

Supplemental Material for “A Video Representation Using Temporal Superpixels”

In the following supplemental material, we expand on five concepts presented in the main paper. In Section A, we describe the digital topology constraints used in the temporal superpixel (TSP) representation. Next, in Section B, we derive the optimal mean parameters for new and old superpixels. The impact of our optimization procedure is then briefly discussed in Section C. In Section D, we consider the image boundary effects in more detail and present our solution to the problem. Finally, we discuss details of the metrics used in the paper in Section E. Additional video results are also included in the supplement.

A Topology Constraints

In typical definitions of digital topology, there are distinct foreground (FG) and background (BG) regions with associated neighborhood connectivity. These connectivities are defined in a pair to avoid topological paradoxes in the implicit continuous shapes [4]. In 2D, one can choose a 4-connected FG and 8-connected BG or an 8-connected FG and 4-connected BG. Figure 1 illustrates one situation where using the same connectivity for both the foreground and background is inconsistent with the underlying continuous curves. In particular this example violates the digital topology principle that two disjoint regions must be separated by one connected region.

Given a particular connectivity, a pixel can move from one region to the other while preserving the topology if and only if it is a *simple point* [1]. As shown in [1], checking if a pixel is a simple point can be done in constant time. This type of concept has successfully been applied to level set based segmentation [3]. However, because of the necessity to define a pair of neighborhood connectivities, M -ary topology is not well defined.

In this work, we are not concerned about the underlying continuous curves. Rather, we use the ideas stemming from digital topology to develop a probabilistic model of superpixels. While the superpixel labels, z , in SLIC were independent, the desired topology of the labels implies that they should be *dependent*. We restrict the distribution of labels such that each unique label is a single 4-connected object. All other configurations have zero probability.

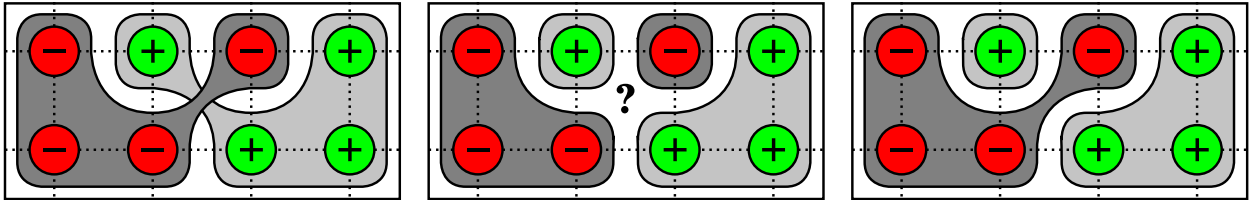


Figure 1: (Left) Topological paradox because the curves overlap. (Center) Topological paradox because part of the image does not belong to the FG or BG. (Right) the pair of different connectivities fixes the problem.

B Optimal Mean Parameters and Joint Likelihoods

New superpixels have a uniform prior distribution over the mean parameters which result in the optimal means equaling the empirical means. In the case of old superpixels, however, the expression for the optimal mean is slightly more complicated. We find the optimal mean parameters and the joint likelihoods of Equations 6-9 and 19-21 in this section.

B.1 New Superpixels

We first note that the likelihood of N Gaussian observations, x_1, \dots, x_N only depends on the joint statistics, $t = \sum_i x_i$ and $T = \sum_i x_i^2$, and can be written as

$$\prod_{i=1}^N \mathcal{N}(x_i; \mu, \sigma^2) = \mathcal{N}\left(\mu; \frac{t}{N}, \frac{\sigma^2}{N}\right) N^{-\frac{1}{2}} (2\pi\sigma^2)^{\frac{1-N}{2}} \exp\left[\frac{\frac{t^2}{N} - T}{2\sigma^2}\right]. \quad (\text{B.1})$$

In the case of new superpixels, the optimal mean is the empirical mean and the joint log likelihood of parameters and observations can be expressed as

$$\begin{aligned} \log p(x, \mu) &\stackrel{C}{=} \log \left[\mathcal{N}\left(\frac{t}{N}; \frac{t}{N}, \frac{\sigma^2}{N}\right) N^{-\frac{1}{2}} (2\pi\sigma^2)^{\frac{1-N}{2}} \right] + \frac{\frac{t^2}{N} - T}{2\sigma^2} \\ &= \log \left[\sqrt{\frac{N}{2\pi\sigma^2}} N^{-\frac{1}{2}} (2\pi\sigma^2)^{\frac{1-N}{2}} \right] + \frac{\frac{t^2}{N} - T}{2\sigma^2} \\ &= -\frac{N}{2} \log 2\pi - N \log \sigma + \frac{\frac{t^2}{N} - T}{2\sigma^2} \end{aligned} \quad (\text{B.2})$$

As we shall see, the $-\frac{N}{2}$ term exists for all types of superpixels (new or old), and thus can be treated as a constant. This leads us the final expression

$$\mathcal{L}_n(x_{\mathcal{I}_k,d}) = \log p(x_{\mathcal{I}_k,d}, \hat{\mu}_{k,d}) \stackrel{C}{=} -N_k \log \sigma_d + \frac{t_{k,d}^2 - N_k T}{2N_k \sigma_d^2}, \quad (\text{B.3})$$

which matches Equation 8 in the paper.

B.2 Old Superpixels

We now derive the optimal mean parameters and joint log likelihood for old superpixels. We begin by expressing the product of two Gaussian distributions with different means and variances as:

$$\mathcal{N}(x; \mu_1, \sigma_1^2) \cdot \mathcal{N}(x; \mu_2, \sigma_2^2) = (2\pi(\sigma_1^2 + \sigma_2^2))^{-\frac{1}{2}} \exp\left[-\frac{(\mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right] \mathcal{N}(x; \hat{\mu}, \hat{\sigma}^2), \quad (\text{B.4})$$

where the resulting mean and variance are

$$\hat{\mu} = \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}\right) \quad , \quad \hat{\sigma}^2 = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1}. \quad (\text{B.5})$$

Using this relationship with Equation B.1, the joint likelihood can then be expressed as

$$\begin{aligned}
p(x, \mu|\theta) &= p(\mu|\theta) \prod_i p(x_i|\mu), \\
&= \mathcal{N}(\mu; \theta, \delta^2) \cdot \mathcal{N}\left(\mu; \frac{t}{N}, \frac{\sigma^2}{N}\right) N^{-\frac{1}{2}} (2\pi\sigma^2)^{\frac{1-N}{2}} \exp\left[\frac{\frac{t^2}{N} - T}{2\sigma^2}\right] \\
&= N^{-\frac{1}{2}} (2\pi\sigma^2)^{\frac{1-N}{2}} \exp\left[\frac{\frac{t^2}{N} - T}{2\sigma^2}\right] \left(2\pi\left(\delta^2 + \frac{\sigma^2}{N}\right)\right)^{-\frac{1}{2}} \exp\left[-\frac{(\theta - \frac{t}{N})^2}{2(\delta^2 + \frac{\sigma^2}{N})}\right] \mathcal{N}(\mu; \hat{\mu}, \hat{\sigma}^2) \\
&= (2\pi\sigma^2)^{\frac{1-N}{2}} (2\pi(\delta^2 N + \sigma^2))^{-\frac{1}{2}} \exp\left[\frac{\delta^2 t^2 + 2\sigma^2 \theta t - N\sigma^2 \theta^2}{2\sigma^2(\delta^2 N + \sigma^2)} - \frac{T}{2\sigma^2}\right] \mathcal{N}(\mu; \hat{\mu}, \hat{\sigma}^2), \tag{B.6}
\end{aligned}$$

where the optimal parameters are

$$\hat{\mu} = \frac{\theta\sigma^2 + t\delta^2}{N\delta^2 + \sigma^2}, \quad \hat{\sigma}^2 = \left[\frac{1}{\delta^2} + \frac{N}{\sigma^2}\right]^{-1} = \frac{\delta^2\sigma^2}{\sigma^2 + \delta^2 N}. \tag{B.7}$$

Using the optimal mean in the log likelihood simplifies to

$$\begin{aligned}
\log p(x_k, \mu_k|\theta_k) &= \frac{1-N}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log(2\pi(\delta^2 N + \sigma^2)) + \frac{\delta^2 t^2 + 2\sigma^2 \theta t - N\sigma^2 \theta^2}{2\sigma^2(\delta^2 N + \sigma^2)} - \frac{T}{2\sigma^2} - \frac{1}{2} \log(2\pi\hat{\sigma}^2) \\
&= -\frac{N+1}{2} \log(2\pi) + \frac{1}{2} \log\left(\frac{\sigma^2\sigma^{-2N}}{\delta^2 N + \sigma^2} \cdot \frac{\sigma^2 + \delta^2 N}{\delta^2 \sigma^2}\right) + \frac{\delta^2 t^2 + 2\sigma^2 \theta t - N\sigma^2 \theta^2}{2\sigma^2(\delta^2 N + \sigma^2)} - \frac{T}{2\sigma^2} \\
&= -\frac{N+1}{2} \log(2\pi) - \log(\delta\sigma^N) + \frac{\delta^2 t^2 + 2\sigma^2 \theta t - N\sigma^2 \theta^2}{2\sigma^2(\delta^2 N + \sigma^2)} - \frac{T}{2\sigma^2} \\
&= -\frac{N}{2} \log(2\pi) - \log(\delta\sigma^N \sqrt{2\pi}) + \frac{\delta^2 t^2 + 2\sigma^2 \theta t - N\sigma^2 \theta^2}{2\sigma^2(\delta^2 N + \sigma^2)} - \frac{T}{2\sigma^2} \tag{B.8}
\end{aligned}$$

Similar to new superpixels, the first term exists in both cases and can be treated as a constant. This allows us to express the log likelihood as

$$\mathcal{L}_o(x_{\mathcal{I}_k,d}) = \log p(x_{\mathcal{I}_k,d}, \hat{\mu}_{k,d}|\theta_{k,d}) \stackrel{C}{=} -\log(\delta_d \sigma_d^N \sqrt{2\pi}) + \frac{\delta_d^2 t_{k,d}^2 + 2\sigma_d^2 \theta_{k,d} t_{k,d} - N_k \sigma_d^2 \theta_{k,d}^2}{2\sigma_d^2(\delta_d^2 N_k + \sigma_d^2)} - \frac{T_{k,d}}{2\sigma_d^2} \tag{B.9}$$

which matches Equation 20 in the paper.

C Joint Optimization

We chose to use an optimization scheme that consisted of proposing joint moves in the label and parameter space. Alternatively, one could have used an iterative scheme similar to k -means. The plot in Figure 2 shows the posterior log likelihood over all hidden variables for different optimization schemes for a single frame. In particular we consider an iterative method, only using joint local moves, and using all joint moves. While the gain of using joint local moves as compared to an iterative approach is marginal, we find that the split and merge moves greatly help in finding a better mode.

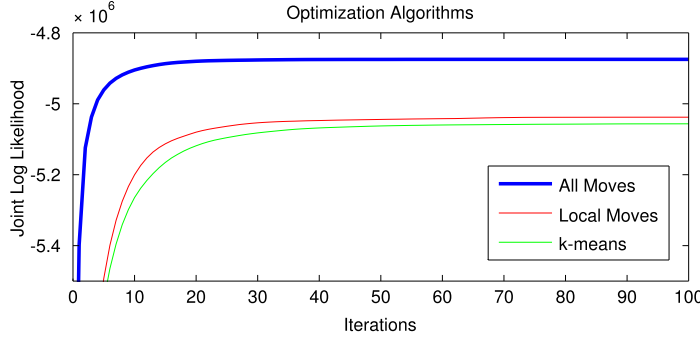


Figure 2: Posterior log likelihood for different optimization algorithms.

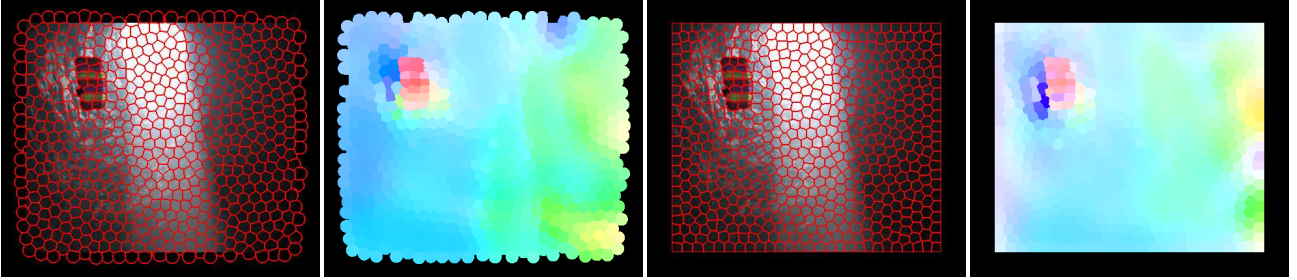


Figure 3: Example superpixels with (left) and without (right) representing support outside of the observed image domain. Corresponding vector difference in means is shown for each case.

D Boundary Effects

As stated in the main paper, boundary effects can cause the resulting mean location and flow to be incorrect. We therefore represent the full support of any superpixel that contains any pixels in the image domain. Because the appearance for pixels outside the domain are not observed, we do not include the appearance likelihood in our calculations. Unlike the appearance, however, the position is known. Since we are not concerned about connectivity in the domain outside of the image, and we do not even need all pixels to be labeled there, we slightly modify the likelihood calculation. We define a null region with label -1 , that pixels only outside of the image domain can belong to.

Given the variance parameter on location and the desired area of a superpixel, the prior says that all pixels within a circle of radius $r = \sqrt{N/(\pi M)}$ should be included in that superpixel. Likewise, all pixels outside r should be included in some other superpixel. We define the probability of a pixel belonging to the null region as

$$p(z_i = k) = \mathcal{N}(r; 0, \sigma_\ell^2) \cdot \mathcal{N}(0; 0, \sigma_\ell^2) \quad (\text{D.1})$$

and use this likelihood to trade off between belonging to a superpixel or null region in the likelihood calculations. At the end of a frame, if a superpixel lies completely outside of the image domain, it is declared to be dead.

An example result with and without the modified representation is shown in Figure 3. Though the effect of the flow on the boundaries is slight, the small errors can accumulate over time and cause large errors.

E Metrics

In this section, we give a mathematical expression for each of the six metrics discussed in the paper. We use the following notation: N is the number of pixels in a frame, T is the number of frames, z are the superpixel labels, and g are the ground truth labels.

3D Undersegmentation Error

The 3D undersegmentation error of [5] can be expressed as

$$\text{UE}(g, z) \triangleq \frac{1}{L} \sum_{l=1}^L \frac{\sum_{\{k | \mathcal{I}_k \cap \mathcal{J}_l \neq \emptyset\}} |\mathcal{I}_k| - |\mathcal{J}_l|}{|\mathcal{J}_l|}, \quad (\text{E.1})$$

where $\mathcal{I}_k = \{(i, t); z_i^t = k\}$ is the set of indices in space and time for superpixel k and $\mathcal{J}_l = \{(i, t); g_i^t = l\}$ is the set of indices in space and time for ground truth segment l .

3D Segmentation Accuracy

The 3D segmentation accuracy of [5] can be expressed as

$$\text{ACC}(g, z) \triangleq \frac{1}{L} \sum_{l=1}^L \sum_{k=1}^K \frac{\max(|\mathcal{I}_k \cap \mathcal{J}_l|, |\mathcal{I}_k \cap \overline{\mathcal{J}}_l|)}{|\mathcal{J}_l|}, \quad (\text{E.2})$$

where $\overline{\mathcal{J}}_l$ denotes the complement of \mathcal{J}_l .

Boundary Recall Distance

The boundary recall distance can be expressed as

$$\text{BRD}(g, z) \triangleq \frac{1}{\sum_t |\mathcal{B}(g^t)|} \sum_{t=1}^T \sum_{i \in \mathcal{B}(g^t)} \min_{j \in \mathcal{B}(z^t)} d(i, j), \quad (\text{E.3})$$

where z^t and g^t are the superpixel label and ground truth label at frame t , $\mathcal{B}(\cdot)$ is the set of boundaries for the label (\cdot) , and $d(\cdot, \cdot)$ is the Euclidean distance between the two arguments.

Temporal Extent

The temporal extent can be expressed as

$$\text{TEX}(z) \triangleq \frac{1}{KT} \sum_{k=1}^K \sum_{t=1}^T \mathbb{I}[\mathcal{I}_k^t \neq \emptyset], \quad (\text{E.4})$$

where $\mathcal{I}_k^t = \{i; z_i^t = k\}$ are the (possibly empty) indices for superpixel k at time t , and $\mathbb{I}[\cdot]$ is the indicator function that evaluates to one iff the argument is true.

Superpixel Size Variation

The superpixel size variation, while easy to calculate, is slightly difficult to express because of the indexing. We define s as the vector of sizes where each element contains the size of a single superpixel in a single frame. The superpixel size variation metric can then be expressed as

$$\text{SZV}(z) \triangleq \sqrt{\frac{1}{|s|} \sum_i s_i^2 - \left[\frac{1}{|s|} \sum_i s_i \right]^2} \quad (\text{E.5})$$

Label Consistency

To define the label consistency, we begin by denoting F as the vector valued ground truth flow field, $F \circ z$ as the warping of z under F , and $[(\cdot)]_i$ as the i^{th} pixel of (\cdot) . The label consistency can be expressed as

$$\text{LC}(F, z) \triangleq \sum_{t=2}^T \sum_{i=1}^N \mathbb{I}[z_i^t = [F \circ z^{t-1}]_i] \quad (\text{E.6})$$

References

- [1] G. Bertrand. Simple points, topological numbers and geodesic neighborhoods in cubic grids. *Pattern Recogn. Lett.*, 1994.
- [2] V. Chalana and Y. Kim. A methodology for evaluation of boundary detection algorithms on medical emages. *IEEE Trans. on Medical Imaging*, 1997.
- [3] X. Han, C. Xu, and J. L. Prince. A topology preserving level set method for geometric deformable models. *PAMI*, 2003.
- [4] T. Y. Kong and A. Rosenfeld. Digital topology: introduction and survey. *Comput. Vision Graph. Image Process.*, 1989.
- [5] C. Xu and J. Corso. Evaluation of super-voxel methods for early video processing. *CVPR*, 2012.
- [6] C. Xu, C. Xiong, and J. J. Corso. Streaming hierarchical video segmentation. *ECCV*, 2012.