

Recognition with Second-Order Topographic Surface Features

Robert B. Fisher

Dept. of Artificial Intelligence, University of Edinburgh
5 Forrest Hill, Edinburgh EH1 2QL, Scotland, United Kingdom

Abstract

We previously[3] defined eight second-order volumetric primitives and then showed[4] that they can be extracted from range data. This paper shows that by using them model matching is more efficient, because the shape vocabulary reduces the combinatorial generation of hypotheses. With the model-to-data correspondences, accurate three dimensional model location is possible.

1 Introduction

While there is already much research on object recognition using 3D surface patches, there is less using volumetric features (e.g. [1], [6], [7]). We have investigated the representation[3] and extraction[4] of second order volumetric features from range data. These features define small shapes that embellish the surface of objects, but are too small to be described by the surface patch methods. They are "second-order" in two senses: (1) they are smaller features that add detail to the surface, rather than specify the overall shape, and (2) they denote more specific, higher-level descriptive shapes. The positive (extruding) features are the BUMP, SPIKE, RIDGE and FIN and the negative (intruding) features are the DENT, HOLE, GROOVE and SLOT.

Features extracted according the previously reported procedure were labeled, but no symbolic description was made. Each feature is defined with respect to a local reference frame, and can be used in a scene description containing surface patches, volumetric groupings and second-order volumetric features. With symbolic descriptions of both the image data and the model, establishing model-to-data pairings is more efficient, because the typing of the features reduces the potential for combinatorial explosion.

2 Describing The Features

Previous research[4] classified extracted features based on their length, depth and whether they extended into or out of the surrounding surface. The shape and position parameters of the classified feature are now used to create a symbolic model parameterized by the length (L), width (W) and depth (D) parameters:

BUMP:	HEIGHT	D	DENT:	DEPTH	D
	MAJOR_RADIUS	L/2		MAJOR_RADIUS	L/2
	MINOR_RADIUS	W/2		MINOR_RADIUS	W/2

SPIKE:	LENGTH	D	HOLE:	LENGTH	D
	CROSS_RADIUS	L/2		CROSS_RADIUS	L/2
RIDGE:	LENGTH	L	GROOVE:	LENGTH	L
	CROSS_RADIUS	W/2		CROSS_RADIUS	W/2
FIN:	LENGTH	L	SLOT:	LENGTH	L
	CROSS_RADIUS	W/2		CROSS_RADIUS	W/2
	HEIGHT	D		DEPTH	D

After creating individual model primitives, an ASSEMBLY[2] that contains the features is created by estimating the reference transformations that map the primitive feature's local reference frames onto that of the ASSEMBLY. The reference frame estimate is based on the L , W and D parameters plus: the minimum range of the feature from the viewer (Z_0), an estimate of the X-Y image plane orientation (θ) of elongated features, and the (X, Y, Z) of the 3D center-of-mass.

The orientation is estimated by using the central moments μ_{pq} :

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right)$$

and the rest are extracted directly from the data. The position is based on a RST(rotation, slant, tilt) rotation[2] and a TRANS(x, y, z) translation specification. The position specification for each feature type is:

BUMP: RST(0, 0, 0) TRANS($X, Y, Z_0 + (L + W)/4$)

DENT: RST(0, 0, 0) TRANS($X, Y, Z_0 - (L + W)/4$)

RIDGE: RST($\theta, 0, 0$) TRANS($X - L\cos(\theta)/2, Y - L\sin(\theta)/2, Z_0 + W/2$)

GROOVE: RST($\theta, 0, 0$) TRANS($X - L\cos(\theta)/2, Y - L\sin(\theta)/2, Z_0 - W/2$)

FIN: RST($\theta, \frac{\pi}{2}, \frac{\pi}{2}$) TRANS($X - L\cos(\theta)/2, Y - L\sin(\theta)/2, Z_0 + D/2$)

SLOT: RST($\theta, \frac{\pi}{2}, \frac{\pi}{2}$) TRANS($X - L\cos(\theta)/2, Y - L\sin(\theta)/2, Z_0 - D/2$)

SPIKE: RST(0, $\frac{\pi}{2}, \pi$) TRANS($X, Y, Z_0 + D$)

HOLE: RST(0, $\frac{\pi}{2}, 0$) TRANS($X, Y, Z_0 - D$)

3 Object Recognition Using the Second-Order Features

Most model matching algorithms involve an element of interpretation tree generation, whereby candidate models are paired with data features (or *vice-versa*). This process often leads to combinatorial explosions; hence much work has involved using local constraints to prune unlikely interpretations early[5]. The main causes of the problem are: (1) ordinarily, each data feature can have a large number of matching model features and (2) the use of a "wild-card" to allow unmatched features (as segmentation can fail in the presence of noise). If there are D data features, and M model features, and if each model feature can be labeled as being one of the data features or 'none' then approximately

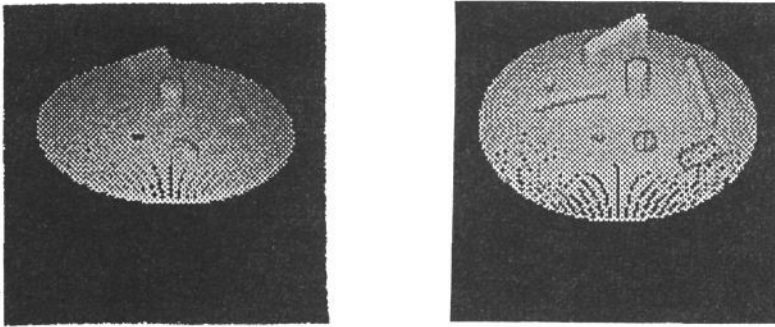


Figure 1: Symbolic Model Constructed from Extracted Features (left is automatic, right is by hand)

$(D + 1)^M$ possible feature labelings are considered in the standard algorithm (i.e. before the use of constraints to reduce this[5]). However, if the data and model features are typed, and matching occurs only between features of the same type, then the number is reduced. If there are approximately equal numbers of T types (here $T = 8$), then the number of hypotheses is reduced to $(\frac{D}{T} + 1)^M$ or T^M times smaller.

We use the standard interpretation tree search algorithm, with a test for global geometric consistency at the end. Model and data features of the same type are compatible if their non-nominal shape parameters are compatible. Pairwise consistency constraints (distance between features) prune geometrically inconsistent pairings. All valid ASSEMBLY interpretations must have data features explaining at least 40% of the model features and no interpretations are subsets of previously verified interpretations. The estimated object rotation is the transformation that maximizes the pairwise alignment of the model and data direction vectors formed between pairs of features. A verification phase ensures that the transformed model features lie close to the data features and that elongated features have proper alignment.

4 Experiments

We previously[4] reported the features extracted from a hill shaped test scene with a variety of surface features. The symbolic model scene model automatically constructed for this test scene is shown in the left half of Figure 1. Comparison with the hand made model shown in the right half shows that a decent description has been constructed.

Applying the matching process using the model on the test scene, 1113 nodes were generated in the interpretation tree (as compared to a maximum of approximately 16 million = $(7 + 1)^8$ nodes), of which 127 remained after parameter compatibility testing and only one complete set of pairwise consistent feature correspondences reached the geometric verification stage. All data features were correctly paired with their corresponding model features. Figure

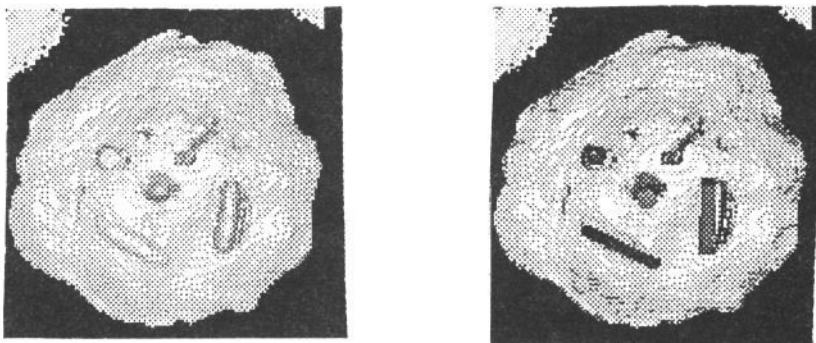


Figure 2: Test Scene and Model in Estimated Position

2 shows the estimated position of the model superimposed in black on the raw data. (Only positive features are shown because the display program could not produce images of negative features in isolation.) When using other models, no sets of consistent feature correspondences with enough pairings were generated.

The low number of hypotheses generated shows the effective use of strong typing and the perfect verification results from the strong constraint of consistent global geometry. Hence, the second-order volumetric features can be used as a reasonable basis for object recognition.

References

- [1] Brooks, R. A., "Symbolic reasoning among 3-D models and 2-D images", *Artificial Intelligence*, Vol 17, pp 285-348, 1981.
- [2] Fisher, R. B., "SMS: A Suggestive Modeling System For Object Recognition", *Image and Vision Computing*, Vol 5, No 2, pp 98-104, 1987.
- [3] Fisher, R. B., "Modeling Second-Order Volumetric Features", *Proc. 3rd Alvey Vision Conference*, pp 79-86, 1987.
- [4] Fisher, R. B., "Extracting Second-Order Topographic Surface Features From Range Data", *Proc. 1st British Machine Vision Association Conference*, pp 241-246, Oxford, 1990.
- [5] Grimson, W. E. L., Lozano-Perez, T., "Model-Based Recognition and Localization from Sparse Range or Tactile Data", *International Journal of Robotics Research*, Vol. 3, pp 3-35, 1984.
- [6] Pentland, A. P., "The Parts of Perception", in Brown (ed), *Advances in Computer Vision*, Vol 2, pp 121-176, Lawrence Erlbaum Associates, New Jersey, 1988.
- [7] Shapiro, L., Moriarty, J., Mulgaonkar, P., Haralick, R., "Sticks, Plates, And Blobs: A Three-Dimensional Object Representation For Scene Analysis", *Proc. Amer. Assoc. for Artificial Intelligence*, pp 28-30, 1980.

This work was performed under funding from the University of Edinburgh.