Modeling Sensor Detectability and Reliability
in the Configuration Space
for Model-Based Vision

Katsushi Ikeuchi, Takeo Kanade
July 1987
CMU-CS-87-144ᶫ

ABSTRACT

The model-based vision requires object appearances in the computer. How an object appears in the image is a result of interaction between the object properties and the sensor characteristics. Thus, in model-based vision, we ought to model the sensor as well as modeling the object. In the past, however, the sensor model was not used in the model-based vision or, at least, was contained in the object model implicity.

This paper presents a framework between an object model and the object appearances. We consider two aspects of sensor characteristics: sensor detectability and sensor reliability. Sensor detectability specifies what kind of features can be detected and in what area the features are detected; sensor reliability specifies how reliable detected features are. Commonly available sensors are briefly examined in terms of their sensor characteristics. We define the configuration space to represent sensor characteristics. We propose a representation method of the sensor dectectability in the configuration space. Sensor reliability distribution is also discussed in the configuration space. Under this framework, we characterize the photometric stereo and the lightstripe range finder as examples.

Table of Contents

1. INTRODUCTION

The model-based vision requires object models in the computer. Various researchers propose many kinds of object models, ranging from generic models such as generalized cylinders [5, 31, 9, 41], extended Gaussian images [42, 23, 21], and super quadric models [38] to specific models such as aspect model [28, 13, 24], region-relation model [4, 37, 6], and smooth local symmetry [7, 8].

The object appearances, however, are determined by a *product* of an object model with a sensor model. Thus, in the model based vision, it is insufficient to consider only an object model; it is essential to exploit a sensor model as well. On the other hand, modeling sensors for model-based vision has attracted little attention; quite often, researchers who are familiar with the sensors they use tended to construct object appearances by implicitly incorporating their sensor behavior. This paper, in contrast, explores a general framework for explicitly incorporating sensor models which govern the relationship between object models and object appearances.

A sensor model must be able to specify two important characteristics: sensor detectability and sensor reliability. The sensor detectability specifies what kind of features can be detected and in what condition the features are detected. The sensor reliability specifies how reliable the detected features are. This paper, thus, present a method for modeling sensors with sensor detectability and sensor reliability. Commonly available sensors are briefly examined in terms of their sensor characteristics. Then, representation techniques for sensor characteristics are explored. We choose the projected Quaternion space using the configuration space to represent sensor characteristics. Finally, we consider two aspects of sensor characteristics: sensor detectability and sensor reliability. We propose a representation space on which a sensor's detectability is expressed in the uniform way. Sensor reliability analysis consists of reliability distribution and error propagation from observed data to geometric features. Under this framework, we characterize the photometric stereo and the light-stripe range finder as examples.

2. SENSORS IN THE MODEL BASED VISION

This section gives a brief survey of commonly available sensors in the model-based vision. We include both passive and active sensors. The following sensors are often used: edge detector [40, 30, 11], shape-from-shading [20, 26], binocular stereo [32, 16, 3, 36], time-of-flight range finder [27, 18], light-stripe range finder [1, 37], trinocular stereo [35], photometric

stereo [44, 25], polarimetric light detector [29],and SAR (Synthetic Aperture Radar) [14, 43, 34].

Each sensor is a mapping function from object features to sensor features. Object features such as faces, edges, and vertices exist in the three-dimensional object space. These object features in the object space are mapped to sensor features such as regions, lines, and points in the sensor space. Then, these sensor features are transformed into geometric features. Note that the sensor features has no meaningful interpretation. For example, a sensor feature such as a region should be considered as a collection of points which are not grouped as one meaningful unit by a sensor. The conversion process from sensor features to geometric features will group a collection of points into a geometric feature such as a region.

Table 1 gives the summary of detectable features in the object space by commonly available sensors. For example, an edge detector can detect some edges in the object space as lines in the sensor space. Since it does not need special light sources, it is classified as a passive sensor.

Table 1 Detectable Object Features				
Sensor	Vertex	Edge	Face	active/passive
Edge Detector	no	yes	no	passive
Shape-from-shading	no	no	yes	passive
SAR	yes	yes	yes	active
Time-of-Flight Range Finder	no	no	yes	active
Light-stripe Range Finder	no	no	yes	active
Binocular Stereo	no	yes	no	passive
Trinocular Stereo	no	yes	no	passive
Photometric Stereo	no	no	yes	active
Polarimetric light detector	no	no	yes	active

Since the brightness change does not occur over any face, passive sensors have usually difficulty in detecting faces. An active sensor projects lights over the scene; stronger reflection can be obtained from faces; weaker or no reflection is given from edges and vertices. Thus, most active sensors are good in detecting faces, while they can detect neither edges nor vertices.

Sensor features in the sensor space are summarized in Table 2. In Table 2, a line means a line-shaped collection of detected points and a region means a region-shaped collection of detected points. For all sensors except SAR the correspondence between the object feature and

the sensor feature is one-to-one. For example, an edge detector generates one line-shaped sensor feature corresponding to either one edge or one reflectance discontinuity line of a object feature in the object space. Photometric stereo generates a surface orientation distribution as sensor features which corresponds to one physical face in the object space. On the other hand, SAR generates either line-shaped sensor feature or point sensor feature from one edge in the object space depending on the sensor configuration. A precise discussion of SAR will be found elsewhere [14, 43, 34].

Table 2 Detected Sensor Features			
Sensor	Vertex	Edge	Face
Edge Detector	-	line	-
Shape-from-shading	-	-	region
SAR	point	point/line	line
Time-of-Flight Range Finder	-	-	region
Light-stripe Range Finder	-	-	region
Binocular Stereo	-	line	-
Trinocular Stereo	-	line	-
Photometric Stereo	-	-	region
Polarimetric light Detector	-	-	point

While this summary tells in general what object features are detectable in what forms by various sensors, it is also important to characterize in what viewing conditions those features are actually detectable and how reliable the detected features are. For that, we need to develop a representation tool for relation between object coordinates and sensor coordinates.

3. REPRESENTING SENSOR CONFIGURATION

This section defines the sensor configuration space on which sensor detectability and sensor reliability are specified. Sensor detectability and reliability depend on various factors: distance of an object, configuration of an object, reflectivity of an object, transparency of air, and background noise such as the sun's brightness. In the model-based vision, since the target object and its rough distance is a prior known, mainly angular freedom of object affects on detectability and reliability. Thus, we will explore a space to specify the relationship between the sensor coordinate and the object coordinate.

The relationship between the sensor coordinate and the object coordinate can be specified by three degrees of freedoms; two degrees of freedom in the sensor direction and one degree of freedom in the sensor rotation. The representation space must satisfy the following two criteria.

Intuitiveness   The representation space should be intuitive because it is a tool to depict detectability and reliability; sensor users should understand them easily.

Continuity   The representation space should be continuous.

In the following discussion, the sensor coordinate is fixed; we will explore the way to specify the object coordinate with respect to the sensor coordinate.

We will examine four different ways to represent these freedoms from these two criteria.

Euler space   to represent an Euler angle on the sphere

Quaternion space   to use quaternion to specify the relationship and represent it on the four dimensional hyperspherical surface.

Projected quaternion space

to project the four dimensional hyperspherical surface to a three dimensional sphere and represent the relationship on the projected sphere.

Modified projected quaternion space

to map the projected three dimensional sphere to another three dimensional sphere and to represent the relationship on the mapped sphere.

These representations are used to represent relative relations between the sensor coordinate and the object coordinate, and therefore, for the sake of convenience in the following discussion, we fix the sensor coordinate and discuss how to specify the object coordinate with respect to the sensor coordinate.

3.1. Euler Space

Use of an euler angle, (ϕ,θ,ψ), is one of the standard ways to specify the relationship from the sensor to the object coordinate: ϕ,θ, and ψ denote rotation around the z axis, rotation around the new y axis, and rotation around the new z axis, respectively.

The Euler angle can be represented as a point in the sphere. ϕ and θ can be equated to the the azimuth and the zenith angles to denote the direction of the point from the center. The distance of the point from the spherical surface depicts the rotation of the new coordinates around the new z axis, ψ. In this space, the north pole, for which (ϕ θ ψ = (0 0 0)), can correspond to the sensor coordinate and any point in the sphere corresponds to a particular object coordinate.

This space satisfies the intuitiveness criterion, but does not satisfy the continuity criterion.

Representation become discontinuous near the north pole. Unfortunately, points near the north pole corresponds to frequently used coordinates, because the sensor coordinate is expressed as the north pole.
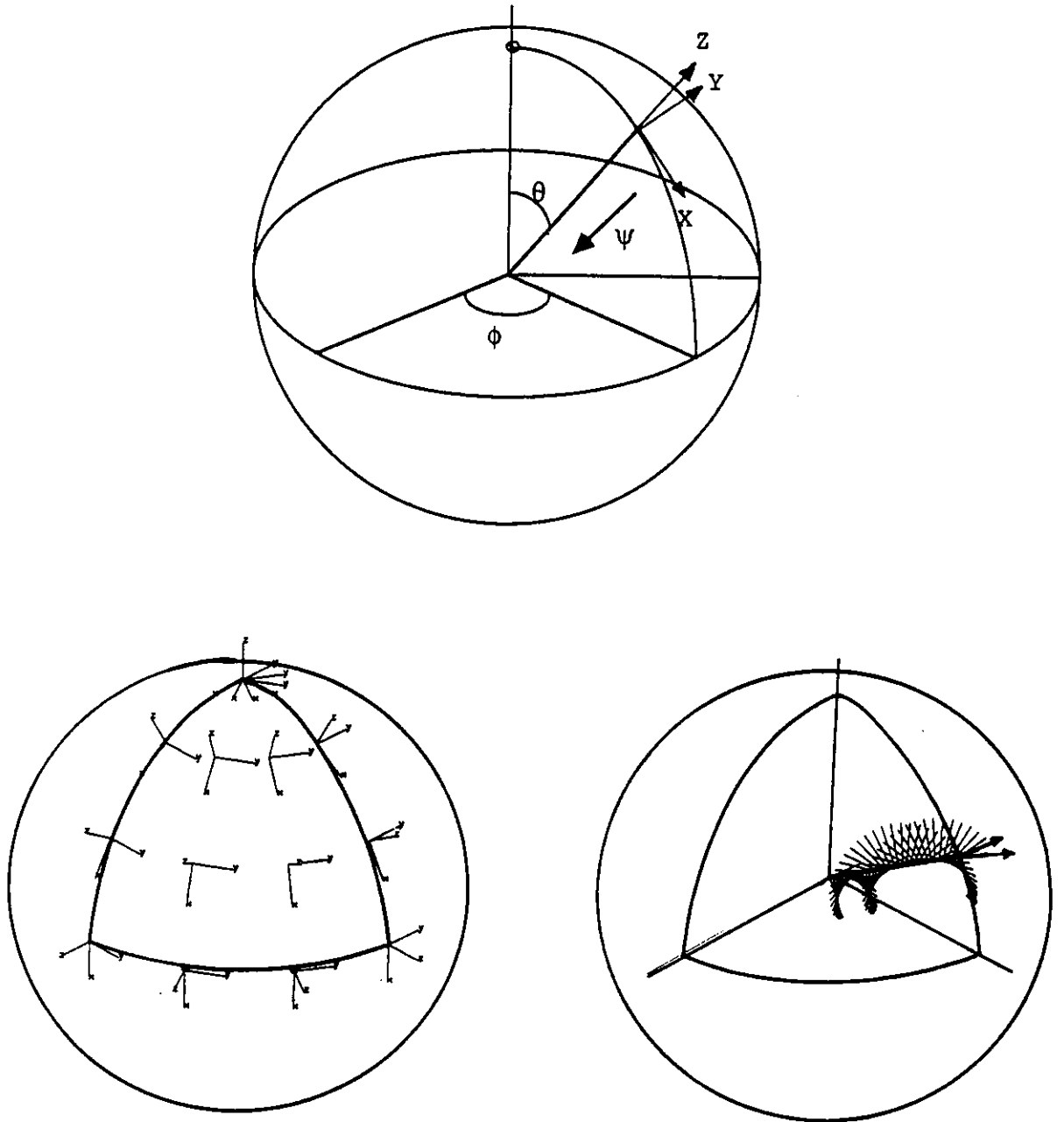

Figure 1 Euler space

3.2. Quaternion Space

Our second candidate is the four dimensional hypersphere which comes from quaternions [17]. Other researchers [10, 39, 12, 22] have also explored the use of quaternions for specifying rotations in computer vision and robotics research.

Quaternions can be represented by a quadruplet of four real numbers,

$$(q_0, q_1, q_2, q_3) = q_0 + q_1 i + q_2 j + q_3 k.$$

By using quaternions, rotation can be represented as

$$q_0 = \cos\left(\frac{\omega}{2}\right)$$

$$q_1 = \sin\left(\frac{\omega}{2}\right) n_x$$

$$q_2 = \sin\left(\frac{\omega}{2}\right) n_y$$

$$q_3 = \sin\left(\frac{\omega}{2}\right) n_z,$$

where ω denotes rotation angle and (n_x, n_y, n_z) denotes the axis of the rotation. Since $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$, any rotation can be represented as a point on the four dimensional hyperspherical surface. In this space, the sensor coordinate corresponds to the north pole of the four dimensional hypersphere, $(1,0,0,0)$. Any point on the upper hyper-hemisphere corresponds to one particular object coordinate. In contrast to Euler space, this space satisfies the continuity criterion, but does not satisfy the intuitiveness criterion because we cannot draw a four dimensional hypersphere.

3.3. Projected Quaternion Space

We can project the four dimensional hyperspherical surface to a three dimensional sphere (four dimensional hyperplane). Before considering the projection, let us examine the two and three dimensional case to draw analogy. A two dimensional spherical surface (circle) can be denoted as

$$y = \cos \alpha$$
$$x = \sin \alpha,$$

where α is the rotation angle. This two dimensional spherical surface (circle) can be projected to a one dimensional line as

$$y = 1$$
$$x = \sin \alpha.$$

Similarly, a three dimensional spherical surface can be denoted as

$$z=\cos\alpha$$
$$y=\sin\alpha\sin\beta$$
$$x=\sin\alpha\cos\beta,$$

where α,β are the zenith and azimuth angles, respectively, and when we project it to a two dimensional plane, we obtain a disc:

$$z=1$$
$$y=\sin\alpha\sin\beta$$
$$x=\sin\alpha\cos\beta.$$

We can use the similar technique for a four dimensional hyperspherical surface. It can be denoted as

$$u=\cos\alpha/2$$
$$z=\sin\alpha/2\cos\beta$$
$$y=\sin\alpha/2\sin\beta\sin\gamma$$
$$x=\sin\alpha/2\sin\beta\cos\gamma,$$

where (u,z,y,x) denotes the quaternion. Following the same way as in the two and the three dimensional cases, we have

$$u=1$$
$$z=\sin\alpha/2\cos\beta$$
$$y=\sin\alpha/2\sin\beta\sin\gamma$$
$$x=\sin\alpha/2\sin\beta\cos\gamma,$$

where α denotes the rotation angle and $(\cos\beta,\sin\beta\sin\gamma,\sin\beta\cos\gamma)$ denotes the direction of the rotation axis. This gives the projection from a four dimensional hyperspherical surface onto a three dimensional sphere $x^2+y^2+z^2 \leq 1$.

The origin of the sphere corresponds to the sensor coordinate, and any point in the sphere corresponds to one particular object coordinate. See Figure 2a. Note that this only projects either the points of upper hyper-hemisphere or the lower hyper-hemisphere to the points of the sphere. However, we only need the upper hyper-hemisphere to represent all possible object coordinates. Let us call cylinders whose axis agree with the axis between the north pole and the south pole as characteristic cylinders. The projected quaternion space has the following relationships with the Euler angle.

Property 1 Points on a characteristic cylinder surface correspond to object coordinates whose zenith angles are the same.

Property 2 A point on the characteristic cylinder rotates on the cylindrical surface depending on the difference between ψ and ϕ.

Property 3 All object coordinates whose corresponding points exist on the intersection line between the characteristic cylinder and the x-y plane satisfy $\phi+\psi=0$ or $\phi+\psi=\pi$.

Property 4 A point on the intersection circle between the characteristic cylinder and the sphere surface corresponds to the same object coordinate as the symmetry point with respect to the sphere center. In other words, the characteristic cylinder connects to its self in a twisted manner.

This space satisfies both the continuity criterion and the intuitiveness criterion. Unfortunately, however, this space has a problem which we refer as the twisted cylinder problem. In this space, coordinates whose z axis make a certain angle with one particular direction are projected as an ellipsoid on a plane parallel to the x-y plane. Then center of the ellipsoid rotates around the z axis depending on the value of z. See Figure 2b.
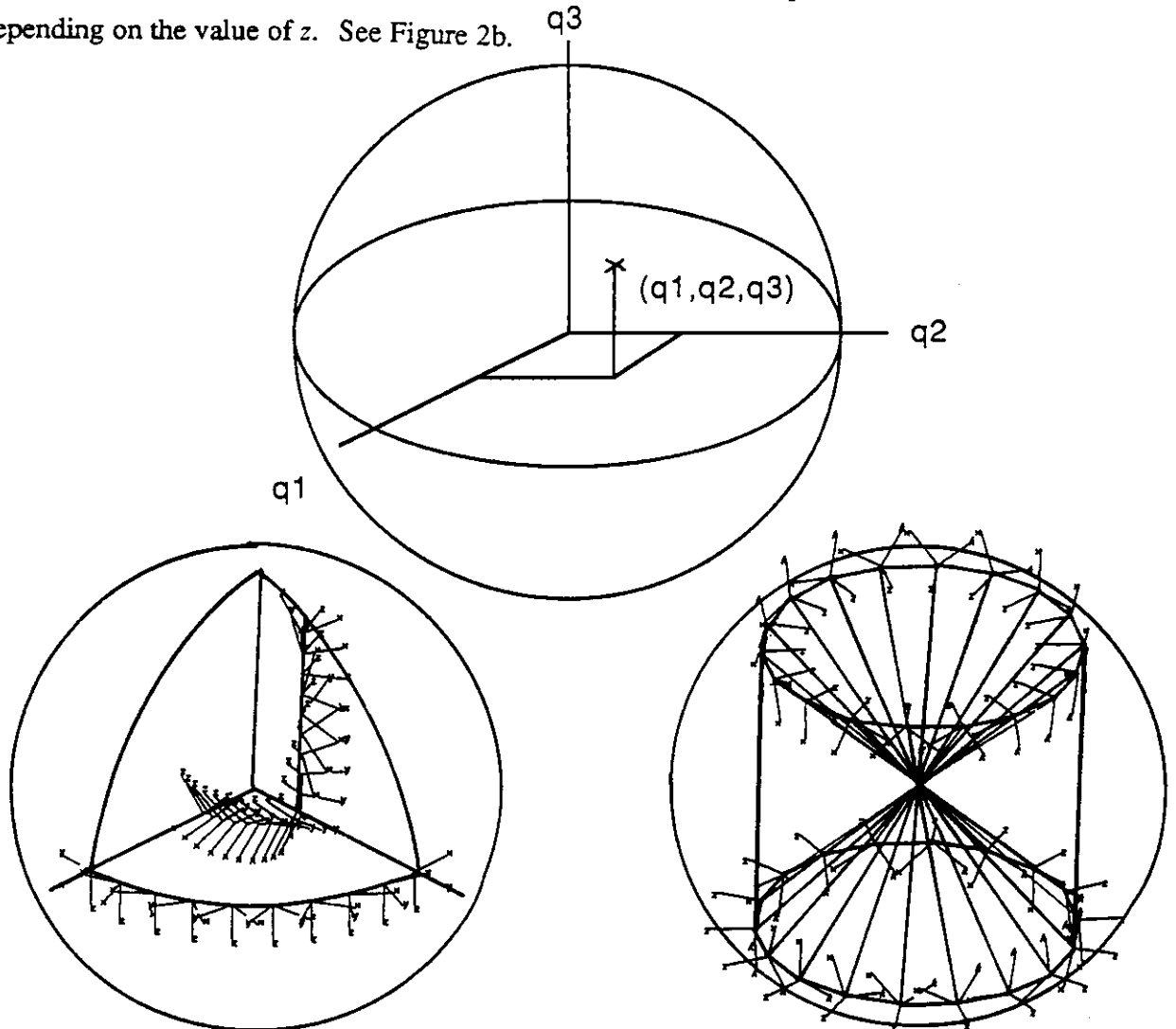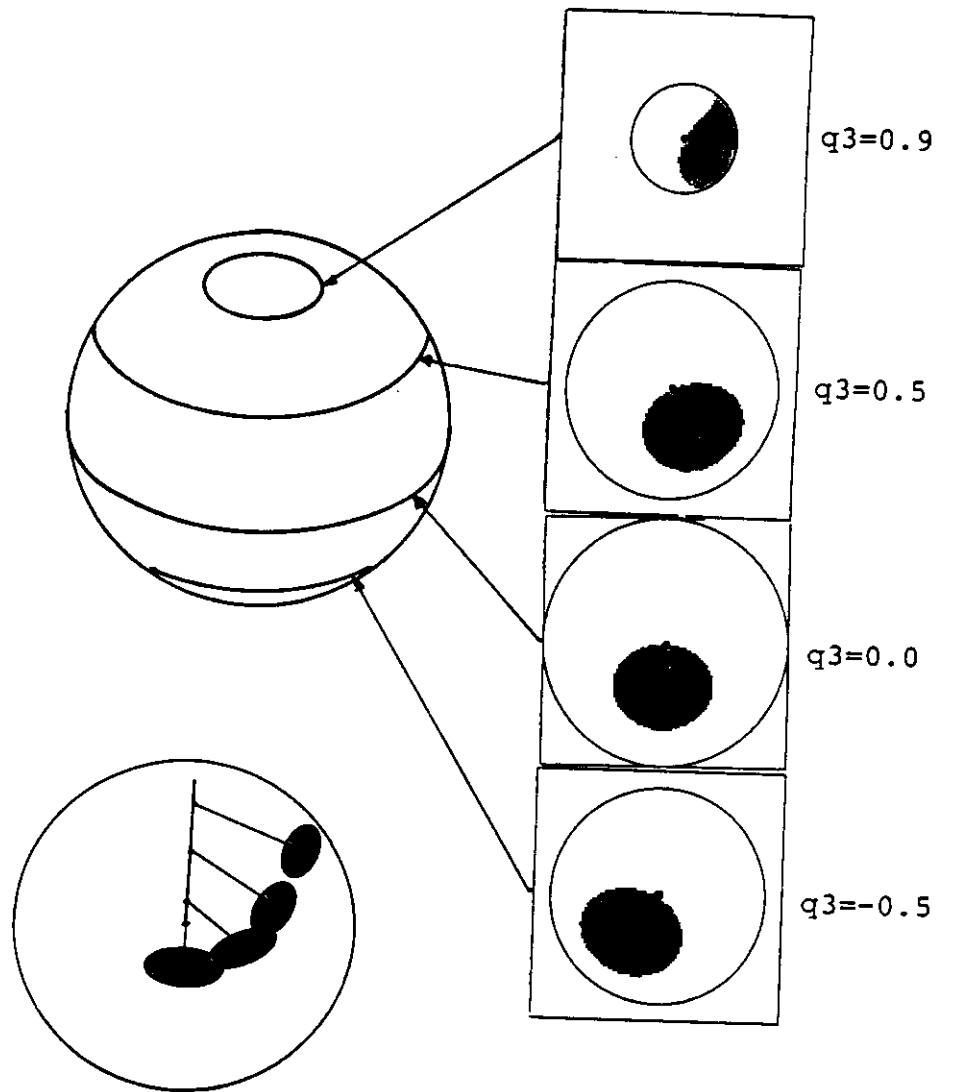

Figure 2a

Figure 2 Projected quaternion space

q3=0.9

q3=0.5

q3=0.0

q3=-0.5

Figure 2b The twisted cylinder problem

Figure 2, continued

3.4. Modified Projected Quaternion Space

The twisted cylinder problem can be solved by we mapping the project quaternion space to another space. We will consider the following sphere. We cut projected quaternion space in the plane at $z=0$ and straighten it up. Then, we convert the plane into a spherical surface covering the contents into the sphere. See Figure 3a. This operation gives a new sphere. On this sphere, the z axis of the corresponding coordinate agrees with the direction of the point from the center. All the coordinates on the spherical surface satisfy $\psi+\theta=0$. The inner points correspond to the coordinates whose z axis direction agrees with the direction from the center to the points, while the z axis rotation, ψ, determines the distance of the point from the spherical surface. This sphere may be considered as a modified Euler sphere. See Figure 3b.

The spherical surface, which comes from the $z=0$ of the projected quaternion space, $\phi+\psi=0$ and has no discontinuity around the north pole. Since the upper hemisphere satisfies the continuity criterion and the intuitiveness criterion, We define and represent various sensor properties on this space. We will refer this space as the configuration space.

We also use the modified projected quaterion sphere to express possible object attitudes on generating appearances of the objects. Since this space corresponds to the original quaternion space in a one-to-one correspondence, we can use quaternions for calculation, and project the result to this space, if necessary. For example, some applications require uniform digitization in terms of rotation. In that case, we can quantize the quaternion space uniformly [10] and then project the tessellated points to this space. This relationship between the quaterion space and the modified projected quaterion space resembles the relationship between the Gaussian sphere and the stereographic plane.
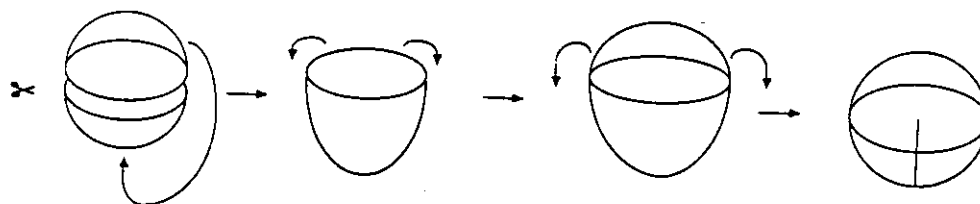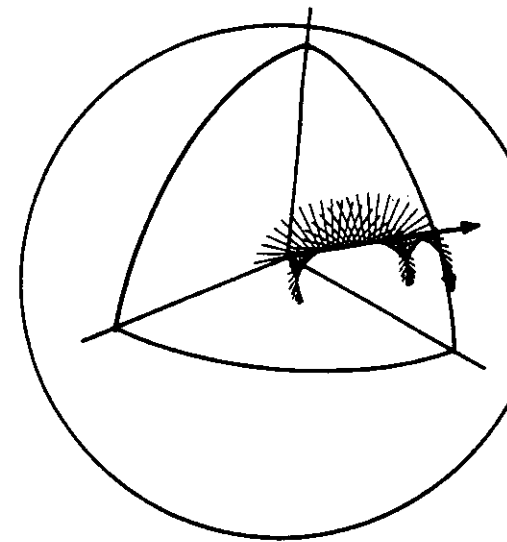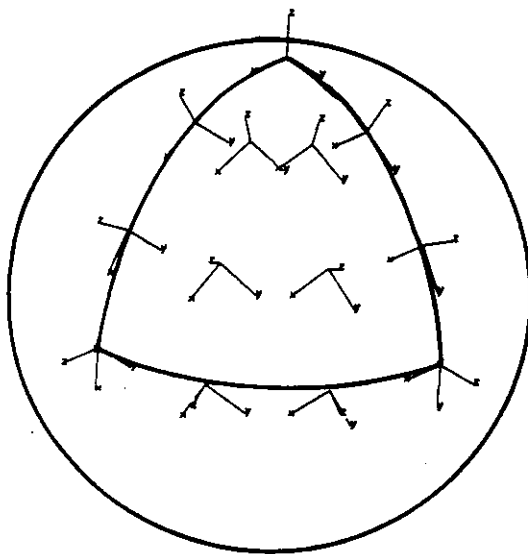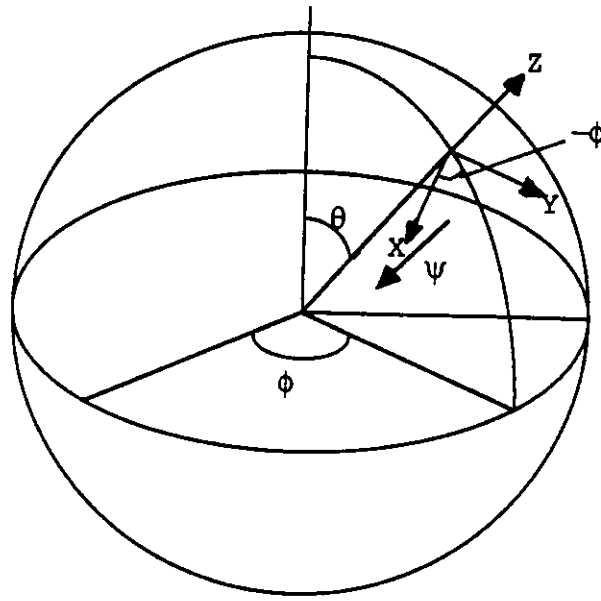

Figure 3a

Figure 3 Modified projected quaternion space

Figure 3b

Figure 3 Modified projected quaternion space

4. DETECTABILITY OF SENSORS

In the previous section, we have defined the way to represent the relationship between the sensor coordinate and the object coordinate. In this section, we will develop a constraint to determine whether an object feature can be detected at each point of the configuration space.

4.1. Constraints in Configuration Space for Feature Detection

Each sensor has two components: sources and detectors. For example, both a time-of-flight range finder and a light-stripe range finder have one source and on detector. Binocular stereo has one source and two detectors; photometric stereo has three sources and one detector. Table 3 summarizes the number of sources and detectors of each sensor.

Table 3 Source and Detector		
Sensor	Number of sources	Number of detectors
Edge Detector	1	1
Shape-from-shading	1	1
SAR	1	1
Time-of-Flight Range Finder	1	1
Light-stripe Range Finder	1	1
Binocular Stereo	1	2
Trinocular Stereo	1	3
Photometric Stereo	3	1
Polarimetric light detector	n	1

One source only illuminates one part of an object; one detector only observes one part of the object. Each sensor which consists of sources and detectors, only detect one part of the object. Thus, in order to specify the detectable area of each sensors, we need to define each source's illuminated area and each detector's visible area. We also need to define a operation method on illuminated areas and visible areas.

In the following discussion, we will consider both sources and detectors as generalized sources (G-sources). Each G-source has two properties: the illumination direction and the illuminated area. In the source case, the illumination direction and the illuminated area are the same as the nominal meanings. In the case of detectors, the illumination direction corresponds to the line of sight of the detector, and the illuminated area corresponds to the visible area from the detector.

In order to define the sensor detectability, we will use the configuration space previously defined. The illumination direction of a G-source is specified as a line in the configuration space; its illuminated area is specified as a volume in the configuration space. We will define two kinds of G-sources in terms of the distribution of illuminated areas: the uniform G-source and the directional G-source. A uniform G-source distributes its light evenly in all directions. An example of a uniform G-source is a usual light source whose illuminated area is located as a hemispherical corn of the sensor space. The center direction of the corn corresponds to the source direction.

We specify a uniform G-source as

(*NS type direction angle*)

The first argument, *type*, specifies what kind of feature the G-source illuminates, and takes one of the values; *face*, *edge*, and *vertex*. The second argument, *direction*, denotes the G-source illumination direction as a vector. The third argument, *angle* defines the illuminated area by specifying the spherical angle between the illumination direction and the surface normals associated with object features. If *type* is *face*, this angle defines the maximum allowable angle between the face surface normal and the illumination direction. If *type* is *edge*, this angle defines the maximum allowable angle of the smaller one of the two angles between the illumination direction and the two normals of incident surfaces to the edge. That is, if either or both faces are well illuminated, then the edge is considered to be illuminated. If *type* is *vertex*, we have to consider at least three faces incident to the vertex. This angle defines the maximum allowable angle of the smallest angle of those angles between the illumination direcption and the normals of incident surfaces. That is, if any of the incident faces of the vertex is illuminated, the vertex is considered to be illuminated.

Another kind of G-source is a directional G-source which projects light depending on the rotation around the light source direction. We specify a directional G-source as

(*DS type direction angle spec–direction spec–angle*)

The first argument, *type*, specifies one of the object features: *vertex*, *edge*, and *face*. The second argument, *direction*, denotes the G-source illumination direction as a vector. The third argument, *angle*, defines the spherical angle of the illuminated area, as for the uniform G-source. The fourth argument, *spec–direction* defines the constraint direction to be used in the following argument. The fifth argument, *spec–angle* defines the maximum allowable angle between the

constraint direction and the principal direction such as the surface normal of a face, the edge direction of an edge, and the average surface orientation around a vertex.

An example of a directional source is a directional edge detector. As mentioned before, a detector is also considered as a source, and its illuminated area corresponds to the detectable area. Since the directional edge detector only detects edges with certain orientations, it is regarded as a directional source. The illuminated area of a directional source becomes a thin slice of the configuration space.

We can specify the sensor characteristic with AND and OR operations of these formal definitions of all component G-sources of the sensors. Figure 4 shows sensor detectability represented by this method for all sensors listed in table 3.

4.2. Use of Feature Detection Constraints

The feature detection constraints are used together with a geometric modeler to predict how the object appears relative to the sensor. A geometric modeler generates possible attitudes of an object corresponding to each point in the configuration space. Then, detectability of each component face, edge, or vertex of the object under this attitude is determined using the constraint. The illumination direction constrains the surface shape to be detected, and the illuminated area constrains the surface orientation to be detected.

More precisely, we can imagine putting the configuration space on each point of the object in order checked whether it can be detected by the sensor. If the illumination direction from that point intersects with any of the surfaces of the object, the point cannot be detected. If the illumination direction does not intersect with any of the surface, the point is detectable. If the surface orientation on the point is contained in the illuminated area, the point is detectable. If the surface orientation is outside of the illuminated area, the point cannot be detected. Figure 5 illustrate the outline of this operation for the illumination direction and illuminated area using this constraint.
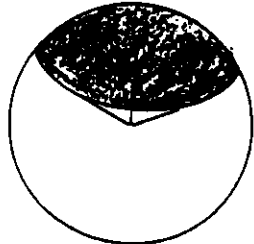
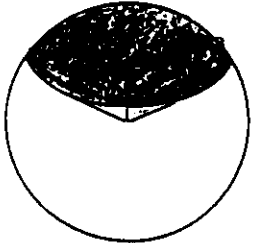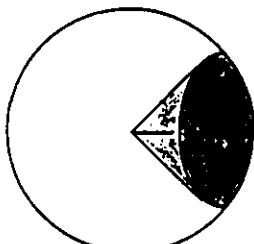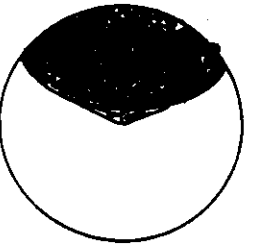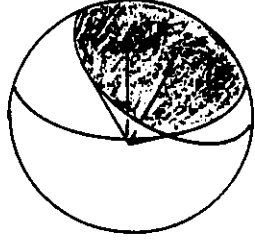| Sensor	Constraints in the formal definition	Constraints in the sensor space
Edge Detector	(AND (NS edge V d) (NS edge V d)) = (NS edge V d)	
Shape-from-shading	(AND (NS face V d) (NS face V d)) = (NS face V d)	

SAR	(OR (NS face V d)
	      (NS edge V d)
	      (NS vertex V d))
	(needs postprocess)	

Time-of-Flight Range Finder	(AND (NS face V d) (NS face V d)) = (NS face V d)	
Light-strip Range Finder	(AND (NS face V1 d) (NS face V2 d))	

Figure 4

Sensor	Constrains in the formal definition	Constraints in the sensor space
Binocular Stereo	(AND (NS edge V1 d1) (DS edge V2 d2 VE de) (DS edge V3 d3 VE de))	
Trinocular Stereo	(AND (NS edge V1 d1) (DS edge V2 d2 VE de) (DS edge V3 d2 VE de) (DS edge V4 d2 VE de))	
Photometric Stereo	(AND (NS face V d1) (NS face V1 d2) (NS face V2 d2) (NS face V3 d2))	
Polarimetric Light Detector	(OR (AND (NS face V d) (NS face V1 d)) (AND (NS face V d) (NS face V2 d)) . . .) where V . V = cos 2d	

Figure 4, continued
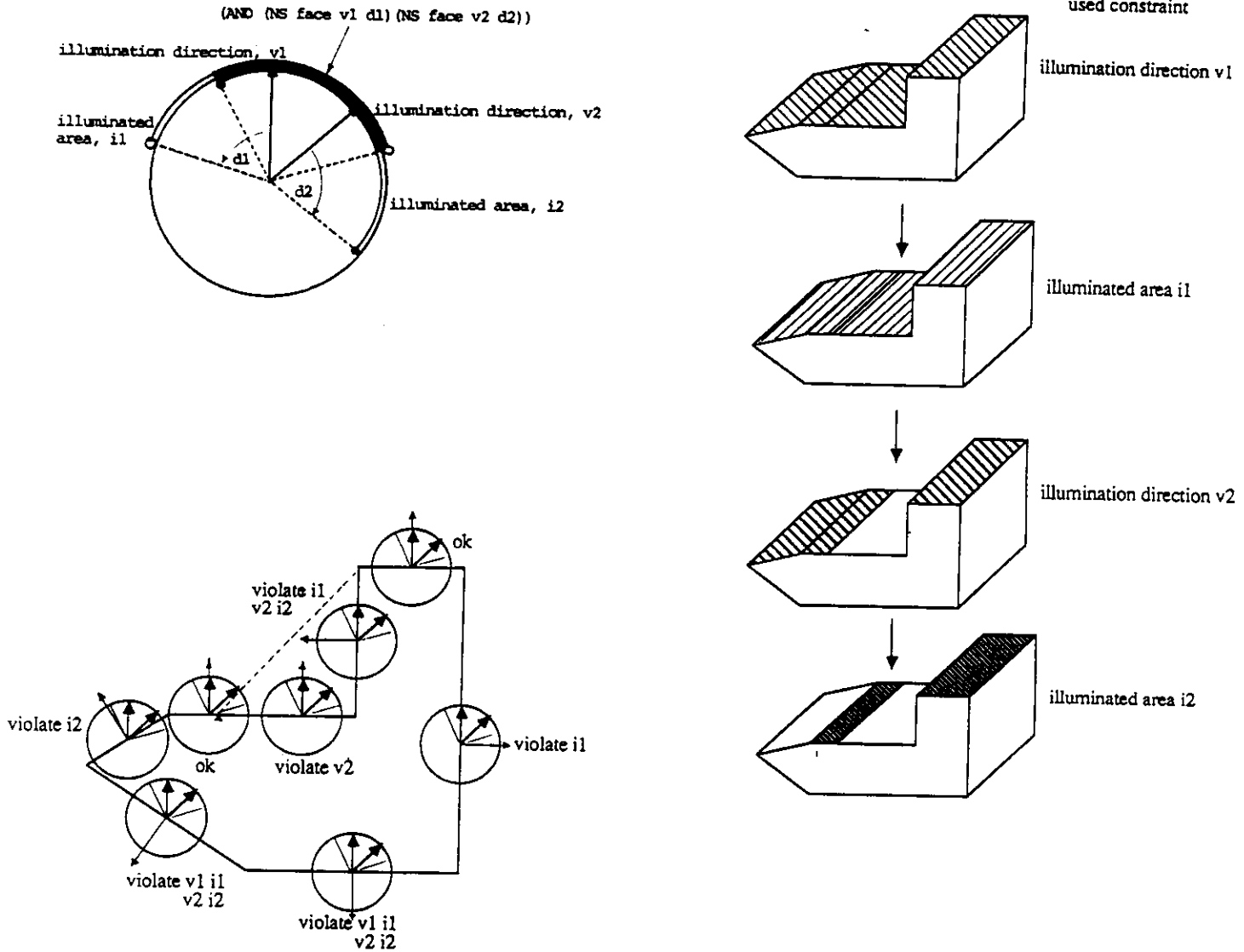
Figure 5 How to use the detection constraints

4.3. Detectability Distribution

The feature detection constraint gives the upper bound of the detectable areas in the configuration space. In some cases, however even though a object feature exists within the detectable area, the feature may be undetected due to noise. We define the detectability distribution such that a feature in the detectable area is actually detected. The probability is

usually high in the central part and low in the peripheral part of the detectable area.

The detectability distribution can be described by multiplication of detectability distributions of the component G-sources. Namely, each G-source has a detectability distribution defined over the illuminated area. In the previous subsection, the constraint was either illuminated or non illuminated or either detectable or non-detectable. We will expand this idea to the continuous case. Namely, each G-source has its own continuous detectable distribution over its illuminated area defined in the configuration space.

Since all sensors detect features based on a brightness distribution, the detectability distribution also depends on a brightness distribution which is detected and converted to sensor features. However, there are two types of sensors in terms of the conversion method; direct sensors and indirect sensors. The direct sensor measures the brightness value and converts it to sensor features, such as surface orientation, directly from the brightness value. The indirect sensor measures the brightness value and positional information of the bright spot if the brightness value is greater than some threshold. The indirect sensor then converts the positional information to sensor features such as depth. Table 4 shows a classification of sensors based on this difference.

Table 4 Measurement method	
sensor	direct/indirect
Edge Detector	direct
Shape-from-shading	direct
SAR	indirect
Time-of-Flight Range Finder	indirect
Light-stripe Range Finder	indirect
Binocular Stereo	indirect
Trinocular Stereo	indirect
Photometric Stereo	direct
Polarimetric light detector	indirect

Since a TV camera is a most typical input device, we will examine its performance before exploring the detectability distribution. Let $P(x|d), P(d|x)$, and $P(x)$ be the conditional probability of a real value x under the observed value d, the conditional probability of the observed data d under a real value x, and the probability of x, respectively. Then, TV camera performance can be

described using Bayes' theorem as

$$P(x|d) = \frac{P(d|x)P(x)}{\int P(d|x)P(x)dx}.$$

If we assume that $P(x)$ is constant, namely the brightness distribution occurs randomly,

$$P(x|d) = P(d|x).$$

The conditional probability of observed data d under the real value x is assumed as the Gaussian distribution whose mean value is x and standard deviation is σ,

$$P(x|d) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-d)^2}{2\sigma^2}}.$$

We can obtain σ from experiments. Our SONY CCD camera has $\sigma=3$, which roughly coincides with a result elsewhere [2].

Since the detectability distribution depends on sensing methods, we will develop the distributions for the photometric stereo as a representative case of the direct sensor, and for the light-stripe range finder as a representative case of the indirect sensor.

4.3.1. Detectability distribution of photometric stereo

An direct sensor such as photometric stereo can be modeled as

$$y = f(x)$$

where x is the input brightness, y is the output feature values, and f is the conversion function. Suppose X^* is the definition area of the function f; ie, the direct sensor outputs a feature value y from any x_i if $x_i \in X^*$. Then, the detectability distribution can be determined as the probability that the input brightness, $x+\delta x$, disturbed by δx, is still contained in the definition area, X^*. In order to be the problem more specific, we will examine the definition area of photometric stereo.

Photometric stereo determines the surface orientation from three images taken from the same position under different lighting directions.

$$I_1 = S_1 \bullet N$$
$$I_2 = S_2 \bullet N$$
$$I_3 = S_3 \bullet N,$$

where I_i, S_i, N are the brightness value under light source i, the i th light source direction vector, and the surface normal vector, respectively. Thus, expressing the brightness as a vector, I, and the light source as a matrix, S,

$I=SN$.

Applying S^{-1} to both sides, we obtain an explicit expression of N,

$N=S^{-1}I$.

This is the basic idea of photometric stereo [44].

Working photometric stereo has, however, two modification [25] to this theory.
1. S^{-1} is determined from calibration and stored in a lookup table rather than calculated from the ideal case.
2. Brightness values are normalized $I/|I|$ so that we can cancel the albedo effect.

We will obtain the detectability distribution of the photometric stereo. At first, we consider light source 1's detectability distribution. Assume a brightness value moves from i_1 to $i_1+\delta i_1$ due to sensor error. The normalization gives $i'_1+\delta i'_1=(i_1+\delta e_1)/(i_1+\delta i_1+i_2+i_3)$. However, the normalized intensity $(i'_1+\delta i'_1, i'_2, i'_3)$ exists on the same plane $i'_1+\delta i'_1+i'_2+i'_3=1$. Since a continuous area on the plane is the solution area for photometric stereo, we can obtain the solution from the new triple $i'_1+\delta i'_1, i_2, i_3$. Than is, we will always succeed to obtain the feature values, ie. we will have a unit detectability distribution for the light source 1. (Though of course the resultant value may be less reliable as will be discussed in the reliability section.) The same discussion is applicable to light source 2 and light source 3. Since the total detectability distribution is given as the multiple of all three detectability distributions of sources, the detectability distribution is a constant distribution over the detectable area in the configuration space. This analysis reveals that the normalization makes the detectability to be a unit value, and thus, helps to detect features in a stable manner.

4.3.2. Detectability distribution of a light-stripe range finder

An indirect sensor projects light on the scene and determines the positional features from the observed image or signal. Thus, the detectability distribution depends on whether a sensor can detect the returned light or not. Usually, to avoid the confusion of the returned value with background noise, threshold operations are applied, such as

i is detected   *if* $i \geq i_0$
i is not detected   *otherwise.*

Let us consider the light-stripe range finder as an example. A light-stripe range finder projects light stripes on the scene and recovers the depth at a point from the distance between two adjacent light stripes. Thus, the detectability function depends on whether the TV camera observes the light stripes or not.

Assuming that the surface is lambertian, the brightness of the stripe is determined by the angle between the surface normal, N, and the light source direction, S. Then, the brightness i is given by $N \bullet S$, while the disturbance factor δi is given by a Gaussian distribution,

$$p(\delta i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\delta i)^2}{2\sigma^2}}.$$

In almost all illuminated areas, $N \bullet S - i_0 >> 3\sigma$ holds, and the viewer direction does not affect the observed brightness. Thus, the detectability distribution is constant over the most part of illuminated area of the light source. In the peripheral area, however,

$$P_{detectable}(i) = P(i + \delta i \geq i_0) = \int_{i_0 - i}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\delta i)^2}{2\sigma^2}} d(\delta i)$$
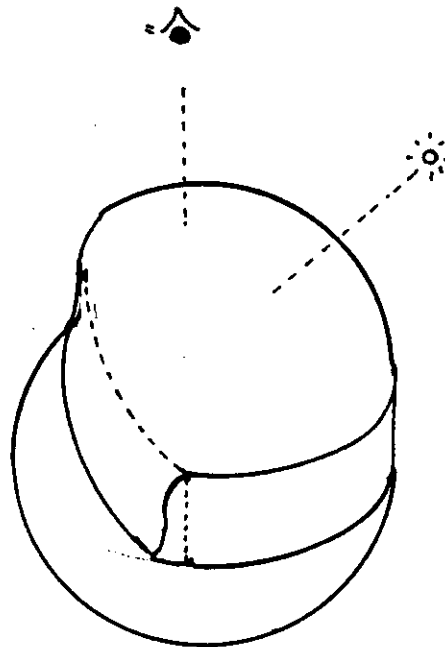
See Figure 6.

Figure 6 Detectability function of a light-stripe range finder

5. RELIABILITY OF SENSORS

Once a sensor feature is detected, then the next question is how reliable the sensor feature is. This section discusses two issues of sensor reliability. The first issue is the reliability of the detected sensor feature; data detected by a sensor always contains measurement error. To determine the bound of the error is important for model based vision. For example, suppose there is a sensor feature which the geometric model takes two nominal value 100 and 90 for two distinct situations or attitudes. If a sensor has an error range of plus/minus 1 for the sensor feature, we can use the feature from that sensor as one of reliable discriminators in the recognition stage. On the other hand, if a sensor has an error range of plus/minus 20, we cannot use the feature from that sensor.

The second issue is propagation of error from sensor features to geometric features, hence the resulting reliability of those geometric features. In some cases, a detected sensor feature from a sensor is used directly as a feature; in most cases, however, geometric features are derived from sensor features and are used as features in model based vision. Thus, it is necessary to determine the error propagation mechanism.

5.1. Reliability Distribution of Sensor Feature

Table 5 shows the main sources that affect reliability of sensor features. In addition to these, various digitization such as phase digitization in a time-of-flight range finder and spatial digitization in binocular stereo [2, 15, 33] must be considered but are omitted for the time being.

Table 5 Main factor of unreliability	
Sensor	Factor
Edge Detector	G-source brightness (TV camera)
Shape-from-shading	G-source brightness (TV camera)
SAR	G-source direction (camera direction)
Time-of-Flight Range Finder	G-source direction (mirror direction)
Light-stripe Range Finder	G-source direction (mirror direction)
Binocular Stereo	G-source direction (camera direction)
Trinocular Stereo	G-source direction (camera direction)
Photometric Stereo	G-source brightness (TV camera, light sources)
Polarimetric light detector	G-source direction (light source direction)

As shown in Table 5, the main error comes from G-source brightness in a direct sensor and from G-source direction in a indirect sensor. Thus, we will analyze the reliability of photometric stereo and the light-stripe range finder as representatives of the direct and indirect sensors, respectively.

5.1.1. Reliability distribution of photometric stereo

For the direct sensor, $y=f(x)$, the disturbance of δx is propagated via f. Namely, the disturbance of the detected value, δy is

$$\delta y = f'(x)\delta x$$

Our photometric stereo can be described as two step processes. First a original brightness triple is converted to a normalized brightness triple.

$$\mathbf{I'} = \mathbf{I}/|\mathbf{I}|.$$

Then, the normalized brightness triple is converted to a surface orientation $\mathbf{N}$.

$$\mathbf{N} = \mathbf{S}^{-1}\mathbf{I'}.$$

Let us denote the brightness disturbance distribution as $N(0,\sigma^2)$. Then the normalized brightness distribution is denoted as $N(I,(\sigma f')^2)$, where f' is the first derivative of f. Figure 7a shows the distribution of f' over the detectable area. Although it is possible to approximate the distribution with polynomial, we assume it is constant (0.004) over the detectable area for simplicity. Since $\sigma=3$, $2\sigma f'=0.03$. This value corresponds to a 1.5 mesh in the lookup table.

We determine $\mathbf{S}^{-1}$ from the real data, because $\mathbf{S}^{-1}$ is represented as a lookup table. Figure 7b shows the angular distance in terms of mesh number. Namely, the figure shows angular differences between two adjacent surface normals in the lookup table. By using this result and a 1.5 mesh error from the brightness distribution, the total error becomes 5 degrees over the detectable area. This agrees with the observation from the experiment, which has plus/minus 5 degrees error in determining surface orientations over the range of detectable surface orientations. See Figure 7c.

5.1.2. Reliability distribution of a light-stripe range finder

In the case of indirect sensors, the main source of unreliability comes not from the G-source brightness but from the G-source direction. The indirect sensor can be modeled as

$$z=f(y(\mathbf{v}_1,\mathbf{v}_2,..\mathbf{v}_n)).$$

$\mathbf{v}_i$ denotes the ith G-source direction, and y denotes the conversion function from G-source
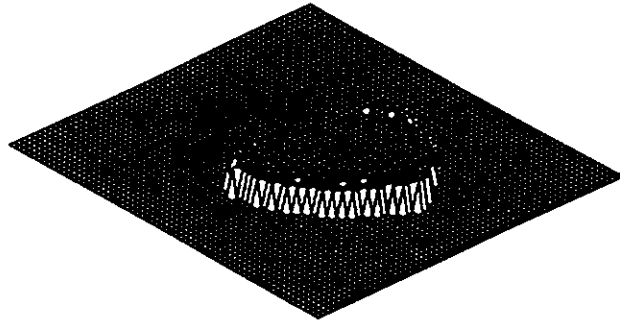
Figure 7a Distribution of *f*

angular
difference
between
two adjacent
cells

5°

50° θ (azimuth
angle)

Figure 7b Angular distance in terms of mesh

Figure 7
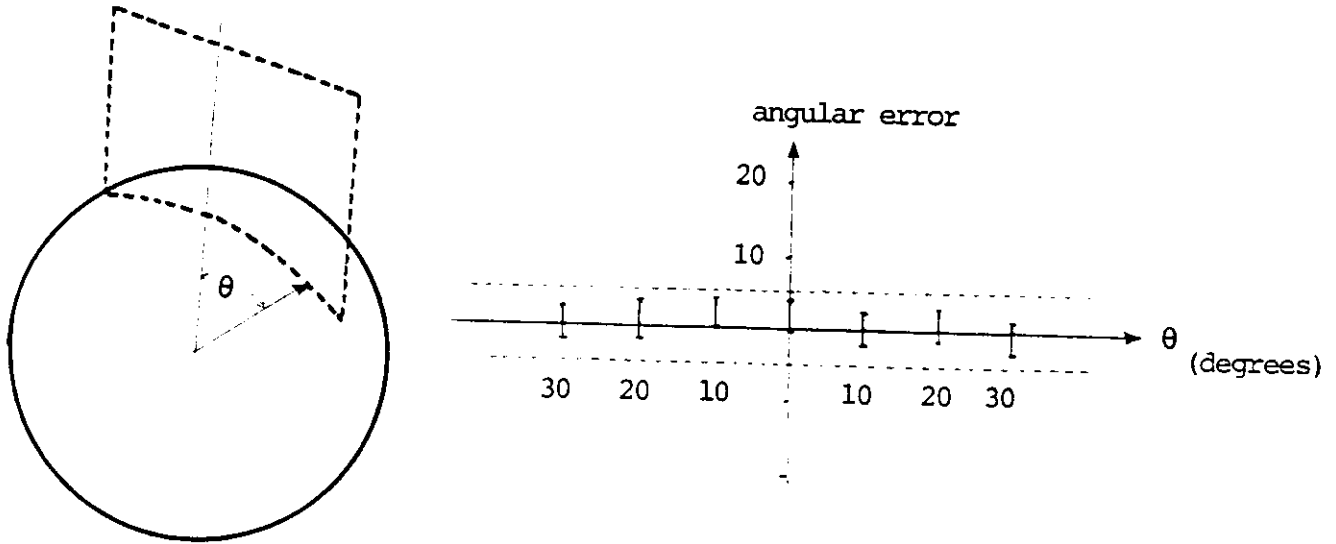
angular error

Figure 7c Angular error over the detectable area

Figure 7

directions to the positional information, while f denotes the conversion function from the positional information such as a bright spot to the observed data, and z specifies the detected data such as depth or surface orientation.

$$\delta z = \Sigma_i f \frac{\partial y}{\partial v_i} \delta v_i.$$

Thus, we will examine the disturbance based on the G-source direction.

In the light-stripe range finder, we will calculate $f \frac{\partial y}{\partial v_i}$ directly from the system analysis. The angular error in the mirror is propagated to the observed error with the physical conversion process. The propagation process can be divided into three parts: mirror error to positional difference in the stripe, positional difference in the stripe to image difference, and image difference to the converted distance difference. The propagation process can be obtained analytically as follows.

Let us denote the angular error as $r\delta\theta$, where r is the distance from the light source to the physical point. At the physical place A, the laser light is intercepted. Then, due to the angular error, the physical difference, δy occurs.

$$\delta y = \frac{r}{\cos \alpha} \delta\theta,$$

where α is the angle between the light source, **S**, and the surface normal, **N**. See Figure 8a.

This physical difference is observed from the TV camera, and the image difference, δi occurs.

$$\delta i = (\cos\beta)\delta y,$$

where β is the angle between the surface normal, **N**, and the viewer direction, **V**.

The third step is the propagation from the image plane to the distance. For the simplicity, we will assume that the camera model be orthographic projection. Then, the horizontal difference, δi, is propagated into the distance error δz as

$$\delta z = \frac{\delta i}{\tan\gamma},$$

where γ, is the angle between the viewer direction, **V**, and the light source direction, **S**. See Figure 8a.

Finally, we obtain

$$\delta z = \frac{\cos\beta}{\cos\alpha\tan\gamma} r\delta\theta$$

$$= \frac{(\mathbf{N}\bullet\mathbf{V})(\mathbf{S}\bullet\mathbf{V})}{(\mathbf{N}\bullet\mathbf{S})\sqrt{1-\mathbf{S}\bullet\mathbf{V}}} r\delta\theta.$$

Figure 8b shows the reliability distribution over the detectable area.

5.2. Propagation of Reliability to Geometric Features

Usually raw data detected by a sensor is converted into geometric features such as distance, area, and inertia. This process propagates errors into the geometric features due to two reasons: the detectability distribution and the reliability distribution. Since this conversion process depends on the detected data, we will concentrate on faces as detected data, because most of the active sensors detect faces as the primal features.

5.2.1. Error propagation from detectability distribution

Most active sensors detect physical patches as detected pixels. Usually, these detected pixels will be grouped and converted into isolated regions. If the sensor fails to find a detectable pixel, the measured area will be reduced from the nominal area size given by a model. This process can be modeled as follows:

Suppose the detectability probability is p over a region and the nominal area size of the region is n. Under this condition, the probability to observe x pixels out of n pixels over the region is
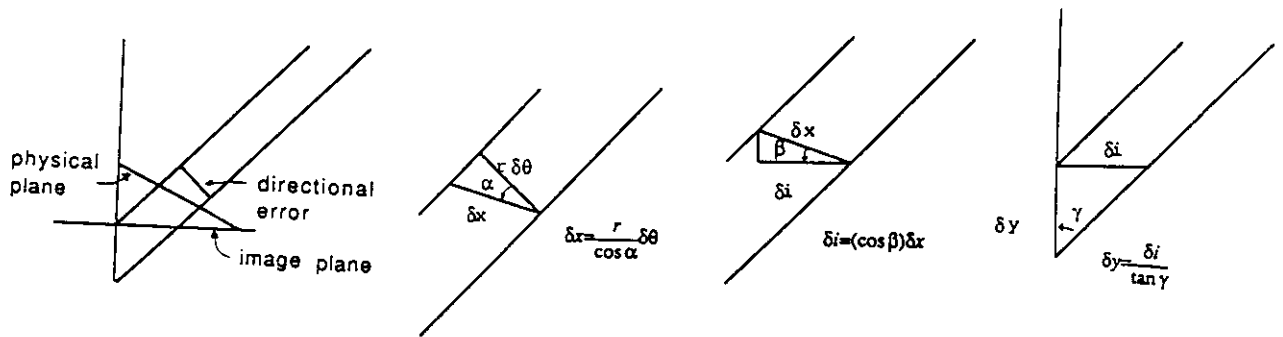
Figure 8a
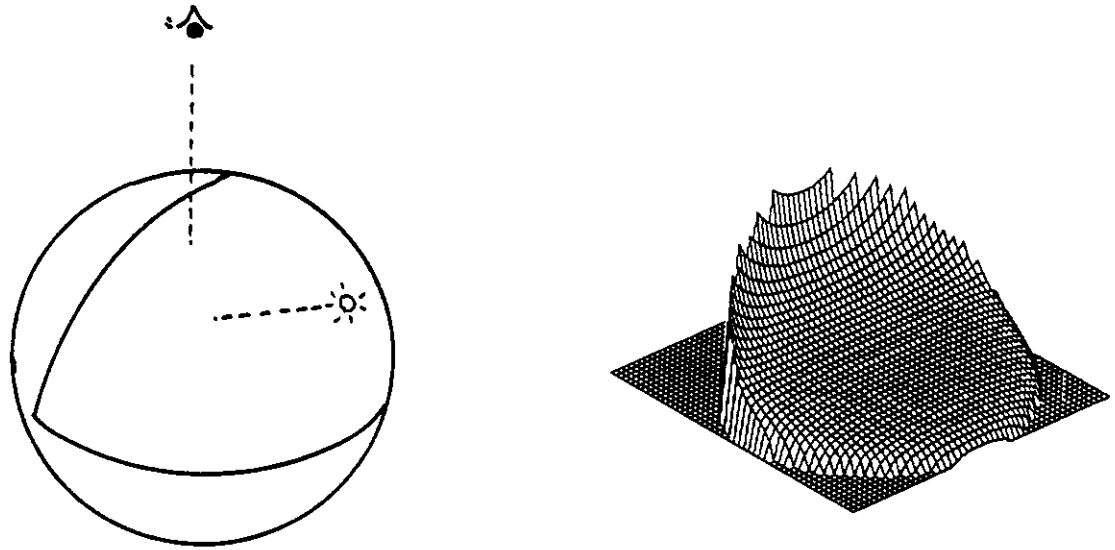
Figure 8b

Figure 8 Reliability distribution of a light-stripe range finder

$$P(x)=\binom{n}{x}p^x q^{n-x}$$

Namely, this probability denotes that the system executes n trials and succeeds to detect x pixels under that condition that success probability is p and fail probability is q, where $p+q=1$. This process is a binomial distribution; the mean and variance of this distribution are

$m=np$
$\sigma^2=npq$

This gives the error propagation (area reduction ratio) for the detectability distribution.

Since both photometric stereo and a light-stripe range finder give $p=1$, neither sensor causes an area reduction due to the detectability distribution.

5.2.2. Error propagation from reliability distribution

The Next factors to be considered is the reliability of detected data. We recover the geometric features such as distance, area, and inertia from skewed raw data by an affine transform, based on the observed surface orientation in either the photometric stereo or light-stripe range finder. Thus, if the raw data are erroneous, the obtained geometric features are also erroneous.

Let d be the real distance, and $d+\delta d$ be the observed one. Then, the physical system generates an observed distance $d\cos\theta$, while due to the sensor error, we will measure this surface surface orientation as $\cos(\theta+\delta\theta)$, where θ is the angle between the viewer and surface normal. Thus, for $\delta\theta$ small, we get

$$d+\delta d = d\cos\theta/\cos(\theta+\delta\theta) = d(1+\delta\theta\tan\theta).$$

In the area case,

$$a+\delta a = a(1+2\delta\theta\tan\theta).$$

In the inertia case,

$$i+\delta i = i(1+3\delta\theta\tan\theta).$$

These formulas give error propagation from angular error to features at each pixel. We will obtain geometric features from a region which consists of n pixels. Thus, the system will execute n trials of measuring $\delta\theta_i$ which is approximated as a Gaussian distribution, $N(0,\sigma^2)$, and observe the total $\Sigma_i\delta\theta_i$. From the theorem of the Gaussian distribution, $\delta\Theta=\Sigma_i\delta\theta_i$ is a Gaussian distribution, $N(0,n\sigma^2)$.

By using these formulas, we calculate error ratio of areas and inertia for photometric stereo as shown in Table 6. Predicted results are obtained based on the reliability of photometric stereo developed previously and the formulas of section. We use $n=70$, $\sigma=0.045$. Observed results are obtained from the distribution of the real data sampled five times. Similar results are expected in the other sensors.

Table 6	Reliability of Geometric Features	
Feature	Observed	Predicted
Area	0.02	0.045

Inertia	0.05	0.067

6. CONCLUDING REMARKS

This paper discussed modeling sensors for model-based vision. Our sensor model consists of two characteristics: sensor detectability and sensor reliability. Sensor detectability specifies under what conditions a sensor can detect a feature, while sensor reliability denotes how reliable the obtained measurement is over the detectable area.

We have proposed to use a modified projected quaternion space as the configuration space which represents the relationship between sensor coordinates and object coordinates. The sensor detectability and the sensor reliability are expressed in this configuration space. Constraints in the configuration space involved in detecting features have been developed by using G-source illuminated area and G-source illumination direction. We have shown how to compute the sensor detectability distribution and the sensor reliability distribution for photometric stereo and a light-stripe range finder as examples.

In model based vision, expected values of various features can be computed from 3D geometric model. Those values are, however, nominal values that they should take in ideal cases or should be sensed by ideal sensors. On the other hand, actually observed sensor data contains noises and should be used accordingly. The sensor model bridges the discrepancy between these two values by modeling the distribution of the sensed value based on the characteristics of a given sensor. In model-based vision, it is possible to precompile a given 3D geometric model into a recognition strategy [24]. This precompilation cannot generate an optimal strategy without knowing each feature's reliability, because the strategy should use the most stable features at each recognition step. Thus, the sensor model is an essential component in model-based vision. We have to explore more reliable sensor models for this purpose.

We also have analyzed the error propagation mechanism from detected data to the geometric features. This is important, because quite often we are interested in geometric features derived from the detected sensor features. Once we establish the error propagation mechanism from detected sensor features to geometric features, we can also assess the reliability of the geometric features, hence we can construct a recognition system more systematically and reliably. Further study is required in this area.

To calculate detectable features of an object under the constraints of various sensors is a tedious job when we use a conventional geometric modeler. The better way is to interface a geometric modeler with the sensor model proposed. We call this a sensor modeler. The traditional geometric modeler only allows users to generate a 3D object by combining primitive objects and to display its views. In this sense, the traditional modeling system has only one sensor model which is projection. The sensor modeler we propose can generate various 2D representations under given sensor specifications. Part of this facility is being implemented in our new geometric/sensor modeler VANTAGE [19].

7. ACKNOWLEDGEMENT

References

[1]   Agin, G.J. and Binford, T.O.
        Computer description of curved objects.
        In *Proc. of 3rd Intern. Joint Conf. on Artificial Intelligence*, pages 629-640. Stanford,
            CA, August, 1973.

[2]   R. Bajcsy, E. Krotkov, M. Mintz.
        Models of errors and mistakes in machine perception.
        In *Proc. of Image Understanding Workshop*. DARPA, 1987.

[3]   Baker, H.H. and Binford, T.O.
        Depth from edges and intensity based stereo.
        In *Proc. of 7th Intern. Joint Conf. on Artificial Intelligence*. 1981.

[4]   Barrow, H.G. and Popplestone, R.J.
        Relational description in picture processing.
        Machine Intelligence 6.
        Edinburgh University Press, Edinburgh, Scotland, 1970.

[5]   Binford, T.O.
        Visual perception by computer.
        In *Proc. IEEE Systems Science and Cybernetics Conf.*. IEEE, 1971.

[6]   Bolles, R.C. and Horaud, P.
        3DPO: A three-dimensional part orientation system.
        In Kanade, T. (editor), *Three-Dimensional Machine Vision*. Kluwer, Boston MA, 1987.

[7]   Brady, J.M. and Asada, H.
        Smoothed local symmetries and their implementation.
        The International Journal of Robotics Research 3(3), 1986.

[8]   Brady, M., Ponce, J., Yuille, A., and Asada, H.
        Describing surfaces.
        In Hanafusa, H. and Inoue, H. (editors), *Proc. 2nd International Symposium on Robotics
            Research*. MIT Press, Cambridge, MA, 1985.

[9]   Brooks, R.A.
        Symbolic reasoning among 3-D models and 2-D images.
        Artificial Intelligence 17(1-3), 1981.

[10]   Brou, P.
        Using the Gaussian image to find the orientation of object.
        The International Journal of Robotics Research 3(4), 1983.

[11]   Canny, J.F.
        Finding edges and lines in images.
        Technical Report AI-TR-720, Massachusetts Institute of Technology, Artificial
            Intelligence Laboratory, 1983.

[12]   Canny, J.F.
        Collision detection for moving polyhedra.
        Technical Report AI memo 806, Massachusetts Institute of Technology, Artificial
            Intelligence Laboratory, October, 1984.

[13]   Chakravarty, I. and Freeman, H.
       Characteristic views as a basis for three-dimensional object recognition.
       In *Proc. The Society for Photo-Optical Instrumentation Engineers Conference on Robot Vision*. SPIE, Bellingham, Wash., 1982.

[14]   Cutrona, L.J.
       Synthetic Aperture Radar.
       Radar Handbook.
       McGraw Hill, New York, 1970, Chapter 23.

[15]   Faugeras, O.D., Ayache, N., Faverjon, B. and Lustman, F.
       Building visual maps by combining noisy stereo measurement.
       In *Proc. of Intern. Conf. on Robotics and Automation*, pages 1433-1438. San Fransisco, April, 1986.

[16]   Grimson, W.E.L.
       From Images to Surfaces: a computational study of the human early visual system.
       MIT Press, Cambridge, MA, 1981.

[17]   Hamilton, W.R.
       Elements of Quaternions.
       Chelsca, New York, 1969.

[18]   Hebert, M. and Kanade, T.
       Outdoor scene analysis using range data.
       In *Proc. of Intern. Conf. on Robotics and Automation*, pages 1426-1432. IEEE Computer Society, San Francisco, April, 1986.

[19]   Hoffman, R., Ikeuchi, K., Kanade, T., Kumar, B., and Robert, J.C.
       VANTAGE.
       Technical Report, Carnegie-Mellon University, Robotics Institute, 1987.
       in preparation.

[20]   Horn, B.K.P.
       Obaining Shape from Shading.
       In Winston, P.H. (editor), *The Psychology of Computer Vision*. McGraw-Hill, New York, 1975.

[21]   Horn, B.K.P.
       Extended Gaussian Images.
       Proc of the IEEE 72(12), December, 1984.

[22]   Horn, B.K.P.
       Robot Vision.
       MIT Press, Cambridge, MA, 1986.

[23]   Ikeuchi, K.
       Recognition of 3-D objects using the extended Gaussian image.
       In *Proc. of 7th Intern. Joint Conf. on Artificial Intelligence*. 1981.

[24]   Ikeuchi, K.
       Generating an Interpretation Tree from a CAD Model for 3-D Object Recognition in Bin-Picking Tasks.
       International Journal of Computer Vision 1(2), 1987.

[25]   Ikeuchi, K., Nishihara, H.K., Horn, B.K.P., Sobalvarro, P., and Nagata, S.
       Determining grasp points using photometric stereo and the PRISM binocular stereo
           system.
       The International Journal of Robotics Research 5(1), 1986.

[26]   Ikeuchi, K. and Horn, B.K.P.
       Numerical shape from shading and occluding boundaries.
       Artificial Intelligence 17(1-3), 1981.

[27]   Jarvis, R.A.
       A laser time-of-flight range scanner for robotic vision.
       IEEE Tran, Pattern Analysis and Machine Intelligence PAMI5(5), 1983.

[28]   Koenderink, J. J. and Van Doorn, A. J.
       Geometry of binocular vision and a model for stereopsis.
       Biological Cybernetics 21(1), 1976.

[29]   Koshikawa, K.
       A polarimetric approach to shape understanding of glossy objects.
       In *Proc. of 6th Intern. Joint Conf. on Artificial Intelligence.* 1979.

[30]   Marr, D. and Hildreth, E.
       Theory of edge detection.
       Proc. of the Royal Society of London B 207, 1980.

[31]   Marr, D. and Nishihara, H.K.
       Representation and recognition of the spatial organization of three-dimensional shapes.
       Proc. of Royal Society of London B 200, 1978.

[32]   Marr, D. and Poggio. T.
       A computational theory of human stereo vision.
       Proc. of the Royal Society of London B 204, 1979.

[33]   Matthies, L. and Shafer, S.A.
       Error modelling in stereo naviagtion.
       Technical Report CMU-CS-86-140, Carnegie-Mellon University, Computer Science
           Department, Pittsburgh, PA, 1986.

[34]   Mensa, D.L.
       High Resolution Radar Imaging.
       Artech House, Dedham MA, 1981.

[35]   Milenkovic, V.J. and Kanade, T.
       Trinocular vision: using photometric and edge orientation constraints.
       In *Proc. Image Understanding Workshop.* DARPA, Miami Beach, FL, December, 1985.

[36]   Ohta, Y. and Kanade, T.
       Stereo by intra- and inter-scanline search using dynamic programming.
       IEEE Trans Pattern Analysis and Machine Intelligence PAMI-7(2), 1985.

[37]   Oshima, M. and Shirai, Y.
       A model based vision for scenes with stacked polyhedra using 3D data .
       In *Proc. Intern. Conf. on Advanced Robot (ICAR85).* Robotics Society of Japan, 1985.

[38]   Pentland, A. P.
       Perceptual Organization and the Representation of Natural Form.
       Artificial Intelligence 28(2), 1986.

[39]   Pervin, E. and Webb, J. A.
       Quaternions in computer vision and robotics.
       In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. IEEE Computer Society,
           Washington, D.C., June, 1983.

[40]   Roberts, L.G.
       Machine perception of three-dimensional solids.
       In Tipplett, J.T. (editor), *Optical and Electro-Optical Information Processing*. MIT Press,
           Cambridge, MA, 1965.

[41]   Shafer, S. A. and Kanade, T.
       *The Theory of Straight Homogeneous Generalized Cylinders, and A Taxonomy of
           Generalized Cylinders.*
       Technical Report CMU-CS-83-105, Carnegie-Mellon University, Computer Science
           Department, January, 1983.

[42]   Smith, D.
       Using enhanced spherical images.
       Technical Report AI Memo 451, MIT Artificial Intelligence Laboratory, 1979.

[43]   Tomiyasu, K.
       Tutorial review of Synthetic-Aperture Rader(SAR) with applications to imaging of the
           ocean surface.
       Proc. of the IEEE 66(5), May, 1978.

[44]   Woodham, R.J.
       Reflectance Map Techniques for Analyzing Surface Defects in Metal Castings.
       Technical Report AI-TR-457, Massachusetts Institute of Technology, Artificial
           Intelligence Laboratory, Cambridge, MA, 1978.