University at Buffalo The State University of New York | REACHING OTHERS

ICCV 2013

PDF LIBSVX 3.0

Flattening Supervoxel Hierarchies by the Uniform Entropy Slice

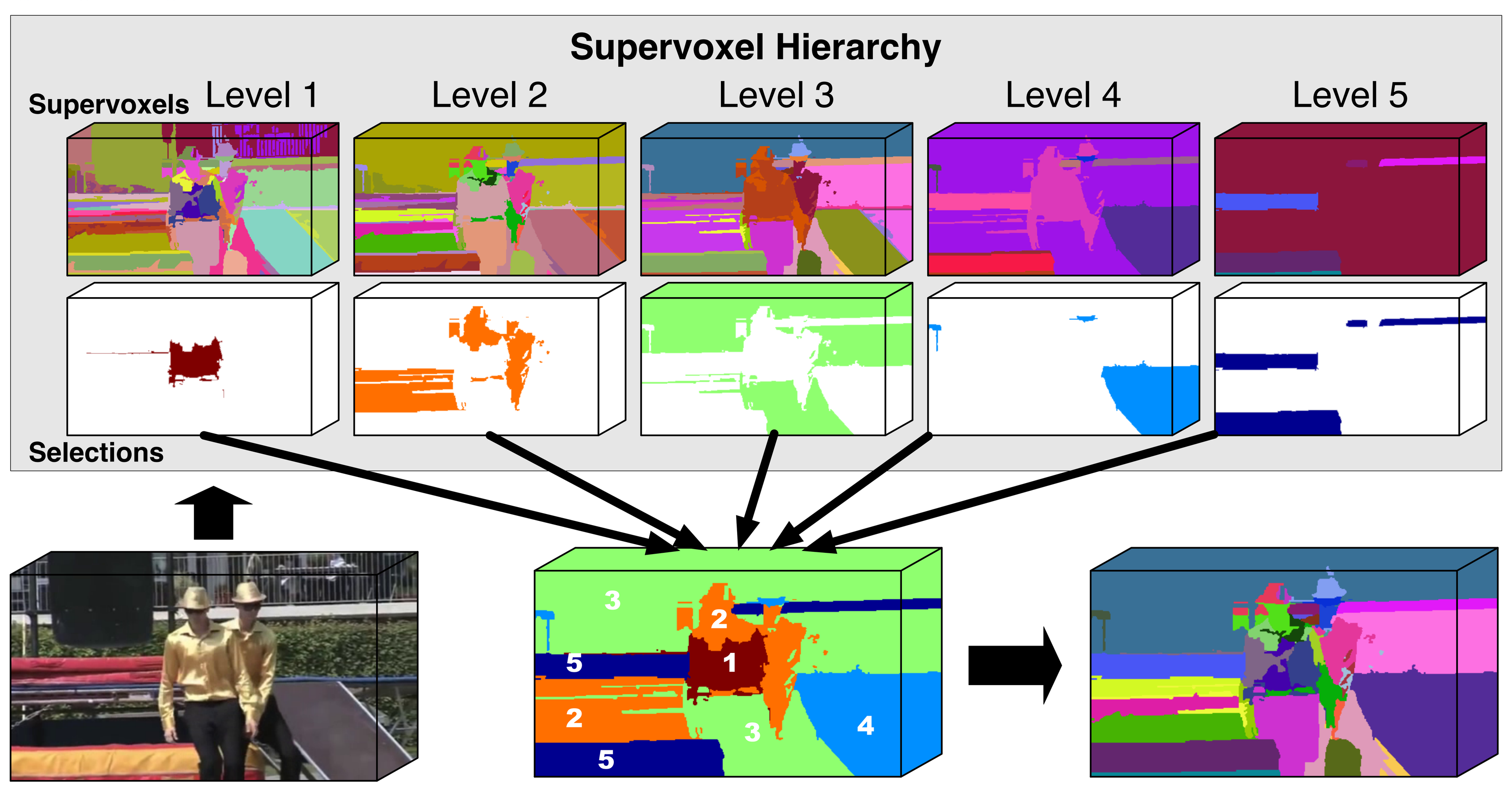
Chenliang Xu, Spencer Whitt and Jason J. Corso
Department of Computer Science and Engineering - SUNY at Buffalo, Buffalo, NY

Objective: **Hierarchical video segmentation has matured in recent years, but it provides too much redundant information than needed.**

- Flatten a supervoxel hierarchy into a single segmentation such that it overcomes the limitations of trivially selecting an arbitrary level.

- Select supervoxels in the hierarchy that balance the relative level of information in the final segmentation based on various post hoc feature criteria, such as motion, object-ness or human-ness.

Overview

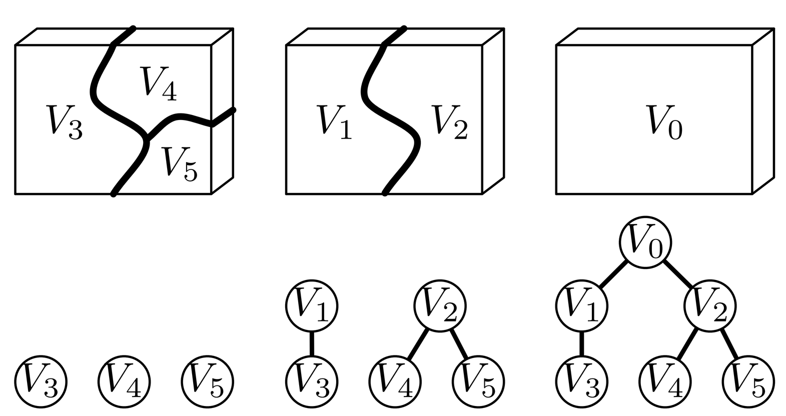


Input Video | UES Selection | UES Flattening

Supervoxel Hierarchy Flattening Problem

Supervoxel Tree Construction

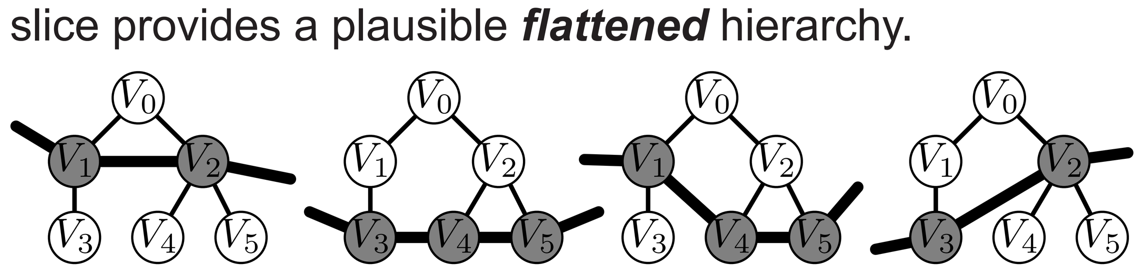
- ***Tree structure hierarchy*** such as the one generated by graph-based hierarchical segmentation (GBH).
- ***Non-tree structure hierarchy*** such as segmentation by weighted aggregation (SWA). A tree can be built by enforcing the boundry agreement from coarse to fine segmentations or vice-versa.



An Example Supervoxel Tree

Supervoxel Tree Slice

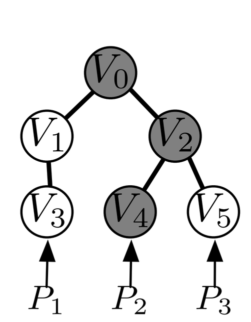
- Define a ***tree slice****** as a set of nodes from the hierarchy such that on each root-to-leaf path in the hierarchy, there is one and only one node in the slice set.
- Each such slice provides a plausible ***flattened*** hierarchy.



Possible Supervoxel Tree Slices

- Consider a binary variable x_s for each node V_s in the tree, and it takes value 1 if the node is a part of the slice and value 0 otherwise. Denote the full set of these over the entire tree as $\mathbf{x}$. Therefore we have the following constraint for a ***valid tree slice***.

$$\mathcal{P}\mathbf{x} = \mathbf{1}_p$$



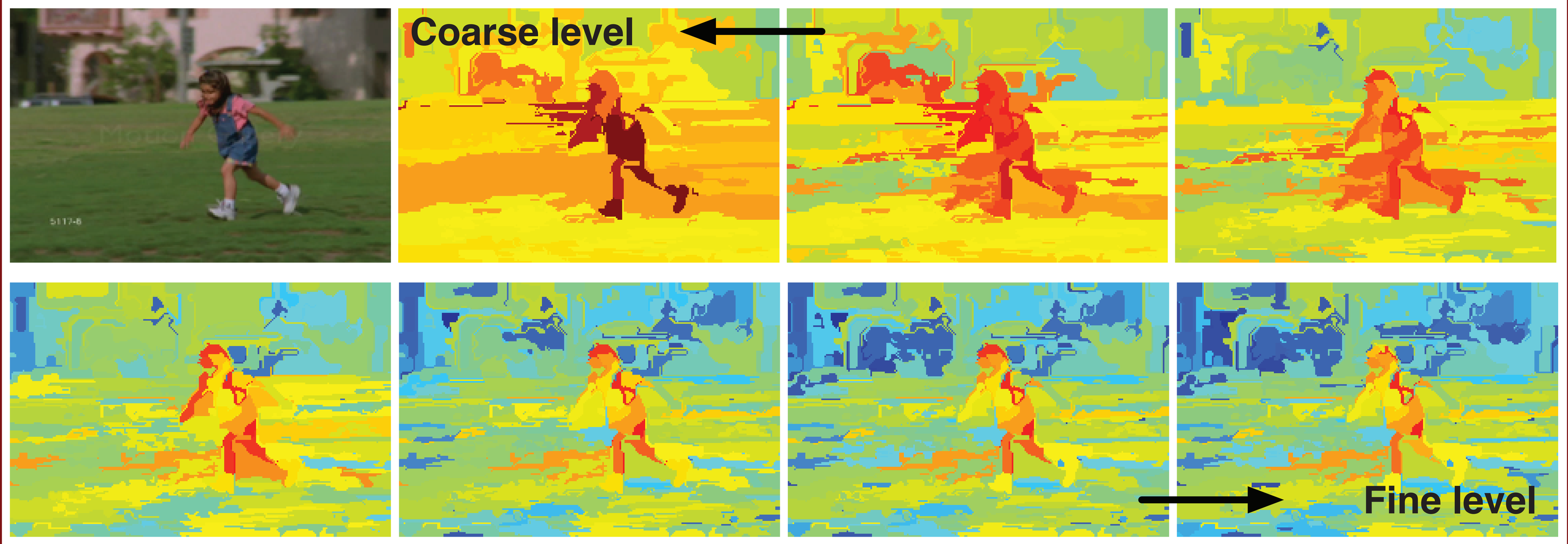
Supervoxel Tree

	V_0	V_1	V_2	V_3	V_4	V_5
P_1	1	1	0	1	0	0
P_2	1	0	1	0	1	0
P_3	1	0	1	0	0	1

Path Matrix $\mathcal{P}$

*Note that we call this a tree slice rather than a ***tree cut*** to distinguish it from conventional use of the term cut, which generally indicates a set of edges and not nodes as we have in the slice.

The Uniform Entropy Slice



Node entropy in a hierarchy for the motion feature criterion

The intuitive idea behind is that we want to select nodes in the tree that balance the information contribution to the overall slice. We want a slice that is able to retain the greatest amount of information relative to the number of selected supervoxels: select bigger supervoxels from coarse levels when there is little information content and conversely, select smaller supervoxels from fine levels when there is high information content.

The ***information content*** of each node in the hierarchy:

$$E(V_s) \doteq -\sum_{\gamma} P_{\mathcal{F}(V_s)}(\gamma) \log P_{\mathcal{F}(V_s)}(\gamma)$$

The ***uniform entropy objective***:

$$\mathbf{x}^* = \arg\min \sum_{V_s, V_t \in \mathcal{T}} |E(V_s) - E(V_t)| x_s x_t$$

Feature Criteria $\mathcal{F}(\cdot)$ can be defined by:
- Unsupervised: Motion.
- Supervised, Category-Independent: Object-ness.
- Supervised: Human-ness, Car-ness, etc..

Problem Formulation: UES as a Binary QP

We reformulate the objective as the following binary quadratic program, which we call the ***uniform entropy slice (UES)***.

$$\begin{aligned} \text{minimize} \quad & \sum_s \alpha_s x_s + \sigma \sum_{s,t} \beta_{s,t} x_s x_t \\ \text{subject to} \quad & \mathcal{P}\mathbf{x} = \mathbf{1}_p \\ & \mathbf{x} = \{0,1\}^N \end{aligned}$$

The ***linear term*** makes the slice prefer simpler segmentations when possible, i.e., prefer coarser levels in the hierarchy rather than finer levels in the hierarchy.

$$\alpha_s = |V^i| \quad \text{if } V_s \in V^i$$

where $|V^i|$ means the total number of supervoxels in i th level of the tree.

The ***quadratic term*** implements the uniform entropy objective:

$$\beta_{s,t} = |E(V_s) - E(V_t)||V_s||V_t|$$

where $|V_s|$ and $|V_t|$ denote the volume of the supervoxels V_s and V_t . By adding the volume factors, we push the selection down the hierarchy unless a uniform entropy has already been achieved.

Experiments and Results

Quantitative Evaluation

- Benchmark Matrics: LIBSVX supervoxel benchmark (segmentation accuracy, undersegmentation error, boundary recall, and boundary precision).
- Dataset: SegTrack dataset, which provides a set of human-labeled single-foreground objects.
- Baselines:
  1) Simple trivial slice that takes a single level from the hierarchy.
  2) Video extension of Li et al. Segmentation by Aggregating Superpixels (SAS).
- Feature Criterion: Motion.

Video	3D ACCU			3D UE			3D BR			3D BP		
	BASE	SAS	UES	BASE	SAS	UES	BASE	SAS	UES	BASE	SAS	UES
birdfall2	9.0	0.0	**69.7**	36.8	38.3	**26.5**	82.1	81.9	**84.9**	0.66	0.65	**0.70**
cheetah	0.0	0.0	0.0	47.4	47.4	47.4	65.7	65.7	65.7	1.93	1.93	1.93
girl	**56.4**	55.9	56.1	7.8	8.2	**5.9**	56.6	56.5	**57.7**	3.36	**3.39**	3.31
monkeydog	0.0	0.0	0.0	52.0	52.2	**51.9**	84.9	**86.8**	86.7	3.32	3.12	**3.35**
parachute	83.7	85.5	**90.3**	23.6	24.4	**22.3**	93.2	93.0	**94.6**	1.66	1.69	**1.72**
penguin	**94.7**	94.4	94.4	**1.8**	1.9	**1.8**	**73.7**	72.3	71.0	1.36	**1.37**	1.27
AVERAGE	40.6	39.3	**51.8**	28.2	28.7	**26.0**	76.0	76.0	**76.8**	**2.05**	2.03	**2.05**

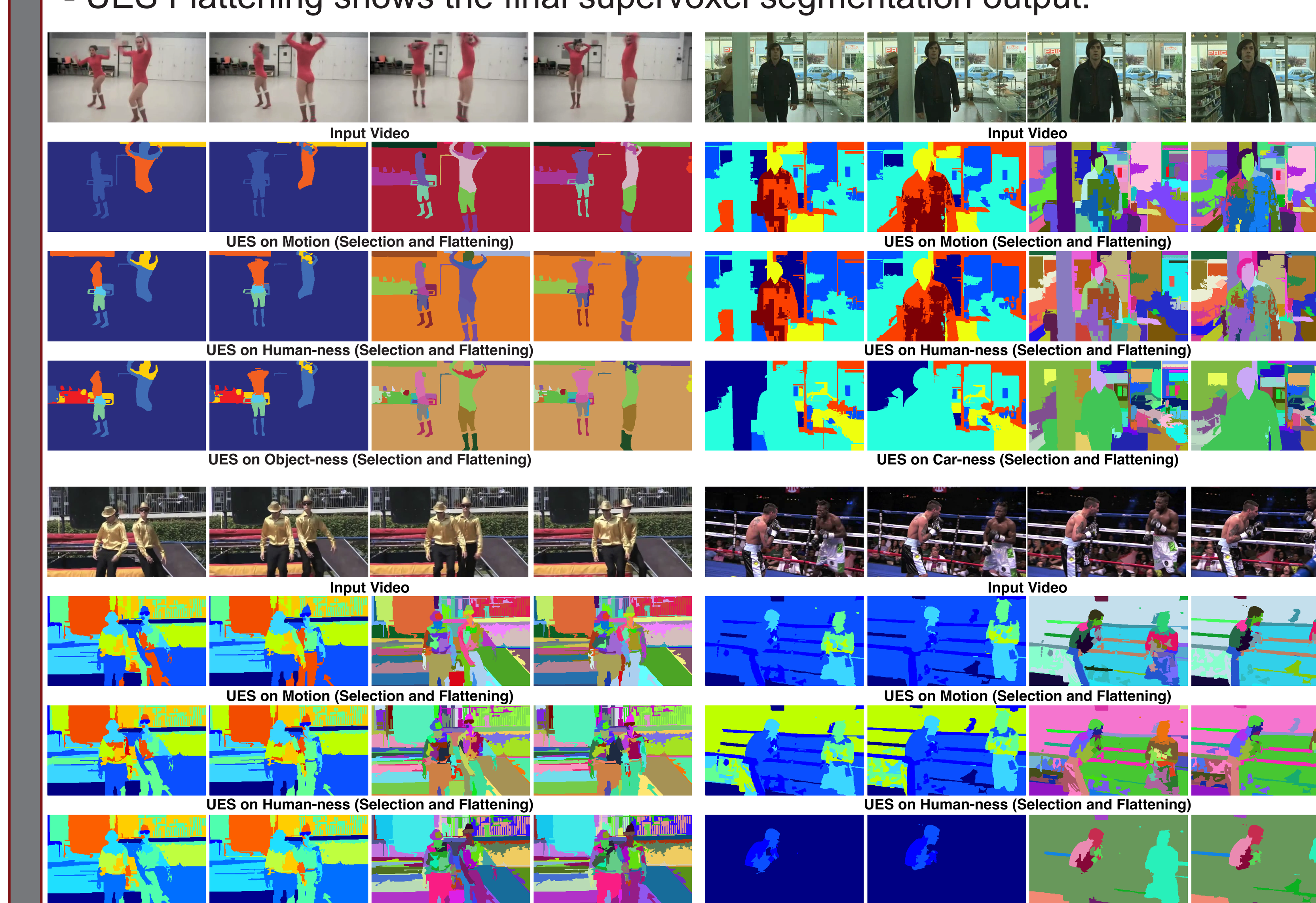
Evaluation on SWA Flattening

Video	3D ACCU			3D UE			3D BR			3D BP		
	BASE	SAS	UES	BASE	SAS	UES	BASE	SAS	UES	BASE	SAS	UES
birdfall2	1.8	0.0	**53.8**	26.9	27.1	**23.2**	74.3	74.0	**82.1**	0.83	0.83	**0.94**
cheetah	30.2	30.2	**39.4**	**31.7**	32.4	34.1	78.3	**79.3**	75.3	1.42	1.43	**1.60**
girl	41.9	**45.6**	41.9	11.2	**11.1**	13.7	54.4	54.1	**58.1**	2.90	2.91	**3.94**
monkeydog	71.9	**79.9**	**79.9**	37.1	**36.6**	43.2	90.7	90.9	**91.0**	2.55	2.47	**2.95**
parachute	89.4	89.4	89.4	38.6	38.6	38.6	87.4	87.4	87.4	10.0	10.0	10.0
penguin	84.7	83.1	**85.0**	2.2	1.9	1.8	**66.7**	65.4	65.5	**1.10**	0.96	0.88
AVERAGE	53.3	54.7	**64.9**	**24.6**	**24.6**	25.8	75.3	75.2	**76.6**	3.14	3.11	**3.39**

Evaluation on GBH Flattening

Qualitative Evaluation

- UES Selection shows which level in the hierarchy a supervoxel is chosen from.
- UES Flattening shows the final supervoxel segmentation output.



Acknowledgements. This work was partially supported by NSF CAREER IIS-0845282, ARO YIP W911NF-11-1-0090 and DARPA Mind's Eye W911NF-10-2-0062.