Border Ownership Assignment in Real Images

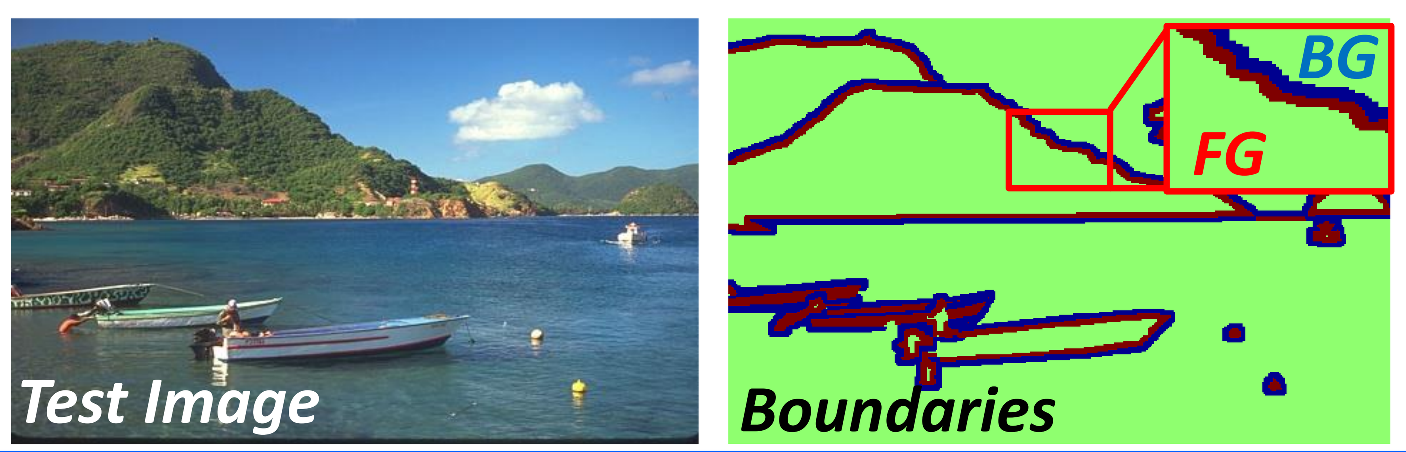
Ching L. Teo , Cornelia Fermüller and Yiannis Aloimonos
University of Maryland Institute for Advance Computer Studies (UMIACS), College Park, MD

Abstract

Although humans can easily understand from single images the 3D structure of the depicted scene, interpreting a complex scene remains an open research question in Computer Vision. Current computational methods employ both top-down (e.g. semantic labels) and bottom-up (e.g. edges) information. In this work, we address the detection of a visual cue that captures relative depth information, the so-called **border ownership**, in real images by analyzing intensity patterns near edges and object boundaries. Leveraging on a fast classification technique known as the Structured Random Forest (SRF), we embed two border ownership cues into the SRF: 1) **dominant Spectral Patterns** and 2) **Gestalt-like grouping patterns**. Experimental evaluation of the proposed approach over two diverse datasets of real images: a) The Berkeley Segmentation Dataset (BSDS) (200 outdoor images) and b) The NYU-Depth (1449 indoor images) shows that our approach is not only more accurate than the previous state-of-the-art, but is able to predict both boundary and border ownership together in real-time: ~0.1s image.

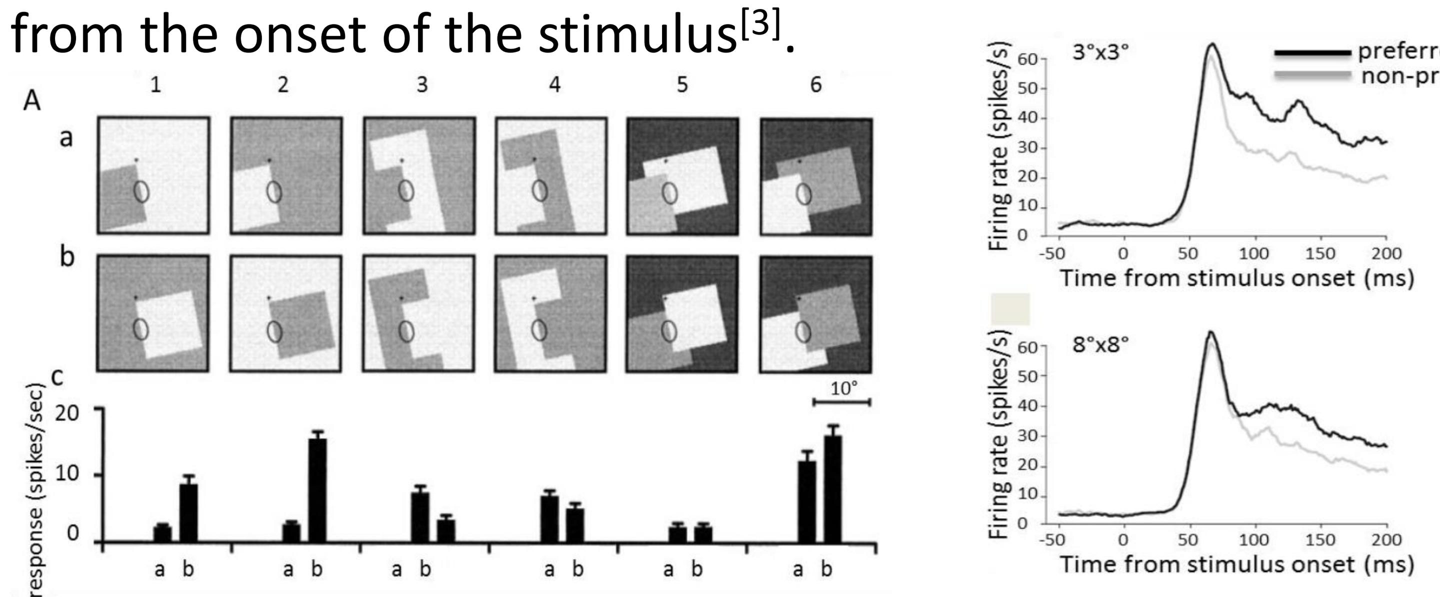
What is Border Ownership?

Given an image and its **boundaries**: regions where objects at different depth meet, the border ownership assignment problem is to determine which **side** of the boundary belongs to the object (foreground – **FG**) and which side is the background (**BG**).



Biological Motivations

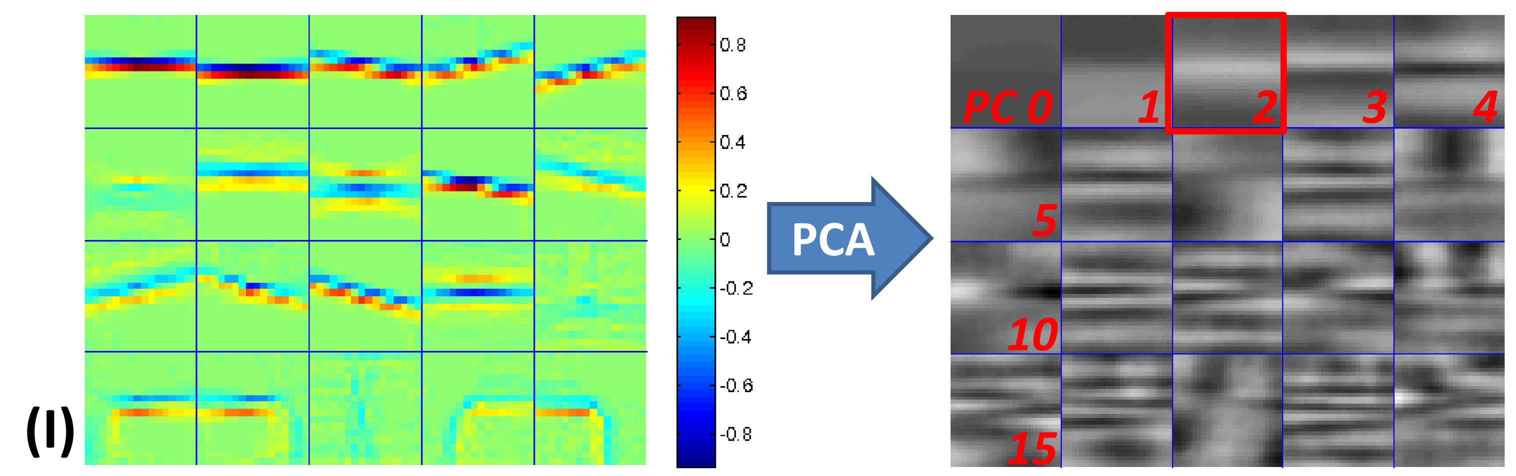
Border ownership sensitive neurons have been demonstrated in the macaque visual cortex[1] from single neuron recordings within the V1, V2 and V4 cortical regions. It was shown that these cells are selective to depth ordering[2] and detection occurs within 75ms from the onset of the stimulus[3].



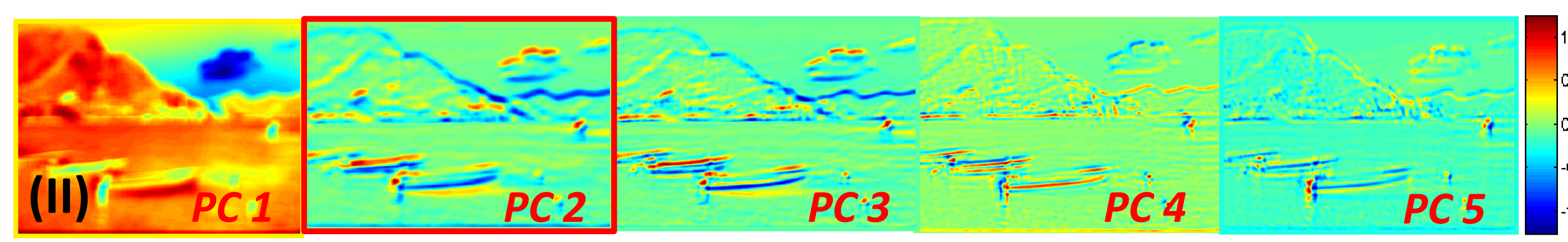
From [1, 3]

Border Ownership Cue 1: Extremal Edges (EE)

Extremal edges (EE), also known as *image folds*, denote the specific change in grayscale intensities that occur along a true boundary of the object, with a distinctive shading at the FG side of the boundary. Psychophysical experiments have shown that EE is one of the strongest cues indicative of border ownership[4].



Aligned horizontal boundary patches (FG at lower half). *Principal components (PC) of the NYU-Depth dataset. PC2 exhibits the distinctive EE shading on the FG-side.*

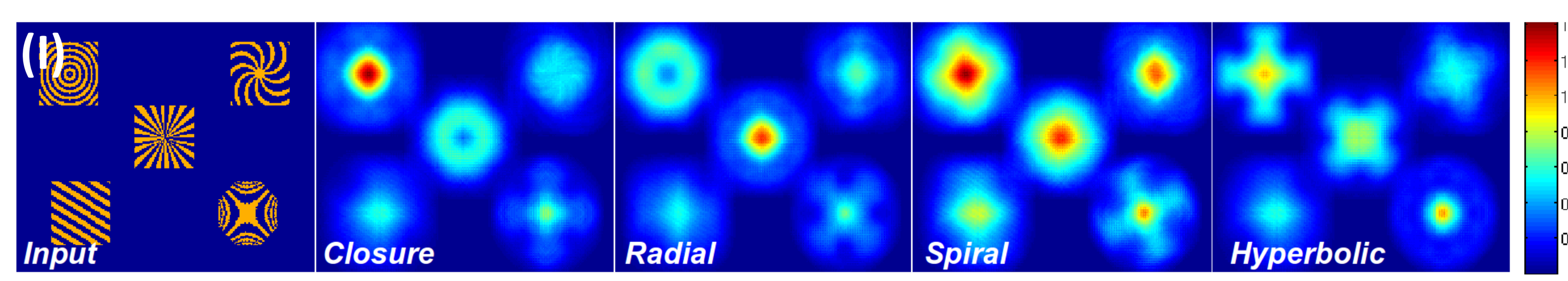


Responses from the projection of the input test image to the top five PCs.

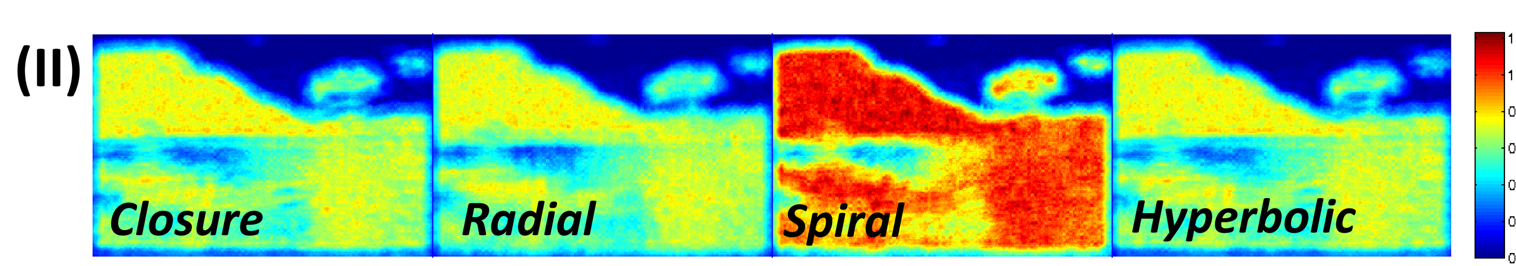
We analyze the intensity patterns within aligned patches using Principal Component Analysis (PCA) **(I)**. The top 5 principal components (PC) are then used as spectral features **(II)**, and the second PC encodes the EE feature.

Border Ownership Cue 2: Gestalt-like patterns

"Gestalt" rules deal with groupings of low-level features (e.g. edges) into *patterns* that encode "object-ness". Such patterns have been observed in area V4 of macaques[5]. We implement through image operators, extending the work in [6], four Gestalt patterns: *closure, radial, spiral* and *hyperbolic* **(I)**. We use the responses of the operator for different patterns over the input image as "Gestalt"-like features **(II)**.

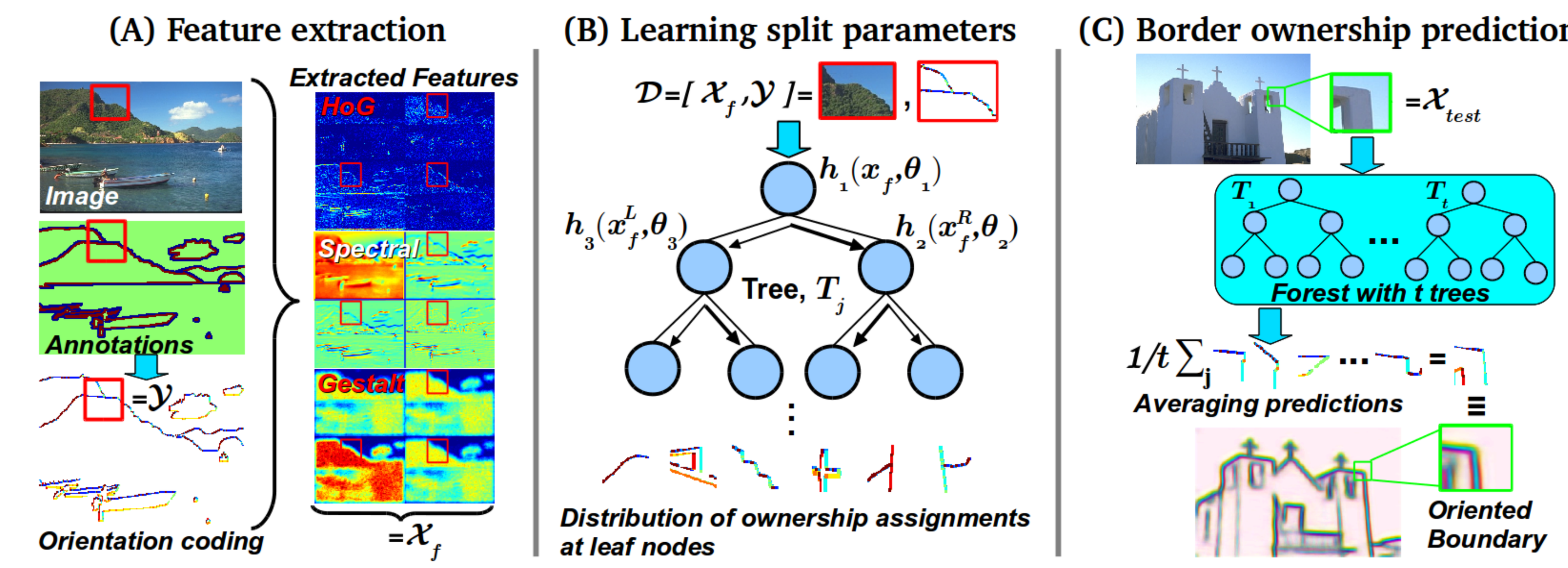


Demonstrating the selectivity of the image operator to four Gestalt patterns.



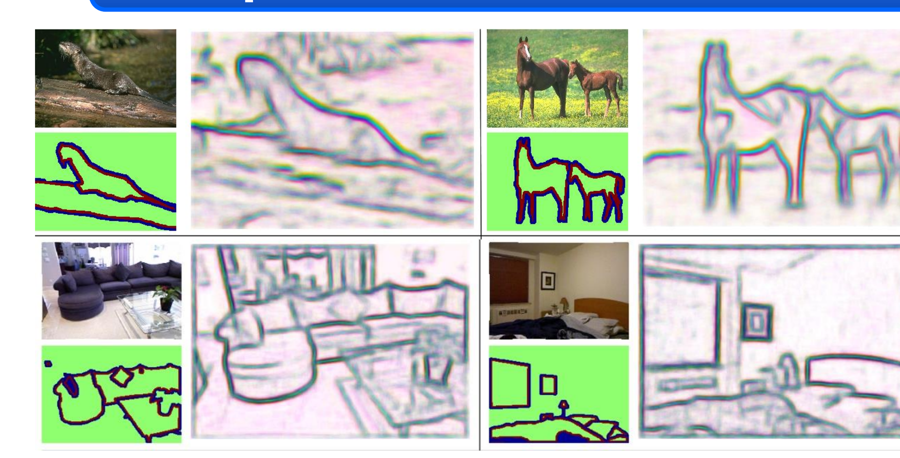
Responses for different Gestalt patterns over the input test image

SRF for Border Ownership Assignment



The SRF consists of a set of *t*=16 decision trees. In order to train the SRF for border ownership assignment, we first extract features from random (16x16) patches **(A)**. In addition to Spectral features and Gestalt-like features, we use Histograms of Gradients (HoG) to localize boundary regions. Next, given the training data, *D*, we learn an optimal splitting threshold, Θ_i associated with a binary split function h_i at every split node **(B)**. The leaves at each tree then encode a distribution of the ownership orientation which we use during inference. Averaging the responses over all *t* trees produces the final boundary and ownership prediction **(C)**.

Experimental Results



Example results: (Top) BSDS dataset and (Bottom) NYU-Depth. Blue: boundary, red: FG, yellow: BG

Feature set	BSDS	NYU-Depth
HoG	72.0%	66.0%
+ Spectral (no contour tokens)	73.1% (72.0%)	67.0% (65.6%)
+ Spectral (contour tokens)	74.0% (72.3%)	68.1% (66.7%)
+ Gestalt patterns	74.4% (72.7%)	**68.4**% (66.7%)
All features + Spectral (NYU)	**74.7**% (72.8%)	-
Global-CRF [7]	69.1%	-
2.1D-CRF [8]	68.9%	-

Feature ablations and comparisons with baselines

Conclusions

A real-time, state-of-the-art approach for border ownership assignment that combines perceptually plausible features with the Structured Random Forest classifier is described. Future works will focus on adding new features (motion and other Gestalt cues) and explore how ownership information can be exploited to improve segmentation and scene understanding.

Acknowledgements

This research was funded in part by the support of the European Union under the Cognitive Systems program (project POETICON++), the National Science Foundation under INSPIRE grant SMA 1248056 and DARPA Grant (W911NF-14-1-0384) under the project: Shared Perception, Cognition and Reasoning for Autonomy.

[1] Zhou et al. "Coding of border ownership in monkey visual cortex". *J. Neuroscience*, 20 (17), 6594–6611, 2000.
[2] Qiu et al. "Figure-ground mechanisms provide structure for selective attention". *Nature Neuroscience*, 10 (11), 1492–1499, 2007.
[3] Sugihara et al. "The speed of context integration in the visual cortex". *J. Neurophysiology*, 106 (1), 374 –385, 2011.
[4] Ghose T. and Palmer S.E. "Extremal edges versus other principals of figure-ground organization". *J. Vision*, 10(8):3, 2010.
[5] Gallant et al. "Neural responses to polar, hyperbolic and Cartesian gratings". *J. Neurophysiology*, 76(4), 2718–39, 1996.
[6] Nishigaki et al. "The image torque: a new tool for mid-level vision". *Conf. Computer Vision and Pattern Recognition*, 502–509, 2012.
[7] Ren et al. "Figure/ground assignment in natural images". *European Conf. Computer Vision*, 614–627, 2006.
[8] Leichter et al. "Boundary ownership by lifting to 2.1d". *Conf. Computer Vision and Pattern Recognition*, 9–16, 2009.