RECOGNIZING OBJECTS USING SURFACE INFORMATION AND OBJECT MODELS

Robert B. Fisher

Department of Artificial Intelligence, University of Edinburgh, UK

Abstract: This paper describes research on recognizing partially obscured objects using three dimensional surface data and geometrical object models as input. The paper shows that surface information is an important input to the visual understanding process. This is because surfaces are the features that directly link perception to the objects perceived (for normal "camera-like" sensing) and because surface understanding makes explicit information needed to understand and cope with some visual problems (e.g. obscured features).

Keywords: object recognition, scene analysis, surface models

1.0 INTRODUCTION

This paper summarizes new results in object recognition based on surface information (e.g. direct 3D scene data) rather than edge information. Here, the surface information is absolute surface depth and orientation organized in a pointillist array aligned with a standard intensity image of the scene. This data can be supplied by several processes currently under development, including stereo and laser rangefinding.

The advantages of using surface information are:

- surface information is explicitly 3D, so 3D properties can be calculated directly, rather than deduced from 2D image properties,
- surfaces can be segmented from both the model and the data according to the same criteria, hence producing directly corresponding features (disregarding scale questions),
- it makes explicit features found on non-blocks-world objects (e.g. curved surface patches),
- it makes explicit where occlusion occurs (depth discontinuity boundaries), and
- it provides larger features for model matching, thus reducing the effect of noise and minor errors.

An object is considered to be recognized if image data can be found that consistently supports the object hypothesis, with the goal of completely explaining the presence or absence of all object features. For partially obscured objects, this requires that the hypothesis be instantiated in spite of missing features and that all remaining evidence is consistent with the hypothesized occlusion.

Recognition (here) is based on pairing data surface regions to model surface regions, which allows direct estimation of the object's spatial position. From this, the program can deduce what object features are expected to be visible, with what aspect and where. To do this, what is needed is:

- surface-based object models,
- criteria for segmenting the image data,
- methods for grouping surface features together,
- methods for estimating an object's position,
- methods for finding data features to pair to the model and
- methods for predicting or overcoming occlusion.

Using these, recognition succeeds with partially obscured, flexibly connected, moderately complicated objects seen from unpredictable viewpoints. Occlusion is thus more manageable provided one uses information made explicit by surface based representations.

The research reported here builds on much research in related areas. For surface extension (section 3), boundary connection methods of Guzman (8) have been extended to use additional constraints embodied in 3D data. These are continuity of surface depth and orientation as well as boundary connectedness. Surface cluster formation (section 4) extends the blocks-world 2D image rules of Waltz (11) and 3D heuristics of Sugihara (10) to group laminar surfaces. Here, the scene constraints were not needed to deduce segment labels as the surface image provides this information directly. Object modeling (section 5) follows ACRONYM (2), except that surface segments are the chosen primitives, rather than volumes. This allows direct model to data pairings, as well as allowing easier feature visibility predictions. The surface-based object models are like those of Faugeras and Hebert (3) except here complete models are used, with hierarchical substructure and flexible attachments. Some work on 3D surface description has started but this has concentrated more on finding continuous surface descriptions (e.g. Brady et al (1)) rather than summary properties. Geometrical reasoning was more advanced in ACRONYM (2), except their interests were in solids and linear image features, whereas here surface reasoning predominates. Faugeras and Hebert (3) used information from planer surface patches for a least squares estimate of object position. This would probably improve on the positional results shown here, which refined bounded estimates in a six dimensional parameter space. The work here also uses constraints from curved surfaces and previously recognized subcomponents. Little previous work has been done on fully recognizing objects, including understanding occlusion and locating missing features.

2.0 SURFACE IMAGES AND SEGMENTATION

Recognition starts from surface data, as represented in a structure called a labeled, segmented surface image (LSSI). This structure is like Marr's $2\frac{1}{2}$D sketch (9) and includes a pointillistic representation of absolute depth and local surface orientation. The surfaces are separated into regions by boundary segments labeled as shape or obscuring edges. Shape segmentation is based on orientation, curvature magnitude and curvature direction discontinuities (5). Obscuring boundaries are placed at depth discontinuities. These criteria segment the surface image into regions of nearly uniform shape, characterized by the two principal curvatures and the surface boundary. As no fully developed processes produce this data yet, the program input is from computer augmented, hand-segmented test images. (Several laboratory systems produce similar data though (3,10).) Figure 2 shows part of the input used for the test scene shown in figure 1. Part (a) shows the cosine of the surface slant for each image point. Part (b) shows the obscuring boundaries.

3.0 COMPLETE SURFACE HYPOTHESES

The image segmentation directly leads to partial or complete object surface segments. Surface completion processes reconstruct obscured portions of surfaces, when possible, by connecting extrapolated surface boundaries behind obscuring surfaces. The advantage of this is twofold - it provides data surfaces more like the original surface for property extraction (section 6) and the extended surfaces give better image evidence during hypothesis completion. Two processes are used for completing surface hypotheses. The first bridges over gaps in single surfaces and the second links two completely segmented surface patches. Merged surface segments must have roughly the same depth and surface characterization. Figures 3a and 3b illustrate both rules in showing the original and reconstructed robot upper arm large surface from the test image.

4.0 SURFACE CLUSTERS

Surface hypotheses are joined to form surface clusters, which are blob-like 3D identity-independent representations. The goal of this process is to partition the scene into a set of 3D solids, without yet knowing their identities. These re useful (here) for aggregating image features into contexts for model matching. They would also be useful for tasks where identity is not necessary, such as object avoidance.

Forming a surface group is based on finding closed loops of isolating boundary segments. Isolating boundaries are generally obscuring and concave surface orientation shape segmentation boundaries. An exception occurs for laminar objects, where the obscuring boundary across the front lip of the trash can (figure 4) does not isolate the surfaces. These criteria determine the primitive surface clusters and larger clusters are formed based on depth ordering relationships. Figure 4 shows some of the primitive surface clusters for the test scene.

5.0 SURFACE-BASED OBJECT REPRESENTATION

Objects are compact, connected solids with definable surface boundaries, where the surfaces are rigid and segmentable at some appropriate scale. The class of objects recognizable by the implemented program may also have rigid subassemblies with possibly flexible interconnections.

Model-based object recognition requires geometric object models. Here, the models are designed for object recognition, not image creation, so the model primitives are based on matchable image features.

The model used here has surface patches as primitives, because the surface is the primary data unit. This allows more direct pairing of data with models, comparison of surface shapes and estimation of model to scene transformation parameters (7). Surfaces are described by their principal curvature parameters and extent boundary. Surfaces have zero, one or two directions of curvature (positive or negative). The segmentation ensures that the shape (e.g. principle curvatures) remains relatively constant over the entire surface segment. The boundary describes the limits of the surface.

Objects are recursively constructed from surfaces or subobjects using coordinate reference frame transformations. Each structure has its own local reference frame transformation and larger structures are constructed by placing the subcomponents in the reference frame of the aggregate. Variable transformations connect subobjects flexibly, by using variables in the attachment relationship. The geometrical relationship between structures is useful for making model to data assignments and for providing the adjacency and relative placement information used by verification.

Illustrated first is the surface definition for the robot upper arm large side panel (uside). The first triple on each line give the starting endpoint for a boundary segment. The last item describes the segment as a LINE or a CURVE (with its parameters in brackets). PO denotes the segmentation point as a Orientation discontinuity point, PC as a Curvature discontinuity point, and BO as an Orientation discontinuity boundary between surfaces. The next to last line for each surface describes the surface type. The final line gives the surface normal at a nominal point on the surface in the feature's reference frame.

```
SURFACE uside =

PO/(0.0,0.0,0.0) BO/LINE
PO/(19.6,0.0,0.0) BO/LINE
PC/(61.8,7.4,0.0) BO/CURVE[7.65,0.0,0.0]
PC/(61.8,22.4,0.0) BO/LINE
PO/(19.6,29.8,0.0) BO/LINE
PO/(0.0,29.8,0.0) BO/CURVE[-22.42,0.0,0.0]
PLANE
NORMAL AT (10.0,15.0,0.0) = (0.0,0.0,-1.0);
```

Illustrated next is a portion of the robot rigid upper-arm assembly (upperarm) with its subsurfaces (e.g. uside) and the reference frame relationships between them. The first triple in the relationship is the (x, y, z) translation and the second gives the (rotation, slant, tilt) rotation. Translation is applied after rotation.

```
ASSEMBLY upperarm =

uside AT ((-17.0,-14.9,-10.0),(0.0,0.0,0.0))
uside AT ((-17.0,14.9,0.0),(0.0,π,π/2))
uendb AT ((-17.0,-14.9,0.0),(0.0,π/2,π))
uends AT ((44.8,-7.5,-10.0),(0.0,π/2,0.0))
uedges AT ((-17.0,-14.9,0.0),(0.0,π/2,3π/2))
uedges AT ((-17.0,14.9,-10.0),(0.0,π/2,π/2))
uedgeb AT ((2.6,-14.9,0.0),(0.173,π/2,3π/2))
uedgeb AT ((2.6,14.9,-10.0),(6.11,π/2,π/2));
```

The assembly that pairs the upper and lower arm rigid structures into a flexibly connected structure is shown below. Here, the lower arm has an affixment parameter that defines the joint angle in the assembly.

```
ASSEMBLY upperasm =

upperarm AT ((0.0,0.0,0.0),(0.0,0.0,0.0))
lowerarm AT ((43.5,0.0,0.0),(0.0,0.0,0.0))
         FLEX ((0.0,0.0,0.0),(jnt3,0.0,0.0));
```

Figure 5 shows an image of the whole robot assembly with the surfaces shaded according to surface orientation.

6.0 THREE DIMENSIONAL FEATURE DESCRIPTION

General identity-independent properties are needed to cue the invocation process; some properties must be extracted before enough evidence exists to suggest the identity of the object, which could then trigger model-directed description processes. Later, these properties are used to ensure that model-to-data surface pairings are correct. The use of 3D information from the surface image makes it possible to compute many object properties directly. Most of the properties measured relate to surface patches and include: local curvature, absolute area, elongation and surface intersection angles. Table 1 lists the values of these properties for the vertical robot base panel, as estimated from the test image.

TABLE 1: Properties Of Robot Base Side Panel

PROPERTY	ESTIMATED	TRUE
adjacent surface angle	4.8	4.7
adjacent surface angle	2.3	3.1
maximum surface curvature	0.127	0.111
minimum surface curvature	0.0	0.0
absolute area	1238	1413
relative area	1.0	1.0
surface eccentricity	3.3	2.0
boundary relative orientation	1.79	1.57
boundary relative orientation	1.48	1.57
number of parallel boundaries	2	2
boundary curve length	27.3	28.2
boundary curve length	46.1	50.0
boundary curve length	51.1	50.0

boundary curvature	0.038	0.11
boundary curvature	0.011	0.0
boundary curvature	0.010	0.0

Model invocation is necessary because of the many potential identities for any image structure, and because generic representation requires suggestive indexing (i.e. there may not be an exact model for the data). Invocation is based on plausibility, rather than certainty, and this notion is expressed through accumulating various types of evidence for objects in an associative network representing both direct property evidence and indirect component or generic evidence. When the plausibility of a structure having a given identity is high enough, a model is invoked.

Further discussion of model invocation can be seen in (4) and (6).

7.0 HYPOTHESIS COMPLETION

Hypothesis completion attempts to find image evidence for each feature of the invoked model. Because of the modeling assumptions, these features are surfaces and recursively defined subcomponents. Invocation provides the data for forming the initial hypothesis, which is used for estimating the 3D location and orientation. This eliminates most substructure search by directly pairing features. All other data comes from within the local surface cluster context.

Hypothesis construction requires global location and orientation estimates for hypothesis completion. The spatial relationship between structures is constrained by the geometrical relationships of the model and inconsistent data implies an inappropriate invocation or feature pairing. Object orientation is estimated by mapping the nominal orientations of pairs of model surface vectors to corresponding image surface vectors. Surface normals and curvature axes are the two types of surface vectors used. Pairs are used because a single vector allows a remaining degree of rotational freedom. Because of data errors, the six degrees of spatial freedom are represented as parameter ranges. Each new model-data feature pairing contributes new spatial information, which helps further constrain the parameter range. Translation is estimated from the allowable range of oriented model surfaces consistent with the image data.

Previously recognized substructures also constrain object position.

Table 2 lists the measured and estimated location positions, orientation angles and flexible attachment angles for the robot in the test image. This data was obtained from an image taken at about 500 cm.

TABLE 2: Measured And Estimated Spatial Parameters (In Global Reference Frame)

Parameter	Measured		Estimated	
X	488	(cm)	486	(cm)
Y	89	(cm)	85	(cm)
Z	554	(cm)	552	(cm)
Rotation	0.0	(rad)	0.24	(rad)
Slant	0.79	(rad)	0.90	(rad)
Tilt	3.14	(rad)	3.64	(rad)
Joint 1	2.24	(rad)	2.29	(rad)
Joint 2	2.82	(rad)	3.07	(rad)
Joint 3	4.94	(rad)	4.34	(rad)

A variety of model driven processes contribute to completing a hypothesis. They are, in order:

(1) decide back-facing surfaces
(2) decide tangential surfaces
(3) predict visibility of remaining surfaces
(4) search for missing surfaces
(5) bind rigidly connected subobjects
(6) bind flexibly connected subobjects
(7) explain some incorrectly segmented surfaces
(8) validate externally obscured structure

Hypothesis completion has a "hierarchical synthesis" character, where data surfaces are paired to model surfaces, surface groups are matched to assemblies and assemblies are matched to larger assemblies. The three key constraints on the matching are: (1) localization in the correct image context (i.e. surface cluster), (2) correct feature identities and (3) consistent reference frame relationships.

Adding a new surface or a rigidly connected subcomponent requires meeting only the above three requirements. Joining together two flexibly connected assemblies also gives the values of the variable attachment parameters, by unifying the respective reference frame descriptions. The parameters must also meet any specified constraints, such as on joint angles in the robot test image.

The construction process tries to find evidence for every portion of the model. Many features are paired during the invocation process. Others, such as the back of the trash can in the test image, need to be paired by a model-directed process. Given the oriented model, the image positions of unmatched surfaces can be predicted. Then, any surfaces in the general area that:

- have not already been previously used,
- belong to the surface cluster and
- have the correct shape and orientation

can be used as evidence for the unpaired model features. Later verifications ensure that correct pairings were made.

Missing structure requires understanding the three cases of occlusion, predicting or detecting its occurrence and showing that the image data is consistent with the expected visible portion of the model. The easiest case of back-facing and tangent surfaces can be predicted using the orientation estimates with known observer viewpoint and the surface normals deduced from the geometrical model. A raycasting technique (i.e. predicting an image from an oriented model) handles self-obscured front-facing surfaces by predicting the location of obscuring surfaces and hence which portions of more distant surfaces are invisible. The final case occurs when unrelated structure obscures portions of the object. Assuming enough evidence is present to invoke and orient the model, occlusion can be confirmed by finding other closer unrelated surfaces responsible for the missing image data.

The self-occlusion visibility analysis for the trash can in the scene is given in table 3. The results are correct. Minor prediction errors occur at edges where surfaces do not meet perfectly.

TABLE 3: PREDICTED TRASH CAN VISIBILITY

SURFACE	VISIBLE PIXELS	OBSC'D PIXELS	TOTAL PIXELS	VISIBILITY
outer front	1479	8	1487	full
outer back	1	1581	1582	back-facing
outer bottom	5	225	230	back-facing
inner front	0	1487	1487	back-facing
inner back	314	1270	1584	partial-obsc
inner bottom	7	223	230	full-obsc

Figure 6 shows the boundaries of the found portions of the object model as predicted by the orientation parameters and superposed over the original intensity image. No hidden line removal was used. Because of minor and cumulative errors from the robot's base position, the position of the lower arm is somewhat away from its observed position. However, when it was initially recognized, its position was closer. Further, the picture shows that the global understanding is correct. In analysis, all features were correctly paired, predicted invisible or verified as externally self-obscured. The numerical results in table 2 also show good performance.

8.0 HYPOTHESIS VERIFICATION

The final step in the recognition process is verification. Verification ensures that instantiated hypotheses are valid physical objects and have the correct identity (i.e. have all object properties). This is necessary because model invocation suggests a particular object, which then acquires rough model instantiation and orientation. It is then necessary to verify the details for correctness. A proper, physical, object is more certain if all surfaces are connected and they enclose the object. Correct identification is more likely if all model structure is accounted for, the model and corresponding image surface shapes and orientations are the same, and the model and image surfaces are connected similarly. The constraints used to ensure correct identities in the test image were:

(for surfaces)
- has approximately correct size
- has approximately correct curvature class

(for solids)
- has no duplicated use of image data
- all predicted back-facing surfaces have no data
- all adjacent data surfaces are adjacent in model
- all subfeatures have correct orientation
- all features predicted as partially self-obscured during raycasting are observed as such (i.e. have appropriate occluding boundaries)

In the example given above, all correct object hypotheses passed these constraints. The only spurious structures to pass verification were single surfaces similar to the correct surface or symmetric subcomponents.

9.0 DISCUSSION

This recognition process was clearly successful on the test image. However, much research is still needed. Objects were represented here at only a single level of scale, but feature descriptions change as a function of observer distance, with larger features dominating at greater distances. The surface data needed to be partly generated by hand, because no surface information was available here. Further, the theory on surface segmentation and description is not well advanced yet. The recognition process is also very slow at present, preventing practical application.

The success of the research lies in demonstrating the ease with which complete explainable object recognition can be achieved using surface information. Further, the type of object fully recognized is significantly more complicated than previously possible (because of its multiple articulated features, curved surfaces, self-occlusion and external occlusion). Further information on this work can be found in (4) - (7).

10.0 BIBLIOGRAPHY

(1) Brady, M., Ponce, J., Yuille, A., Asada, H., "Describing Surfaces", Proc. 2nd Int. Symp. on Robotics Research, 1984.

(2) Brooks, R.A., "Symbolic Reasoning Among 3D Models And 2D Images", AIM-343, Computer Science, Stanford University, 1981.

(3) Faugeras, O.D., Hebert, M., "A 3-D Recognition and Positioning Algorithm Using Geometric Matching Between Primitive Surfaces", Proc. 8th Int. Joint Conf. on Atrif. Intel., pp996-1002, 1983.

(4) Fisher, R. B., "From Surfaces to Objects: Recognizing Objects Using Surface Information And Object Models", PhD Thesis, Dept. of Artificial Intelligence, Univ. of Edinburgh, in preparation.

(5) Fisher, R. B., "Discontinuity Segmentation Of Curves And Surfaces", Working Paper, Dept. of Artificial Intelligence, Univ. of Edinburgh, in preparation.

(6) Fisher, R. B., Orr, M. J. L., "Spatial Configurations in the Model Invocation Process", awaiting publication in "Image and Vision Computing".

(7) Fisher, R. B., "Using Surfaces And Object Models To Recognize Partially Obscured Objects", Proc. 8th Int. Joint Conf. on Atrif. Intel., pp989-995, 1983.

(8) Guzman, A., "Decomposition of a Visual Scene Into Bodies", MIT AI memo #139, 1967.

(9) Marr, D., "Vision", W.H. Freeman and Co. (pub), 1982.

(10) Sugihara, K,, "Range-Data Analysis Guided by a Junction Dictionary", Artificial Intelligence, Vol 12, pp41-69, 1979.

(11) Waltz, D., "Understanding Line Drawings of Scenes with Shadows", in Winston (ed), "The Psychology of Computer Vision", pp19-91, 1975.

Acknowledgements

Portions of the work were supported by a University of Edinburgh studentship. Thanks go to M. Orr for advice with the paper and many others in the department for advice on the research.

Figure 1: Test Scene

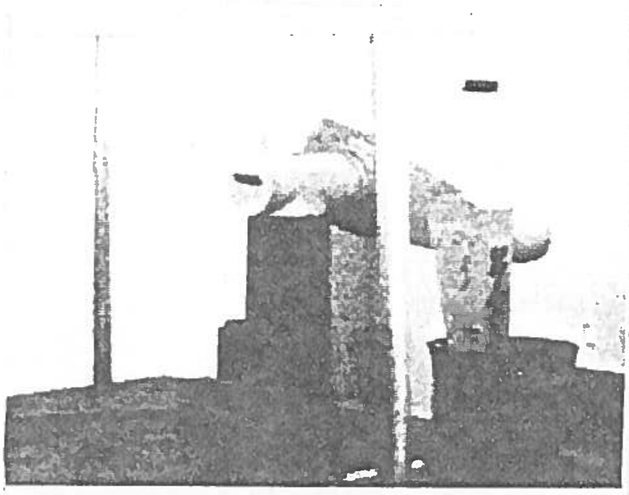
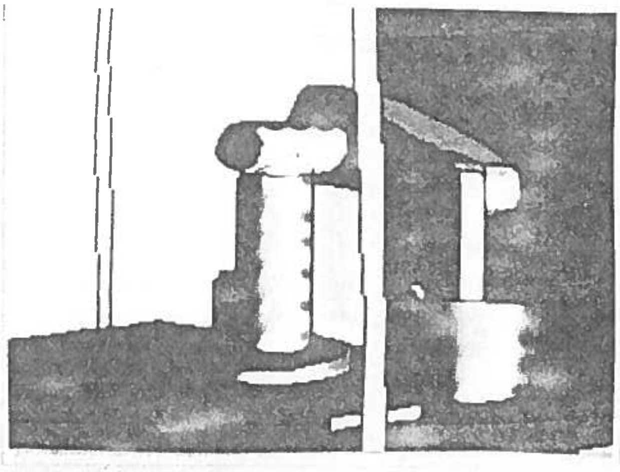


Figure 2: Input Data For Test Scene

(a) Cosine Of Surface Slant



(b) Obscuring Boundaries

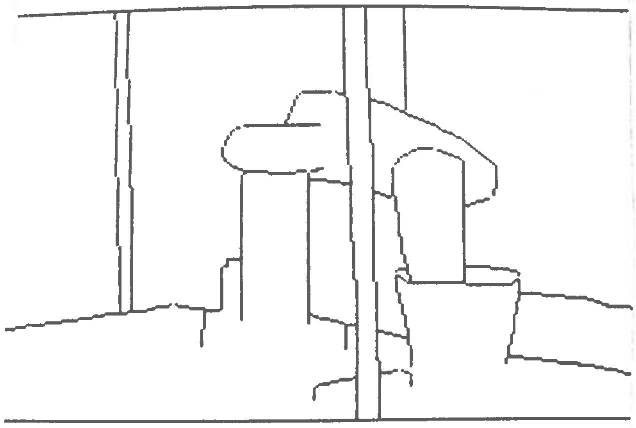


Figure 3: Original and Reconstructed Robot Upper Arm Surface

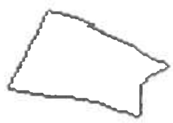


Figure 4: Some Connected Surface Groups

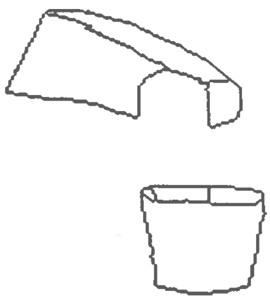


Figure 5: Shaded View of Robot Model

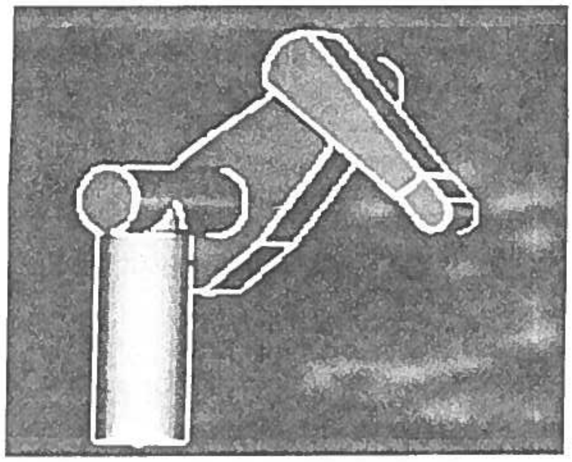


Figure 6: Predicted Surface Boundaries For Found Robot Assembly