A Topological Approach to Hierarchical Segmentation using Mean Shift

Sylvain Paris and Frédo Durand

ABSTRACT

Mean shift is a popular method to segment images and videos. Pixels are represented by feature points, and the segmentation is driven by the point density in feature space. In this paper, we introduce the use of Morse theory to interpret mean shift as a topological decomposition of the feature space into density modes. This allows us to build on the watershed technique and design a new algorithm to compute mean-shift segmentations of images and videos. In addition, we introduce the use of topological persistence to create a segmentation hierarchy. We validated our method by clustering images using color cues. In this context, our technique runs faster than previous work, especially on videos and large images. We evaluated accuracy with a classical benchmark which shows results on par with existing low-level techniques, i.e. we do not sacrifice accuracy for speed.

CONTRIBUTIONS

We describe a fast method to compute a hierarchical segmentation of large, low-dimensional data sets, such as high-resolution digital photographs and videos. We base our work on a new interpretation of the mean-shift algorithm as a topological decomposition of the feature space.

DEFINITION: MEAN-SHIFT SEGMENTATION

Each pixel is associated to a feature point "position & color".

$$X_i = (x_i, y_i, L_i, a_i, b_i)$$

Iteration: Each point moves to the mean of its neighbors.

$$y_j+1 = \sum_{i=1}^{n} K (y_j - x_i) \times x_i$$

Known result: This is an ascent process on the feature point density.

$$D(p) = \sum_{i=1}^{n} K(p - x_i)$$

Segmentation: Points with the same limit are grouped.

RUNNING TIMES (8 Mpixels)

1- Compute the density function on the whole feature space.
2- Extract the density modes.
3- Build a hierarchy using topological persistence.

FAST COMPUTATION OF THE FEATURE POINT DENSITY

We estimate the density function on a regular grid in feature space. Since we use a Gaussian density estimator, the density function is band-limited. Thus, we can use a coarse grid to sample the feature space. Performances are further improved using a separable Gaussian convolution.

FAST EXTRACTION OF THE DENSITY MODES

We use an algorithm akin to watershed. We process grid nodes from high density to low density. We assign labels to a node depending on the number of different labels present in its 1-neighborhood.

0 label: The processed node is a local maximum. We create a new label.
1 label: All the ascending paths go to the same summit. We copy the label.
2+ labels: Boundary between two modes. We put a special marker.

RESULTS

Our algorithm works well for low-dimensional feature spaces and large kernels.

The Berkeley benchmark shows an accuracy on par with existing methods using color cues.

We process 6 seconds of video in 5min51 with 5D feature points, in 40s with 4D feature points. Previous work needs about 10min.