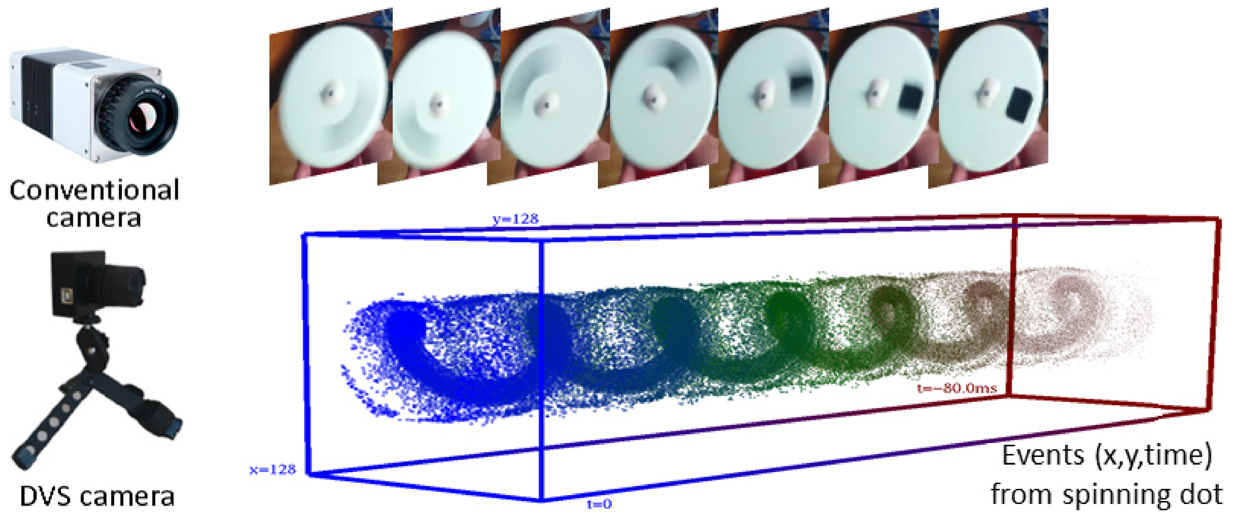
Contour Detection and Characterization for Asynchronous Event Sensors

Francisco Barranco* [1,2], Ching L. Teo*[1], Cornelia Fermüller[1], and Yiannis Aloimonos[1]
[1]Department of Computer Science, University of Maryland, College Park, MD
[2]Department of Computer Architecture and Technology, University of Granada, Spain

Abstract

The bio-inspired, asynchronous event-based **Dynamic Vision Sensor (DVS)** records temporal changes in the luminance of the scene at **high temporal resolution**. Since events are only triggered at significant luminance changes, most events occur at the boundary of objects. The detection of these contours is an essential step for further interpretation of the scene. This work presents an approach to learn the location of contours and their border ownership using **Structured Random Forests (SRFs)** on event-based features that encode motion, timing, texture, and spatial orientations. The classifier integrates information over time by utilizing the classification results previously computed. Experimental results demonstrate good performance in **boundary detection**, **border ownership** and **segmentation**.

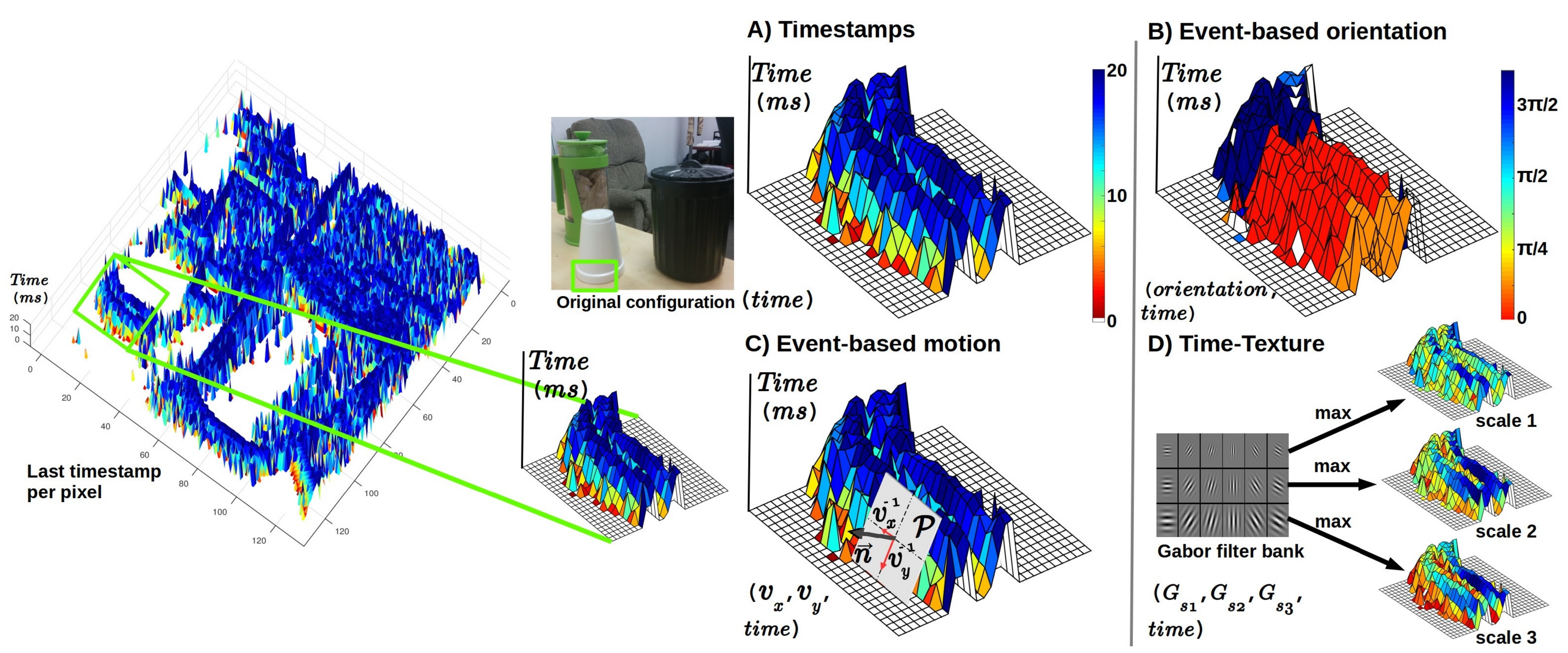
What is a Dynamic Vision Sensor (DVS)?



The DVS[4] provides asynchronous responses at **high temporal resolution** (~15μs) of **where** and **when** changes in the scene occur.

Image motion estimation and **detection of object boundaries** are considered two chicken-and-egg problems. Thus, locating object contours in early stages facilitates further processing such as dense image motion, segmentation, or recognition.

Extraction of event-based features



Motivations:

Event-based motion encodes relative depth information and allows us to detect occlusion boundaries.

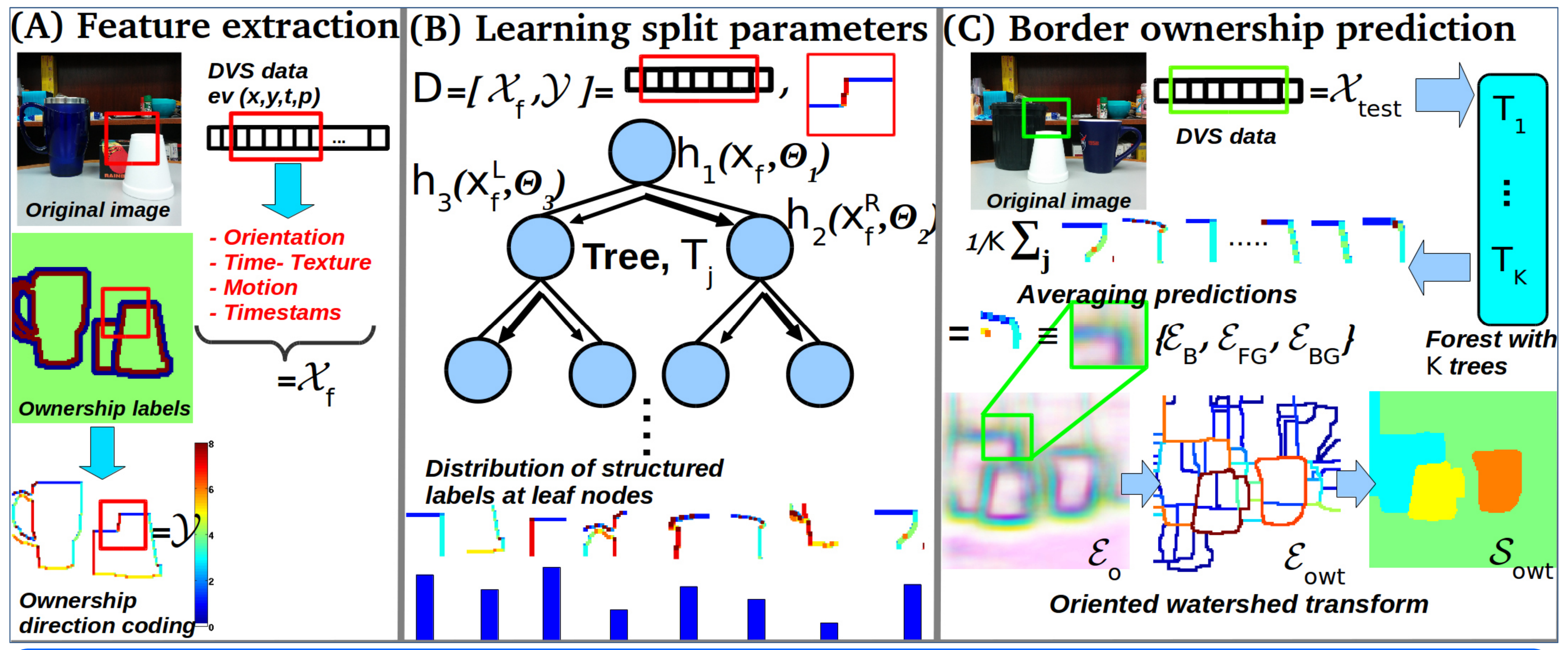
Temporal data provides information for tracking contours.

Orientation is extensively used in boundary detection and ownership[2].

Time texture helps mainly separating foreground and background textures from contours.

Border ownership assignment via SRF

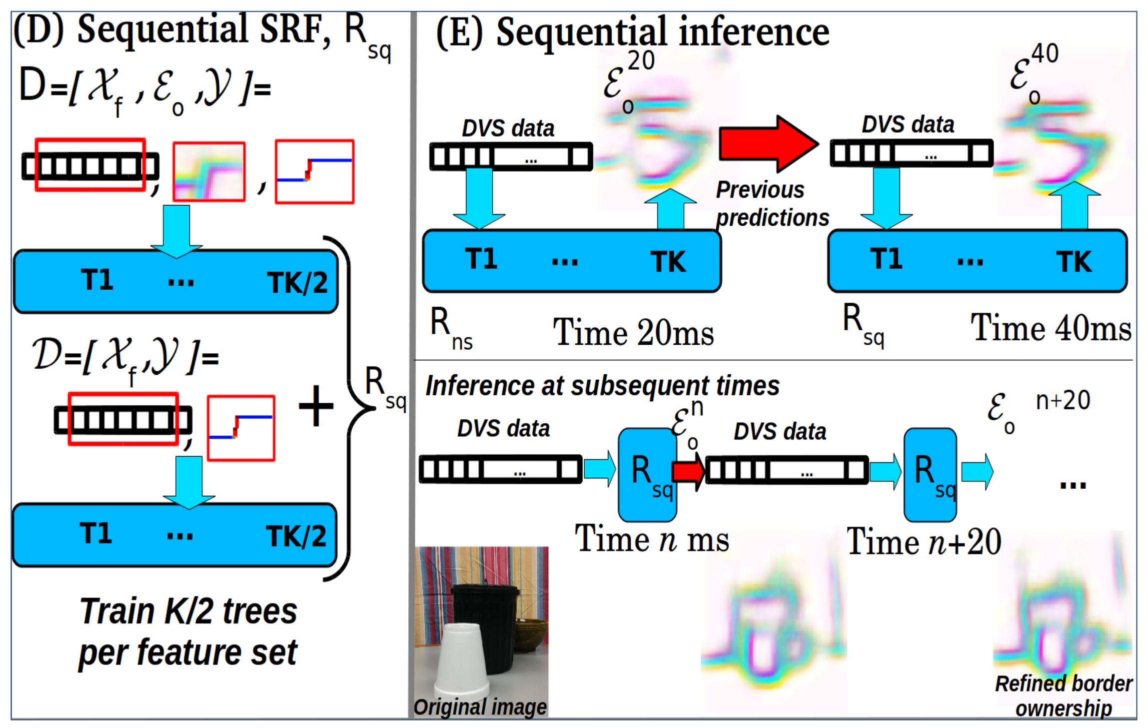
The SRF is trained for **border ownership assignment** using event-based features from random (16x16) patches. **(A)** Given the training data D, we learn an optimal splitting threshold Θ_j, associated with a binary split function h_i at every split node. **(B)** The leaves at each tree T_j encode a distribution of the ownership orientation which we use during inference. Averaging the responses over all K trees produces the final boundary and ownership prediction: $\mathcal{E}_o = \{\mathcal{E}_B, \mathcal{E}_{FG}, \mathcal{E}_{BG}\}$. We then obtain $\mathcal{E}_{owt}$ by applying a watershed transformation over $\mathcal{E}_B$ to construct an initial segmentation $\mathcal{S}_{owt}$ **(C)**.



Refinement and event-based segmentation

We **augment** in **(D-E)** the event-based features with the predictions computed for the *previous time interval.*

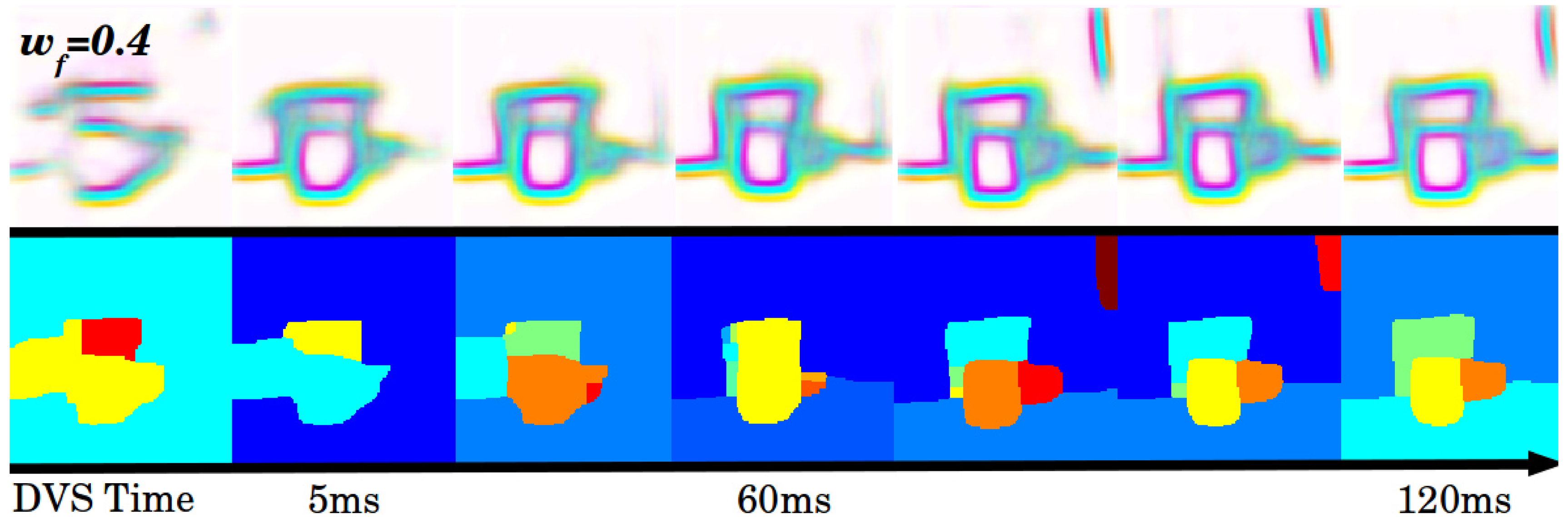
The original, **non-sequential SRF** R_{ns} creates predictions $\mathcal{E}_o^n$, which are used with the features for the next time (*n+1*) as input to the **sequential SRF** R_{sq}.



Refined segmentation

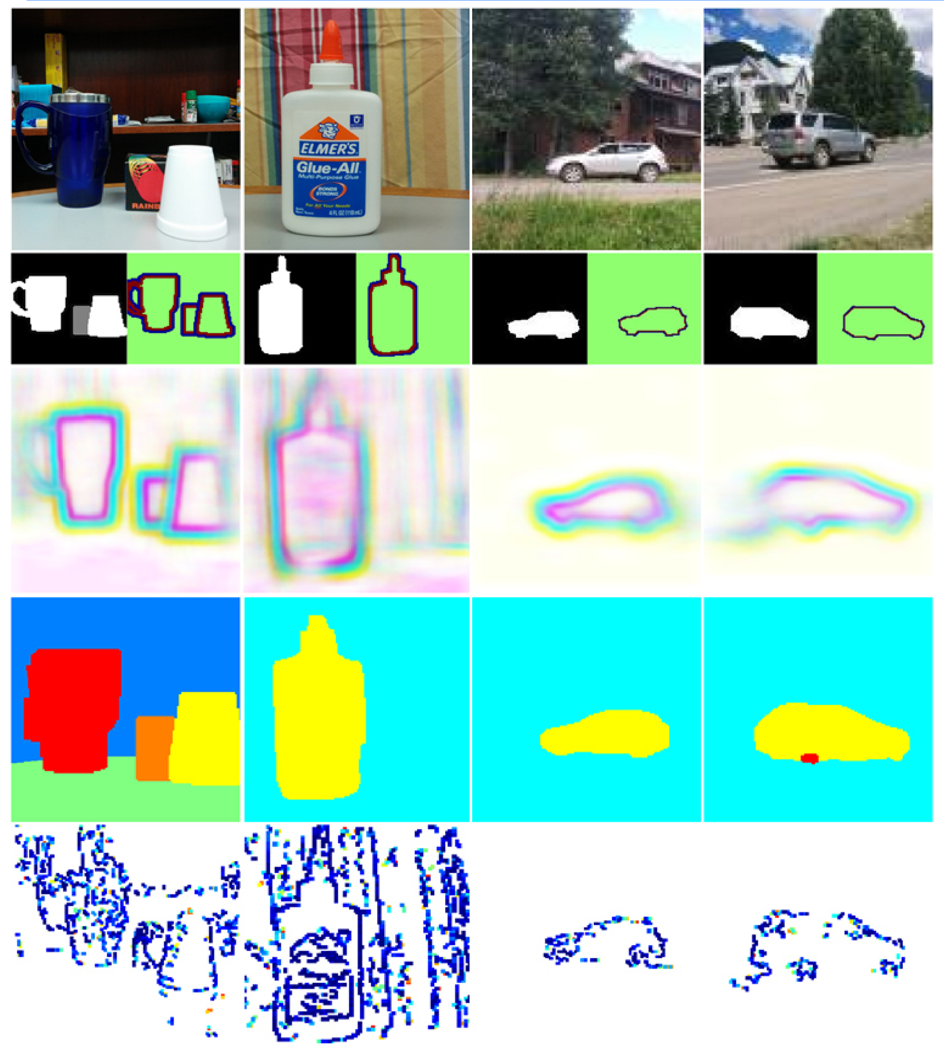
1) **Initial segmentation** $\mathcal{S}_{owt}$ estimated from the predictions $\mathcal{E}_o$ of the SRF.

2) Segments are refined by enforcing **motion coherence** between them.



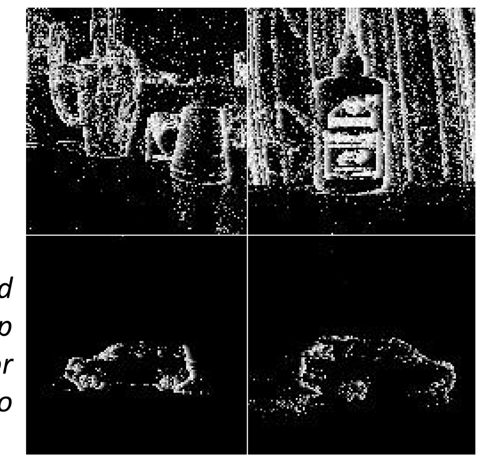
Top: Predictions of sequential SRF R_{sq} for different the weight w_f = 0.4 (blue: boundary, red: FG, yellow: BG). Bottom: Segmentation after refining predictions and using the previous segmentations. Estimations retain some history along time with increasing w_f.

Experiments



1) Indoor sequences of 15 common objects, over varying backgrounds, distances, shape, and sizes with a maximum of 3 depth layers.

2) Outdoor sequences with moving and static cars.



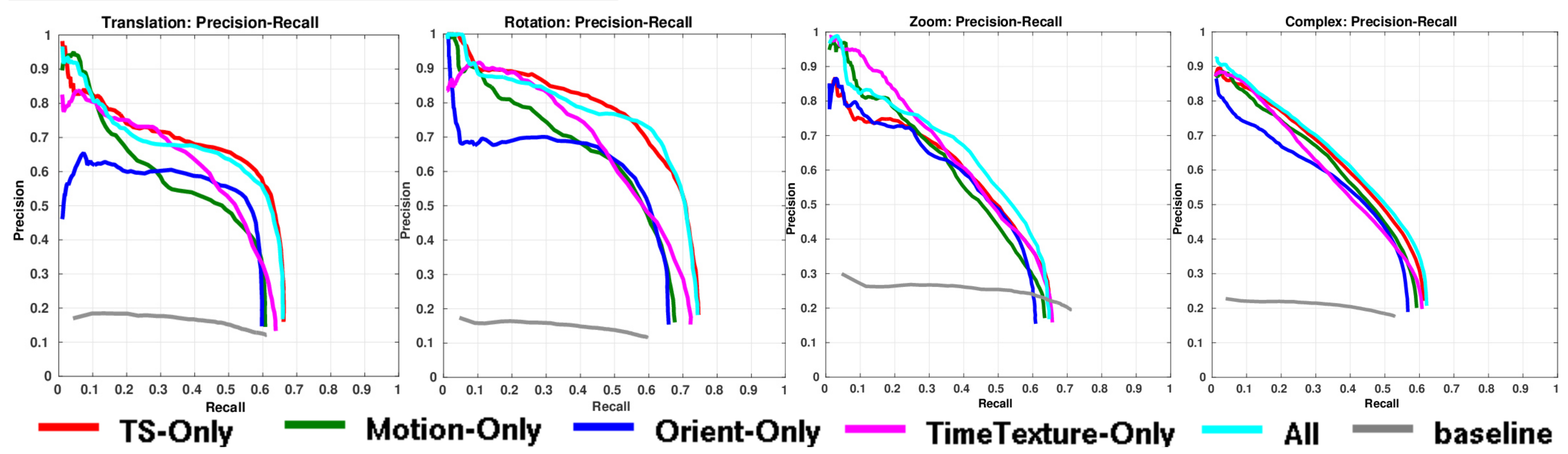
DVS data accumulated during 40 ms. From top to bottom: two indoor scenarios and two outdoor scenarios.

Example results (top to bottom): Original scene, hand-annotated segmentation and border ownership, predicted boundaries (blue: boundary, red: FG, yellow: BG), predicted segmentation, baseline contours.

Feature ablations	Rotation	Translation	Zoom	Complex	NewObj-NewBG	Cars
Timestamp Only	**0.394**, 0.641, **0.517**	0.308, **0.591**, **0.449**	0.239, 0.498, 0.368	0.331, 0.569, 0.450	0.255, 0.473, 0.364	0.343, 0.517, 0.430
Motion Only	0.307, 0.558, 0.433	0.271, 0.492, 0.381	0.251, 0.475, 0.363	0.278, 0.522, 0.400	**0.217**, 0.429, 0.323	0.337, 0.510, 0.423
Orientation Only	0.321, 0.570, 0.445	**0.323**, 0.536, 0.429	0.243, 0.494, 0.368	0.311, 0.525, 0.418	0.232, 0.434, 0.333	0.286, 0.463, 0.375
Time-Texture Only	0.268, 0.552, 0.410	0.197, 0.512, 0.354	0.223, 0.492, 0.358	0.248, 0.472, 0.360	0.193, 0.409, 0.301	0.278, 0.426, 0.352
All features	0.373, **0.661**, **0.517**	0.313, 0.578, 0.445	**0.268**, **0.523**, **0.395**	**0.340**, **0.585**, **0.463**	0.255, **0.478**, **0.366**	†**0.344**, **0.519**, **0.431**
Baseline	–, 0.218, –	–, 0.237, –	–, 0.344, –	–, 0.273, –	–, 0.257, –	–, 0.240, –

Sequence	Before refinement, S_{owt}	After motion refinement, S_M
Rotation	0.91, 0.91, 0.40	**0.93**, **0.93**, **0.37**
Translation	0.92, 0.92, 0.42	**0.93**, **0.93**, **0.40**
Zoom	**0.80**, **0.81**, **0.64**	0.79, **0.81**, 0.91
Complex	0.88, 0.89, **0.51**	**0.91**, **0.91**, **0.51**
Complex-C	0.65, 0.66, 1.45	**0.67**, **0.68**, **1.30**

*Top: For each feature ablation we report {$\mathbf{F}_{own}$, **ODS**, $\mathbf{F}_c$}. For boundaries, we report the maximal F-score (**ODS**) [5] and for ownership the F-score $\mathbf{F}_{own}$ for predictions not further than 0.4% of the image diagonal to the groundtruth [3]. Finally, $\mathbf{F}_c$ measures the average of boundary and ownership.*
*Left: Segmentation accuracy comparing the segmentation $\mathcal{S}_{owt}$ and the refined segmentation including motion information $\mathcal{S}_M$. The metrics reported are GT-Cover (**ODS**), Random Index (**RI**), and Variation Information (**VI**) [1].*



Future work: select features according to the **predominant global motion** and use specific SRF **classifiers tuned** for the predicted motion.

Conclusions

- First approach for locating border contours and assigning border ownership for event-based data.
- The method will be used in future work to develop a complete motion segmentation using as input DVS streams together with classical images (provided by new experimental cameras).

Acknowledgments

This work was supported by an EU Marie Curie grant IOF-33208, the EU Project Poeticon++ under the Cognitive Systems program, the National Science Foundation under INSPIRE grant SMA 1248056, the Junta de Andalucia VITVIR project P11-TIC-8120, and DARPA through U.S. Army grant W911NF-14-1-0384.

Code & data: **www.umiacs.umd.edu/research/POETICON/DVSContours**

[1] P. Arbelaez, et al. "From contours to regions: An empirical evaluation". *IEEE Conf. Computer Vision and Pattern Recognition*, 2294-2301, 2009.
[2] G. Kanizsa and W. Gerbino. "Convexity and symmetry in figure-ground organization". *Vision and artifact*, 25-32, 1976.
[3] I. Leichter and M. Lindenbaum. "Boundary ownership by lifting to 2.1D". *Conf. Computer Vision and Pattern Recognition*, 9-16, 2009.
[4] P. Lichtsteiner et al. "A 128×128 120db 15µs latency asynchronous temporal contrast vision sensor". *IEEE J. Solid-State Circuits*, 43(2): 566-576, 2008
[5] D. R. Martin, et al. "Learning to detect natural image boundaries using local brightness, color, and texture cues". *IEEE T PAMI*, 26(5):530-549, 2004.