Topology-Constrained Layered Tracking with Latent Flow

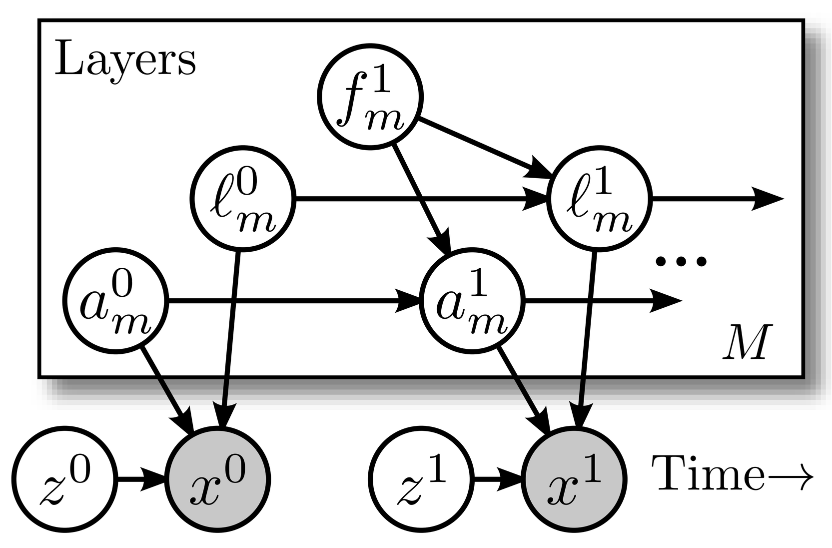
Jason Chang & John W. Fisher III

Overview

In this work, we consider the problem of boundary-accurate object tracking. The approach infers the support, motion, and ordering of each layer within a layered representation We develop an efficient sampling scheme for discrete MRFs with topology constraints that improves upon [2]. We show how to use the sampler for efficient particle filtering without weight decay or the need for resampling techniques.

Layered Model

The graphical model for describing our approach is shown below



f_m : Flow for layer m

a_m : Appearance for layer m

l_m : Support of layer m

x : Observed Frame

z : Layer Ordering

The layer ordering coupled with the layer supports define the visible layer at every pixel:

$$v_i = \arg\min_{\{m|\ell_{m,i}=1\}} z_m$$

The posterior distribution of each observed pixel is modeled to be Gaussian:

$$x_i | a, \ell, z \sim \mathcal{N}(a_{v_i,i}, \Sigma_X)$$

Temporal Dynamics

We denote the previous appearance model warped by the flow as

$$fa_m^{t-1} \triangleq \text{Warp}(f^t, a_m^{t-1})$$

The current appearance model is assumed to evolve under a Gaussian distribution from the previously warped appearance model

$$a_{m,i}^t \sim \mathcal{N}(fa_{m,i}^{t-1}, \Sigma_A)$$

The dynamics on the support are chosen to penalize the curve length and the divergence from the previously warped support

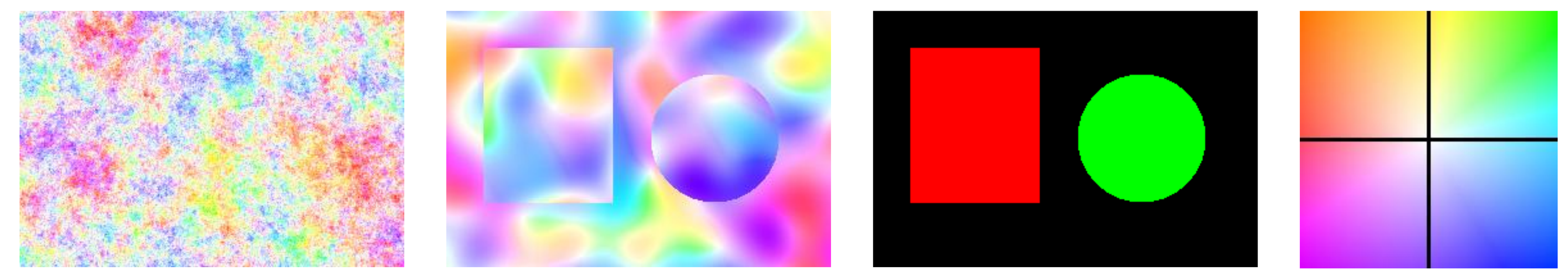
$$p(\ell_m^t | \ell_m^{t-1}, f_m^t) \propto \exp\left[-\alpha_L \oint_{\mathcal{C}_m^t} dl - \alpha_S \sum_i \mathbb{I}\left[\ell_{m,i}^t = f\ell_{m,i}^{t-1}\right]\right]$$

Additionally, we enforce that each layer must be a single connected component.

[1] T. Brox and J. Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Trans. PAMI*, pages 500–513, 2011.
[2] J. Chang and J. W. Fisher III. Efficient topology-controlled sampling of implicit shapes. In *ICIP*, 2012.
[3] M. Grundmann, V. Kwatra, M. Han, and I. Essa. Efficient hierarchical graph based video segmentation. In *CVPR*, 2010.
[4] Y. J. Lee, J. Kim, and K. Grauman. Key-segments for video object segmentation. In *ICCV*, 2011.
[5] C. Liu, W. T. Freeman, E. H. Adelson, and Y. Weiss. Human-assisted motion annotation. In *CVPR*, 2008.
[6] T. Ma and L. Latecki. Maximum weight cliques with mutex constraints for video object segmentation. In *CVPR*, 2012.
[7] D. Tsai, M. Flagg, and J. M.Rehg. Motion coherent tracking with multi-label mrf optimization. In *BMVC*, 2010.

Gaussian Process Flow

Scene motion is often discontinuous at boundaries. We model the motion by combining multiple smooth, layered motion fields. The motion of each layer follows a Gaussian Process (GP) with a stationary covariance kernel. A sample from a layered GP with a squared exponential (SE) kernel compared with the typical L2 neighborhood penalty is below.



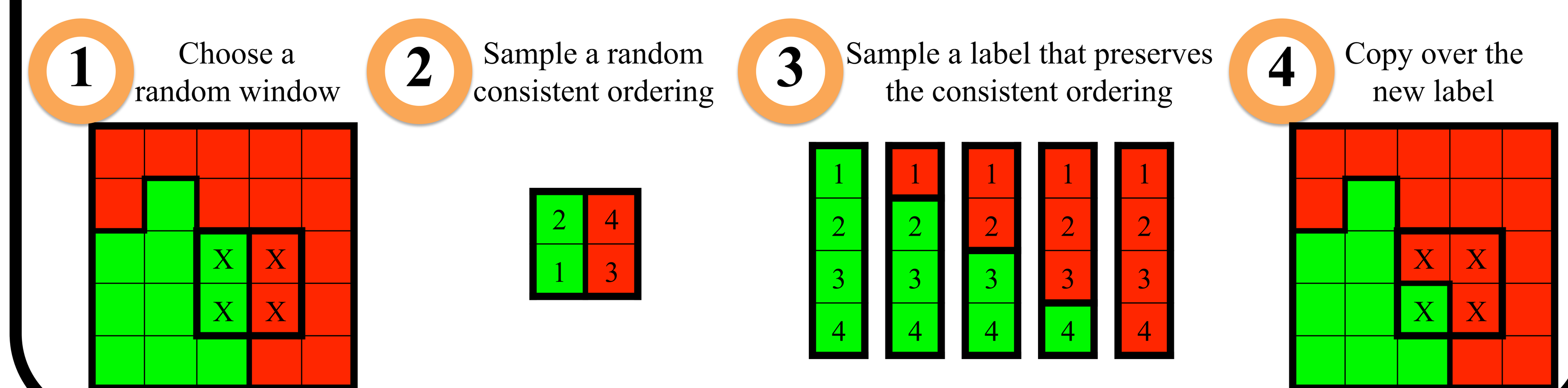
Sparse Neighborhood Precision — Layered GP with SE Kernel — Layer Mask — Flow to Color Map

Since objects may actually self-occlude (e.g. limbs of a person), we choose a kernel that is a composition of the SE kernel and a delta function. The independent portion allows us to model potential discontinuities within a layer.

Sampling-Based Inference

Because the space of layer supports is very large, (~2[N]), we use a particle-filter sampling scheme. However, instead of propagating particles with only the prior, we propagate with both the prior and observation likelihoods. As shown in the paper, this eliminates the need to maintain weights associated with each particle To quickly sample from arbitrary discrete MRFs, we develop the following sampling algorithm.

We show in the paper that the following steps produce a valid Metropolis-Hastings MCMC sampling algorithm for discrete MRFs:



Numerical Comparison

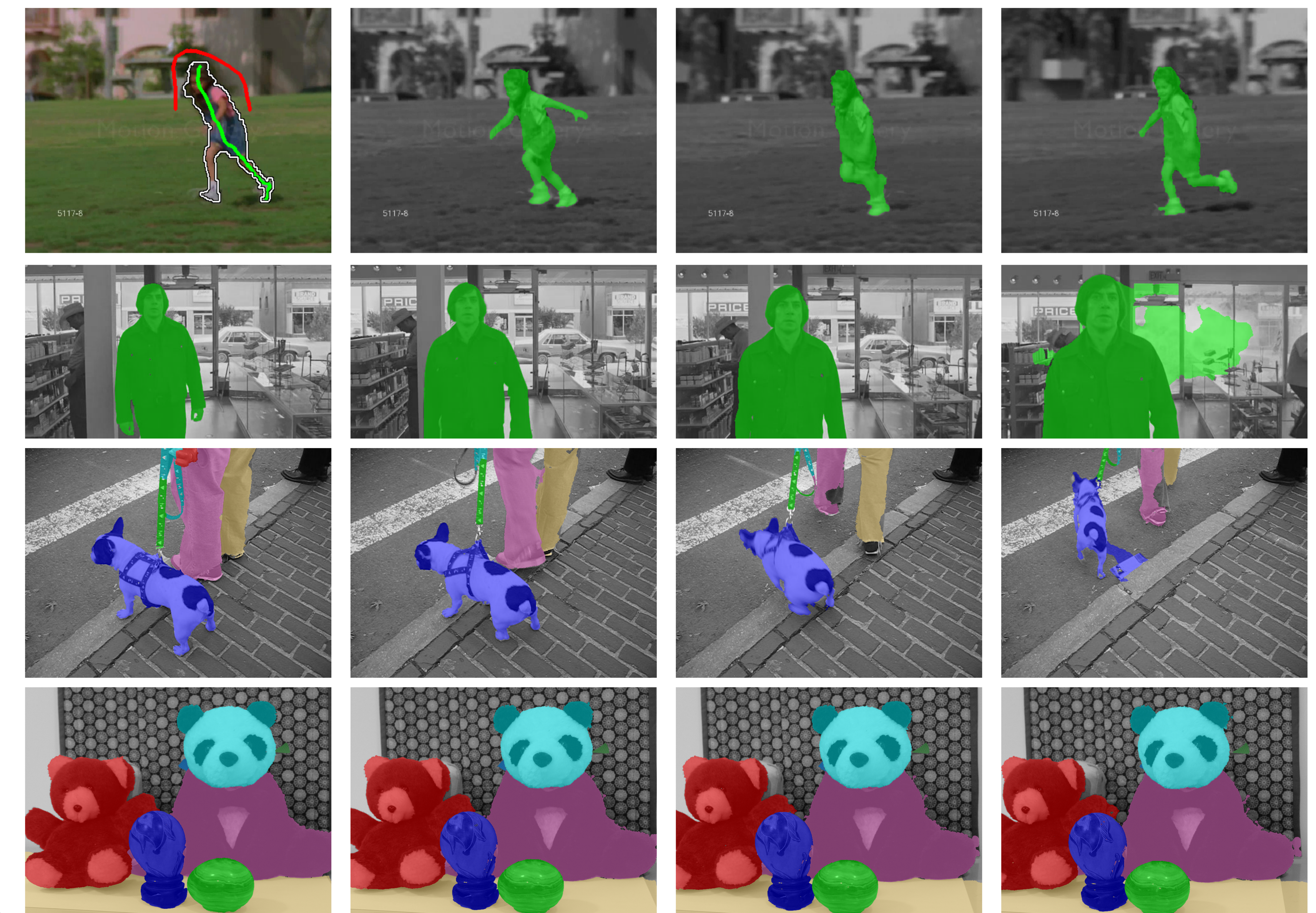
We compare our algorithm to [4, 6, 7] on the Segtrack dataset [7]. Additionally, we annotated the results independently to gauge what is within human error. Pixel errors counts are shown in the table. We note that [4, 6] do not require initializations.

Video	Human	Ours	[6]	[4]	[7]
birdfall	130	265	**189**	288	252
cheetah	308	**570**	806	905	1142
girl	762	**841**	1698	1785	1304
monkeydog	306	**289**	472	521	563
parachute	299	310	221	**201**	235
penguin	279	**456**	-	136285	1705
Mean Error	347	**455**	677*	740*	867

Code available at http://people.csail.mit.edu/jchang7/

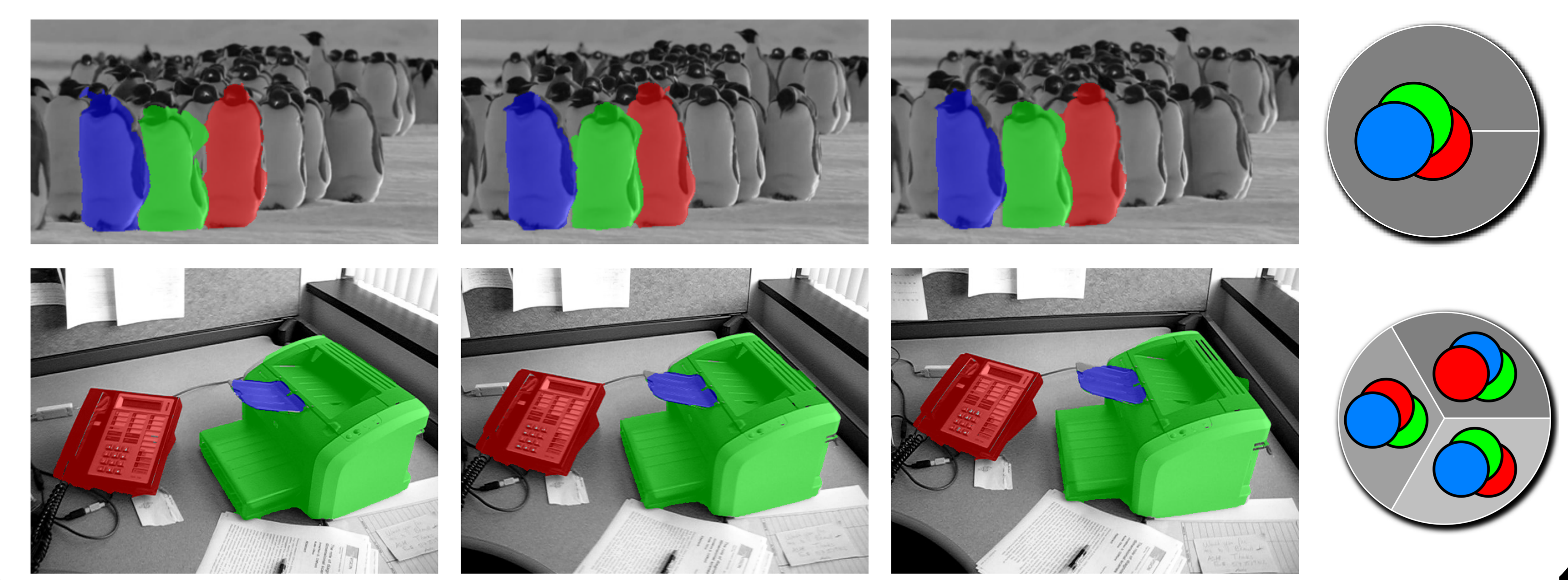
Visualizing Results

Below are sample results using our algorithm.



Layer Order Inference

The following shows results for inferring the layer ordering.



Independent Effects

The following analyzes the impact of the different model components.