A Topological Approach to Hierarchical Segmentation using Mean Shift

Sylvain Paris and Frédo Durand

Massachusetts Institute of Technology
Computer Science and Artificial Intelligence Laboratory

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”.

$$\mathbf{x}_i = (x_i, y_i, L_i, a_i, b_i)$$

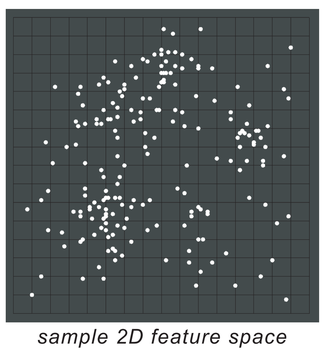
Iteration: Each point moves to the mean of its neighbors.

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^{n} K(\mathbf{y}_j - \mathbf{x}_i)\, \mathbf{x}_i}{\sum_{i=1}^{n} K(\mathbf{y}_j - \mathbf{x}_i)}$$

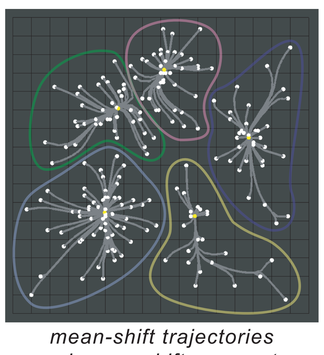
Known result: This is an ascent process on the feature point density.

$$D(\mathbf{p}) = \sum_{i=1}^{n} \tilde{K}(\mathbf{p} - \mathbf{x}_i)$$

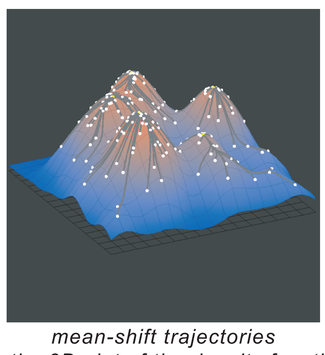
Segmentation: Points with the same limit are grouped.



sample 2D feature space



mean-shift trajectories and mean-shift segments



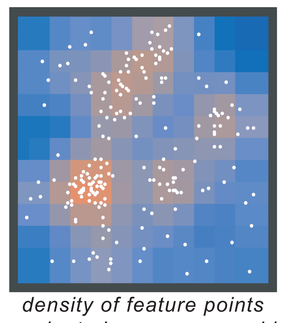
mean-shift trajectories on the 3D plot of the density function

OUR STRATEGY

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.

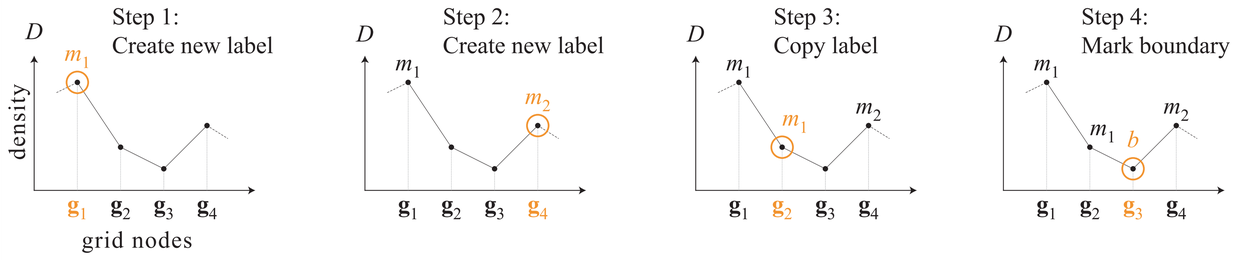


density of feature points evaluated on a coarse grid

FAST EXTRACTION OF THE DENSITY MODES

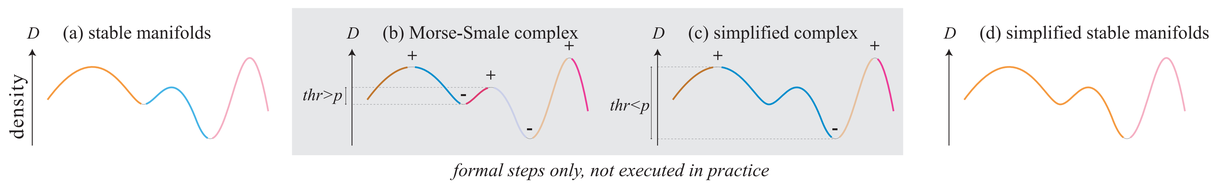
We use an algorithm akin to watershed. We process grid nodes from high density to low density. We assign labels to a node depending on the number of different labels present in its 1-neighborhood.

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.



BUILD A HIERARCHY

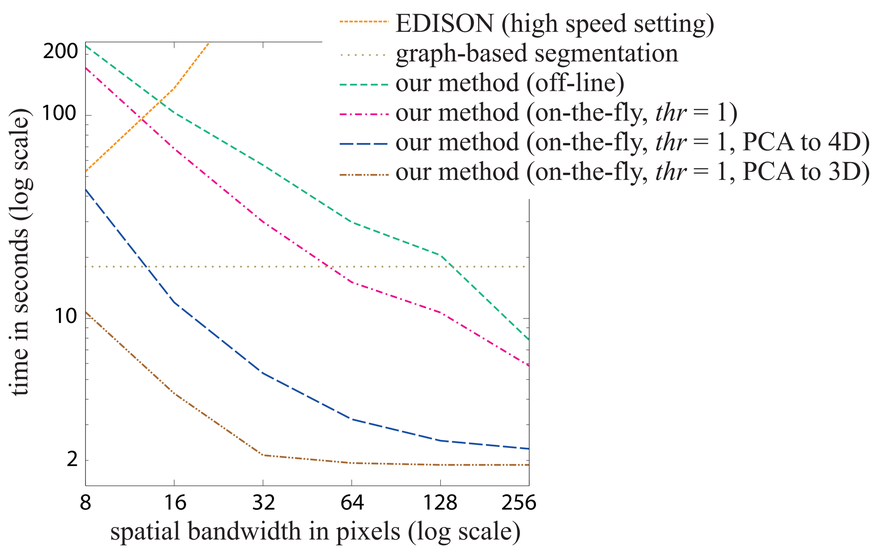
Persistence characterizes the importance of topological features. For a mode, it is defined as the difference between the height of its summit and the height of a saddle point linking to an adjacent mode. A hierarchy is built by merging adjacent modes. Low persistence modes are merged first.



formal steps only, not executed in practice

We prove that this is equivalent to simplifying the underlying Morse-Smale complex, and that mergers can be done on-the-fly.

RUNNING TIMES (8 Mpixels)



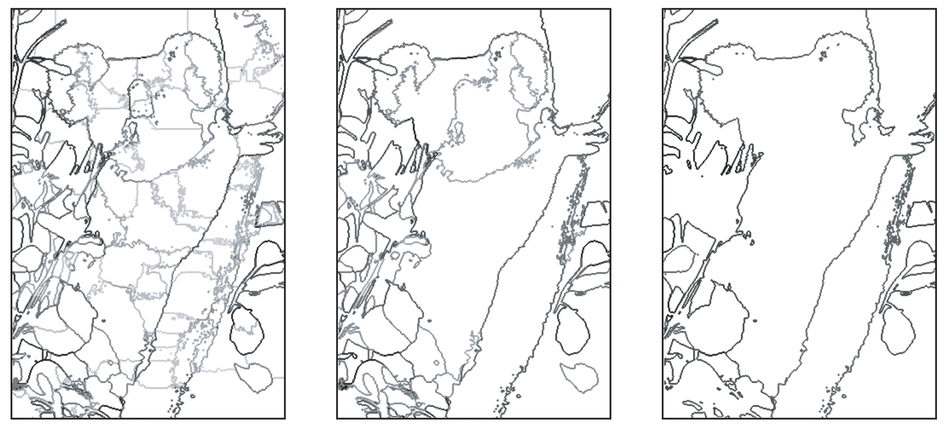
Our algorithm works well for low-dimensional feature spaces and large kernels.

RESULTS

Input

Sample levels of the hierarchy

Our algorithm needs to be run only once to obtain all levels.



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.