ 0.707\ 0 \rangle$. Although many methods had been proposed to recover the original description, the parameter vectors are still hardly usable as environmental knowledge.

In addition to the problem of using knowledge to understand the environment caused by view sensitive descriptions, the surface list of a model provides no information on knowledge completeness, i.e., if system's knowledge of an object is complete or how to make it complete so that environmental knowledge can be automatically acquired. It therefore becomes imperative to modify the representation of environmental knowledge.

3 SPATIAL KNOWLEDGE

View sensitivity is caused by the parameter vector $\mathbf{p}_i$ that changes along with the coordinate system under which function $f(\mathbf{x}, \mathbf{p})$ is defined. As differential geometry has been successfully used in image segmentation because view independent properties capture the spatial properties of surface shapes[10], when view insensitivity of object representation is concerned, the coordinate systems chosen to define object surfaces should come from the surface characteristics that relate with surface shapes only.

Surface centroid is the average of all surface points. It has been proven[12] to be independent of descriptive coordinate systems. In addition, according to differential geometry, there are also other transformation invariant local spatial characteristics at the surface centroid, such as its maximum and minimum curvature, and its Euclidean distance to any point on the surface. When a coordinate system is set up locally on each surface so that its origin coincides with the centroid, the **z** axis points to the average normal direction, the **x** axis to a direction determined by a local surface characteristic, and the **y** axis makes the system a right hand Cartesian system with the other two axes, the local coordinate system on the surface is independent of viewing directions. Furthermore, when described in such a coordinate system, the representation of the surface inherits the view insensitivity. As for the topological relationship between boundary surfaces, it can be specified by a set of homogeneous transformation matrices that relate local coordinate systems to an object coordinate system[13].

For a surface $S_i(j)$ on object $O_i(j)$, $0 \le i \le N-1$, $0 \le j \le m_i - 1$, let its localized surface description and transformation matrix be $\mathsf{s}_i(j)$ and $\tau_i(j)$ respectively. By representing an environmental object $O_i(j)$ with a model which is a list of $\langle \mathsf{s}_i(j), \tau_i(j)\rangle$ pairs, the environmental knowledge $\mathcal{M}$ is the collection of all the object models.

$$\mathcal{M} = \{M_i \mid M_i = (\langle \tau_i(0), \mathsf{s}_i(0)\rangle, \cdots, \langle \tau_i(m_i-1), \mathsf{s}_i(m_i-1)\rangle)\} \quad (3)$$

Since surfaces are given in local coordinate systems, the parameter vector for each boundary surface remains the same independently of descriptive coordinate systems. As the spatial structure of boundary surfaces is related by homogeneous transformation matrices that refer local systems to an object coordinate system, their relations are also transformation invariant. In such a way, environmental knowledge is changed to view insensitive lists, each of which stands for an object in the space, and whose elements are localized surface parameters composed of a local parameter vector and a homogeneous matrix.

4 KNOWLEDGE COMPLETENESS

View insensitivity is an obvious advantage of using localized surface parameters in visual perception. More important is that the representation embodies the knowledge of object closure and surface direction. While the knowledge of environmental objects are obtained by extracting surface parameters from sensory inputs, it becomes particularly useful for visual system to judge on existing knowledge so that consistent and complete object models can be accumulated from different directions[14, 15].

4.1 Mass Vector Chain

For a surface $S_i(j)$, a mass vector $P_{\mathbf{n}}(\mathsf{s}_i(j))$ is derivable from the pair $\langle \mathsf{s}_i(j), \tau_i(j)\rangle$. It is the average surface normal $\mathbf{n}_i(j)$ weighted by the projection area $R_i(j)$ of $\mathsf{s}_j(i)$ onto the **x-y** plane of its local coordinate system, i.e., $P_{\mathbf{n}}(\mathsf{s}_i(j)) = \mathbf{n}_i(j)\, R_i(j)$. Since the sum of all the mass vectors for a closed object model is a zero vector[16], if the mass vector sum of an object equals to a nonzero vector $\mathbf{V}_{gap}$, there must be some surfaces still unprocessed and the total of their mass vectors is the negative of $\mathbf{V}_{gap}$. Suppose the number for processed surface patches is m'_i, then, from equation

$$\sum_{j=0}^{m'_i-1} P_{\mathbf{n}}(\mathsf{s}_i(j)) + \sum_{j=m'_i}^{m_i-1} P_{\mathbf{n}}(\mathsf{s}_i(j)) = \mathbf{0} \quad (4)$$

it is easy to get the result.

$$\sum_{j=m'_i}^{m_i-1} P_{\mathbf{n}}(\mathsf{s}_i(j)) = -\sum_{j=0}^{m'_i-1} P_{\mathbf{n}}(\mathsf{s}_i(j)) = -\mathbf{V}_{gap} \quad (5)$$

Being defined to the average normal vector $\mathbf{n}_i(j)$, each mass vector $P_{\mathbf{n}}(\mathsf{s}_i(j))$ is actually the average visible direction of that surface. Therefore, $\mathbf{V}_{gap}$ provides an estimated direction from which these unprocessed surfaces could be observed.

4.2 The Two-Sphere Method

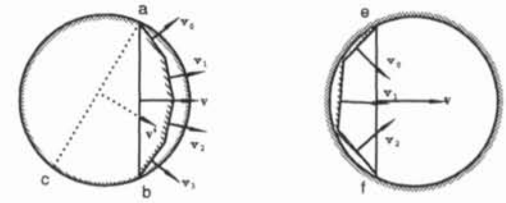

Figure 1: View Estimation

This observation can be used in automatic knowledge acquisition. Consider the two spheres in Fig. 1 as two Gaussian spheres, one for objects with convex surfaces and the other with concave surfaces. As a convex object has a unique Gaussian image[17], a Gaussian sphere covering the object stands for its boundary condition. It is obviously that, if the total mass vector of obtained surfaces is $\mathbf{v}$(Fig. 1(a)), the average visible direction for the unprocessed big arc from b to a is in the opposite direction of $\mathbf{v}$. Even if more surface patches are processed after several views, the negative of the updated total mass vector, which is $\mathbf{v}'$ in the figure, still points to the visible direction of unprocessed portion.

For concave surfaces, a similar Gaussian sphere can also be used to examine the boundary condition for the concave part where only concave surfaces are concerned. Concave surfaces usually do not form a closed mass vector chain by themselves. When a virtual surface is introduced to represent the cutting surface, however, the total mass vector for the concave surfaces plus the virtual surface must be zero since they form a close surface. Suppose there are m_h concave patches in a hole, the following relation exits

$$\sum_{j=0}^{m_h-1} P_{\mathbf{n}}(\mathsf{s}_i(j)) + \mathbf{V}_i(hole) = \mathbf{0} \quad (6)$$

where $\mathbf{V}_i(hole)$ is the mass vector for the virtual surface patch. As a result, an equation similar to Eq. 5 is obtained to estimate the visible direction for unprocessed concave surfaces.

$$\begin{aligned}
&\sum_{j=m'_h}^{m_h-1} P_{\mathbf{n}}(\mathsf{s}_i(j)) \\
&= -(\sum_{j=0}^{m'_h-1} P_{\mathbf{n}}(\mathsf{s}_i(j)) + \mathbf{V}_i(hole)) \\
&= -\mathbf{V}_{hgap} \qquad (7)
\end{aligned}$$

This equation also applies to objects with more than one open hole where more virtual mass vectors, instead of one, are added to the equation to count in all cutting surfaces.

When an object has both convex and concave surfaces, however, the two spheres have to be combined together to investigate the knowledge completeness of boundary surfaces. Since the two types of surfaces form a boundary together, the virtual surface represented by $\mathbf{V}_i(hole)$ in Eq. 7 can be dealt as a virtual surface that makes the object boundary close without the concave surfaces. Therefore, after taking off the virtual surface from Eq. 5, and taking off $\mathbf{V}_i(hole)$ from Eq. 7, the sum of mass vectors for the object forms a relation that view estimation can be made for objects with both convex and concave surfaces in a way similar to simple objects.

$$\sum_{j=0}^{m_i-1} P_{\mathbf{n}}(\mathsf{s}_i(j)) + \sum_{j=0}^{m_h-1} P_{\mathbf{n}}(\mathsf{s}_i(j)) = \mathbf{0} \quad (8)$$

In the equation, all local surfaces $\mathsf{s}_i(j)$ belong to a list $(\langle\tau_i(0), \mathsf{s}_i(0)\rangle, \cdots, \langle\tau_i(m_i-1), \mathsf{s}_i(m_i-1)\rangle)$ in Eq. 3.

It can be concluded from the above discussion that the knowledge of an environmental object is complete only if the mass vector chain of its model list is closed; otherwise, the vector that connects the head to the tail of the chain estimates the visible direction of the missing surfaces. Therefore, the chain of mass vectors of all the obtained boundary surfaces of an object can be used to check the spatial closure of the object and to predict the directions of unprocessed surfaces.

5 THE USE AND ACQUISITION OF KNOWLEDGE

Since no view sensitivity is involved, object recognition based on knowledge representation in Eq. 3 is direct. It is carried out in three steps. Firstly, localized parameter vectors of the examined object are compared directly with those of acquired models. A model is considered as a candidate for the object only if it has surfaces matching with any of the object surfaces. Secondly, those matched model surfaces must have the same topological relations as their object surfaces. Finally, unmatched model surfaces should not be visible if the model is rotated to the orientation of the object[18].

As for the mechanism of automatic knowledge acquisition, its usage in engineering design and manufacturing systems is obvious

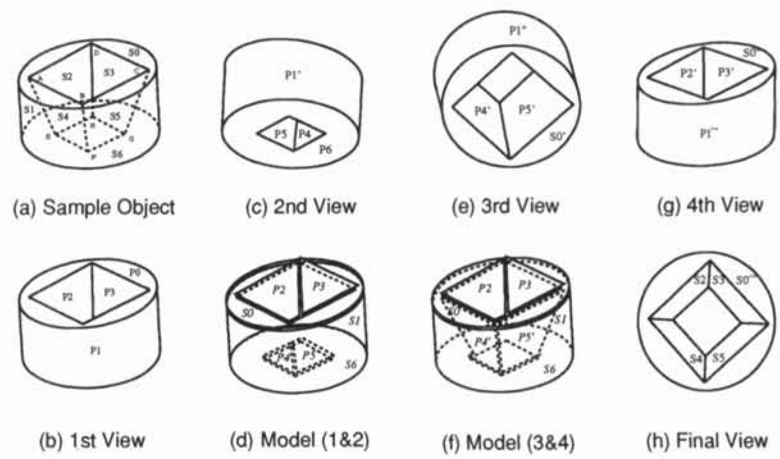

Figure 2: Automatic Knowledge Acquisition

because it enables the systems to learn their working environment by themselves. Its operation can be explained[19] with a sample object(Fig. 2(a)), which could be a machinery part in automatic assembly.

First of all, spatial features are extracted from sensory inputs (e.g., a range image) taken at a random viewing direction. After defining local coordinate systems for all the extracted surfaces, a partial model of the object is then established as a list of localized surface parameters(Fig. 2(b)). Since all surfaces of a convex object should be visible from two opposite directions, the next viewing direction is the direction opposite to the first random direction(Fig. 2(c)), instead of a direction determined by mass vector chains. At this time, the partial model of the object is updated with new information(Fig. 2(d)), during which identical surface descriptions play an important role in finding a surface should or should not be merged with other surfaces.

Following it, the mass vectors of all the surfaces in the partial model are chained together. If the mass vector chain is closed, i.e., the object is closed in space, transformation matrices are finalized with an object coordinate system, and the model is saved as system's knowledge of the environment. Otherwise, the system repositions the sensory device to a direction determined by the mass vector chain, and repeats the previous process to update the object model(Fig. 2(e) and (f)). Again, surface merging is necessary as different parts of same surfaces may be extracted repeatedly during model updating.

View adjustment is necessary when the estimated direction of unprocessed surfaces is blocked by other surfaces. If an updated partial model has little or no difference with the old one(Fig. 2(g)), the next view is determined[13] by the group of blocked surfaces(Fig. 2(h)), rather than the mass vector chain. After the acquisition, the model of the object is the knowledge of the environment. It can then be used to identify objects even in ambiguous situations[18]. In such a way, an engineering system is able to learn its environment and uses what it has learned to perform activities that require visual perceptual capabilities.

6 CONCLUSION

Discussed in the paper is a representation method of spatial domain knowledge, which uses localized surface parameters to describe environmental objects. Because of its view insensitivity, descriptions of objects can be used as environmental knowledge in object recognition and other visual perceptual activities. In addition, the ability of automatic knowledge acquisition based on mass vector chains enables a computer system to learn its environment. The presented technique is useful in autonomous robot for an unknown environment as well as factory automation.

References

[1] Vishvjit S. Nalwa. *A Guided Tour of Computer Vision.* Addison Wesley, 1993.

[2] James D. McCafferty. *Human and Machine Vision: computing perceptual organization.* Ellis Horwood, 1990.

[3] R. Jain and A. K. Jain. Report: 1988 NSF range image understanding workshop. In *Analysis and Interpretation oj Range Images*, chapter 1, pages 1–32. Springer-Verlag, N.Y., 1990.

[4] James Foley and et al. *Computer Graphics: Principles and Practice.* Addison Wesley, second edition, 1992.

[5] Paul J. Besl. The free-form surface matching problem. In *Machine Vision for Three-Dimensional Scenes*, pages 25–72. Academic Press, Inc., 1990.

[6] S. M. Weiss and C. A. Kulikowski. *Computer Systems That Learn.* Morgan Kaufmann Publishers, Inc., 1991.

[7] Nigel Ford. *Expert Systems and Artificial Intelligence.* Library Associate Publishing, 1991.

[8] J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence 4*, pages 463–502, 1969.

[9] James L. Crowley and Fano Ramparany. Mathematical tools for representing uncertainty in perception. In *Spatial Reasoning and Multi-sensor Fusion: Proceeding of the 1987 Workshop*, pages 293–302. Academic Press, New York, 1987.

[10] Paul. J. Besl. *Surfaces in Range Image Understanding.* Springer-Verlap, New York, 1988.

[11] T.J.M. Bench-Capon. *Knowledge Representation: An Approach to Artificial Intelligence.* Academic Press, 1990.

[12] Xiaobu Yuan. LSP: A view insensitive 3D representation method. In *Proceedings of Vision Interface '93*, pages 64–69, Toronto, Ontario, 1993.

[13] Xiaobu Yuan. *3D Reconstruction as An Automatic Modeling System.* Ph.D thesis, Department of Computing Science, University of Alberta, Canada, 1992.

[14] M. Magee and M. Nathon. Spatial reasoning, sensor repositioning and disambiguation in 3D model based recognition. In *Spatial Reasoning and Multisensor Fusion: Proceeding of the 1987 Workshop*, pages 262–271. Academic Press, 1987.

[15] Hugh F. Durrant-Whyte. Sensor models and multi-sensor integration. In *Spatial Reasoning and Multi-sensor Fusion: Proceeding of the 1987 Workshop*, pages 303–312. Academic Press., New York, 1987.

[16] Xiaobu Yuan. Mass vector controlled view estimation for automatic 3D object modeling. In *Proceedings of Vision Interface '93*, pages 177–182, Toronto, Ontario, 1993.

[17] B. K. Horn. Extended Gaussian images. *Proc. IEEE*, 72:1656–1678, December 1984.

[18] Xiaobu Yuan. Object recognition based on localized surface parameters. In *Proceedings of the AMSE International Conference on Information Processing Methodologies and Applications*, pages 71–80, Orlando, Florida, 1993.

[19] Xiaobu Yuan. Auto*model*: An expert system of object modeling and recognition. In *Proceedings of the AMSE International Conference on Information Processing Methodologies and Applications*, pages 81–91, Orlando, Florida, 1993.