Qualitative 3D Surface Reconstruction from Images

Alex Levinshtein[1], Cristian Sminchisescu[2], Sven Dickinson[1]
[1]University of Toronto, [2]University of Bonn

Prior to the advent of appearance-based recognition in the early 1990's, object categorization researchers modeled the prototypical shape of an object, seeking models that were invariant to changes in color, texture, and minor within-class shape deformation. While these categorical models were well-motivated, they could not be reliably recovered from real images of real objects, and eventually gave way to models based on recurring, local, appearance-based features. But while appearance yields powerful exemplar-based recognition features, it is seldom generic to a category.

In this work, we use registered optical and range images of objects to learn a mapping from region appearance to qualitative surface shape. At training time, from a set of extracted regions, we compute features describing both the external region boundary as well as the internal region appearance. The corresponding regions in the registered range image can be fit with second-order surfaces, from which mean (H) and Gaussian (K) curvatures can be computed. The signs of these curvatures, in turn, specify one of 8 possible qualitative surface labels [1]. The set of superpixels, paired with surface labels, forms our training set, and is used to learn the parameters of a conditional random field that accounts for both spatial constraints and image evidence. Figure 1 illustrates our operating regime: the qualitative HK labels of the vocabulary, the superpixel random field, and the surface labeling corresponding to the superpixels.

Learning 3-D shape from realistic images has been attempted recently. Hoeim et al. [3] and Saxena et al. [2] learn to segment an image into planar structures that are suitable for navigation or simple contextual analysis. In contrast, we aim at a richer representation that better supports object recognition, recovering a mid-level, qualitative shape description of an object without invoking an object-specific model, such as a teapot or human. Even a sparse labeling of the shapes of an object's surfaces can yield a powerful set of viewpoint-invariant shape indices for object categorization. We

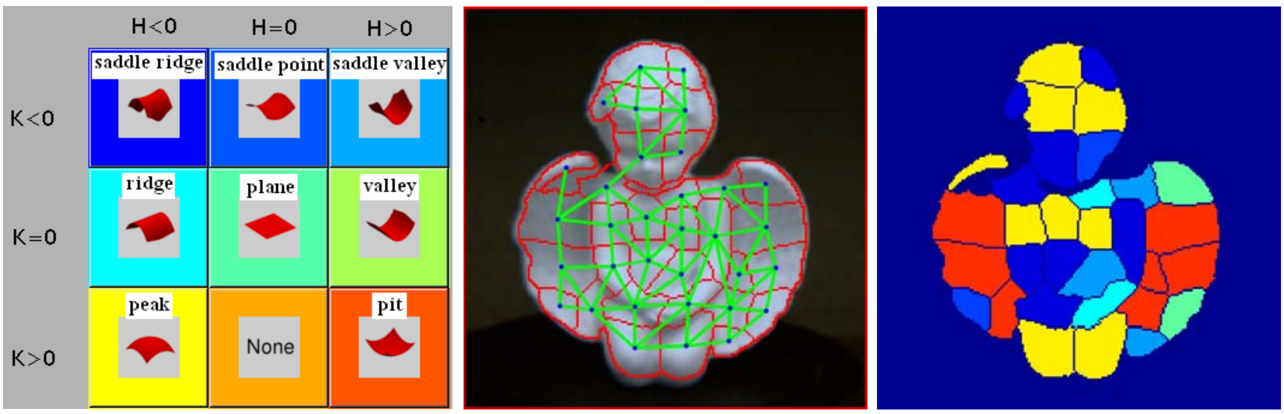

Figure 1: *Left:* Qualitative HK surface curvature labels. *Middle:* (Image) Superpixel graph; *Right:* Surface labeling computed either from the registered 3-D range image during training (ground truth), or from superpixel data, using CRF inference, during testing.

model the surface labels of image superpixels as a conditional random field (CRF), with nodes corresponding to image superpixels and edges spanning superpixels that share a boundary (Figure 1 (center)). Let $\mathbf{X}$ be a set of superpixel labels in the image, K the number of possible shape labels, x_i the label of superpixel i, and $\mathbf{F}_i$ the vector of image features corresponding to superpixel i (all stored in a matrix $\mathbf{F}$). The probability of the labels in the CRF model is:

$$p(\mathbf{X}|\mathbf{F}, \mathbf{w}, \mathbf{v}) = \frac{1}{Z(\mathbf{w}, \mathbf{v})} \exp\{\sum_i \mathbf{w}_i^\top \cdot \mathbf{F}_i + \sum_{ij} \mathbf{v}(i,j)\} \tag{1}$$

where $\mathbf{w}$ and $\mathbf{v}$ are the parameters of the potentials. If S is the number of features per site, then $\mathbf{w}$ is a $K \times S$ matrix of coefficients with $\mathbf{w}_k$ representing the k-th row corresponding to class k, and $\mathbf{v}$ is a symmetric $K \times K$ matrix of binary potentials. The unary potentials associate image features with different surface types, whereas binary potentials model correlations between neighboring surface types (acting as a smoothness prior, as well). We train the model using Conditional ML, and use a Bethe free energy approximation to the partition function in order to calculate the log likelihood and its gradient. Loopy belief propagation is used for inference.

We use a variety of both interior and exterior superpixel features, making up a total of 36 features ($\mathbf{F}_i$). On the interior image data, we compute the responses of local jets [4] as well as a 3x3 HK histogram computed from the surface approximation of the intensity data, while along the exterior, we compute a histogram of boundary curvature weighted by the gradient magnitude across the boundary and normalized by contour length. These are kernelized using dot-products $\mathbf{F}_i \cdot \mathbf{F}_i^\top$ yielding a lower triangular matrix (including the diagonal) of 666 features. Our initial experiments with simple feature selection methods indicate that HK histograms seconded by superpixel-based curvature features were among the most informative features.

We test our algorithm on two data sets of registered range and optical images. The first is obtained using a Minolta range scanner (OSU) and the second is the K2T structured light data set from USF. The Minolta data set contains 50 images and the K2T data set contains 40 images, split in half into training and testing. Ground truth labels for superpixels are generated automatically by analyzing the quadratic surface fitted to the corresponding superpixel's range data. We merge all saddles into one class (due to the difficulty in discriminating among different saddle classes) and obtain 6 possible surface labels. Table 1 reports the precision of our algorithm on both the training and test set for the two datasets. Our preliminary results indicate that better than chance performance can be achieved.

	Train	Test
Minolta	0.60086	0.48203
K2T	0.47739	0.32429

Table 1: Precision of the qualitative surface labeling.

Topic: visual processing and pattern recognition **Preference: poster**

References

[1] P. J. Besl and R. C. Jain, Segmentation through Variable-Order Surface Fitting. *PAMI*, 10(2), 167–192, 1988.

[2] Ashutosh Saxena, Min Sun, and Andrew Y. Ng, Learning 3-D Scene Structure from a Single Still Image. *ICCV workshop on 3D Representation for Recognition (3dRR-07)*, 2007.

[3] D. Hoiem, A.A. Efros, and M. Hebert, Recovering Surface Layout from an Image. *IJCV*, 75(1), 2007.

[4] J. J. Koenderink and A. J. van Doorn, Representation of local geometry in the visual system. *Biological Cybernetics*, 55(6), 367–375, 1987.