Motivation: recognizing parts in 3D shapes is fundamental to several applications in 3D computer vision, computer graphics, and robotics. Earlier work: “hand-engineered” geometric descriptors, heuristic processing stages, etc. In upper layers, different filters are sensitive to different local surface patterns (triangular, circular patches etc.). In upper layers, different filters are sensitive to different local surface patterns.

Overview

Motivation: recognizing parts in 3D shapes is fundamental to several applications in 3D computer vision, computer graphics, and robotics. Earlier work: “hand-engineered” geometric descriptors, heuristic processing stages, etc. In upper layers, different filters are sensitive to different local surface patterns (triangular, circular patches etc.). In upper layers, different filters are sensitive to different local surface patterns.

Challenges: subtlety in 3D geometric cues, arbitrary orientation, noise, varying resolution, arbitrary or no interior, missing texture, non-manifold geometry, shape part variability, need to parse local and global context.

Our approach: combine fully convolutional net (FCN) operating on rendered shape views with surface-based graphical model (CRF)

Method

Rendering stage: infer set of viewpoints that maximally covers the surface of the input shape across multiple scales. To favor rotational invariance, perform in-plane camera rotations.

Views are not ordered, number of viewpoints differ per shape, and no view correspondence across shapes are assumed.

Encode surface position & normals: render shaded images (normal dot view vector) and depth images relative to the cameras.

Render surface reference images: each pixel stores a pointer to a surface element.

ShapePFCN architecture: end-to-end trainable and analytically differentiable.

Key ideas:
- Adaptive view selection per shape to maximally cover its surface
- Multi-scale representation of the surface information
- Initialize network from pre-trained image-based architectures
- End-to-end training of the whole network (FCN & CRF)
- Projective layer for mapping view representations to surfaces

Key advantages:
- High-resolution shape analysis
- Robustness to geometric representation artifacts (noise, irregular tessellation, arbitrary interior, non-manifold geometry)
- Transfer learning from massive image datasets
- Rotational invariance
- CNN representation power is focused on the shape surface

Projects:
- ShapeBoost
- ShapePFCN

Notes: per category training, 50% training / 50% testing, max 500 shapes per class, no assumption on shape orientation

Results

ShapeBoost (16 classes), L-PSB & COSEG (30 classes)

Average labeling accuracy on segmented ShapeNetCore

Experiments: 3D ShapeNet (16 classes), L-PSB & COSEG (30 classes)

Note: per category training, 50% training / 50% testing.

Project page with datasets, results and source code: http://people.cs.umass.edu/~kalo/papers/shapepfcn/index.html