Hierarchical Integration of Local 3D Features for Probabilistic Pose Recovery

Renaud Detty  Justus Piater

Dept. of Electrical Engineering and Computer Science, University of Liège, Belgium

Abstract
We address the problem of 3D object representation. We develop a hierarchical model based on probabilistic correspondences and probabilistic relations between 3D visual features. Features at the bottom of the hierarchy are bound to local observations. Pairs of features that present strong geometric correlation are iteratively grouped into higher-level meta-features that encode probabilistic spatial relationships between their children. The model is instantiated by propagating evidence up and down the hierarchy using a Belief Propagation algorithm, which infers the pose of high-level features from local evidence and reinforces local evidence from globally consistent knowledge. We demonstrate how to use our framework to estimate the pose of a known object in an unknown scene, and provide a quantitative performance evaluation on synthetic data.

1 Introduction
Objects can be characterized by configurations of parts:
• Richer than purely geometric models;
• More expressive than methods purely based on codebook vector relationships;
• More robust than methods based on global appearance.

We have recently presented a framework for unsupervised learning of hierarchical representations that combines local visual appearance descriptors and probabilistic spatial relationships [2].

We are currently developing an extension of the model that:
1. Makes use of a different underlying probability framework to support non-Gaussian relationships;
2. Supports 3D features;
3. Supports multi-modal (visual, haptic) features.

We thus intend to combine visual descriptors with haptic and proprioceptive information, which will be directly applicable to robotic tasks such as grasping and object manipulation.

In this paper, we concentrate on 3D visual features. An observation is an oriented patch in 3-space, annotated by various visual appearance characteristics.

We illustrate our model on the application of object pose estimation.

2 Hierarchical Model
Our object model consists of a set of generic features organized in a hierarchy. Features that form the bottom level of the hierarchy, referred to as primitive features, are bound to local observations. The rest of the features are meta-features which embody spatial configurations of more elementary features.

2.1 Model Parametrization
The hierarchy is implemented using a Pairwise Markov Random Field. The relationship between a meta-feature and one of its children \( z \) is parametrized by a compatibility potential function \( \psi_{ij}(x_i,x_j) \) associated to the edge \( ij \). A compatibility potential specifies, for any given pair of poses of the features it links, the probability of finding that particular configuration for these two features. The statistical dependency between a hidden variable \( z_i \) and its observed variable \( y_i \) is parametrized by an observation potential \( \phi_i(z_i) \).

2.2 Model Instantiation
Model instantiation provides pose densities for all features of the model, indicating where the learned object is likely to be present. It amounts to:
1. Estimate evidence for primitive features from their observations using the codebook;
2. Infer posterior marginal densities for all features of the hierarchy with any inference mechanism.

We present a probabilistic framework for hierarchical 3D object representation. This framework can be used to compute a pose for a known object in an unknown scene without explicit point correspondences, which we demonstrated through a series of experiments on synthetic data. Our framework allows the natural integration of non-visual features and can potentially support cross-modal representations useful for robotic grasping, which we will explore in future work.

3 Pose Estimation
Objects are learned from a reference scene, in a reference pose. To estimate the pose of a known object in an unknown scene, we instantiate the object model in that scene. We then compare the top-level feature poses in the unknown scene to those in the reference scene; the transformation between them is the pose of the object in the unknown scene.

4 Conclusion
We presented a probabilistic framework for hierarchical 3D object representation. This framework can be used to compute a pose for a known object in an unknown scene without explicit point correspondences, which we demonstrated through a series of experiments on synthetic data. Our framework allows the natural integration of non-visual features and can potentially support cross-modal representations useful for robotic grasping, which we will explore in future work.

Acknowledgment
This work was supported by the Belgian National Fund for Scientific Research (FNRS) and the EU Cognitive Systems project PACO-FPLUS (IST-FP6-IP-027867). The authors thank Norbert Krüger and Nicolas Pugeault for providing Molis data and for fruitful discussions.

References

Fig. 1: A Pairwise Markov Random Field representing a feature hierarchy. Features correspond to hidden variables (white). Observed variables (black) compared to observations, bound to bottom-level primitive features.

Fig. 2: Effect of the message product in BP. Messages are in dark red and dark yellow, the product is in purple.

Fig. 3: Synthetic observations of objects in a reference pose on the left, clustered scenes on the right. Color indicates the different primitive feature classes. The plots show in red the error of the translation (relative to the object size) and rotation (in degrees) estimates of the object poses in their clustered scenes as a function of the number of levels.