Fast 2D Border Ownership Assignment

Ching L. Teo, Cornelia Fermüller, Yiannis Aloimonos
Computer Vision Lab, University of Maryland, College Park, MD 20742, USA

Accurate localization and detection of object boundaries in 2D images has long been an active area of research in Computer Vision. This is because boundaries provide important low-level signals that serve as input to higher-level visual processes such as scene segmentation and object recognition. Boundaries, however, provide more than just where objects meet with each other or with the background. In this work, we propose a fast approach for *border ownership* assignment, that is, we predict at regions immediate to the boundary whether they belong to the foreground or background (Fig. 1 (top)). Border ownership thus encodes 3D information in the form of ordinal depth relationships. This can be very useful for further scene interpretation that benefits from geometric information, for example for foreground-background segmentation, and it is also closely related to selective attention.

Unlike previous state-of-the-art approaches [4, 7] that first detect boundaries followed by a separate ownership prediction step, our approach predicts *both* boundaries and ownership in one single efficient step using Structured Random Forests (SRF) [3]. Our method not only is faster: 0.1s vs. 15s [4] for a 320×240 image, but it also provides more accurate ownership predictions than the state-of-the-art: 74.7% vs. 69.1% [7] and 68.9% [4] over a subset of the Berkley Segmentation dataset (BSDS) [7] annotated with ownership labels. In addition, we evaluate our approach over the much larger and more challenging NYU Depth V2 (NYU-Depth) dataset [8], which we have annotated automatically using depth and segmentation information, achieving 68.4% prediction accuracy. Since our approach predicts boundaries in addition to ownership, we evaluate the accuracy of boundary prediction over the larger BSDS-500 [1] and NYU-Depth datasets. Compared to current state-of-the-art boundary detectors [1, 2], our approach predicts boundaries with comparable accuracies even though it was trained on sparser annotated data. All code and data used in this work are available at http://www.umiacs.umd.edu/~cteo/BOWN_SRF.

Key to the approach is the use of local and mid-level features that capture border ownership. For local features, we compute Histograms of Gradients (HoG) to capture local shape/curvature and spectral features obtained via PCA over oriented boundary grayscale patches. The intuition is that foreground ownership tends to be on the concave side of a boundary and exhibits specific gradient variations near the boundary, a cue known in the Visual psychology literature as *extremal edges* [6]. An analysis of the top principal components reveals grayscale variations that are not only indicative of extremal edges but other important ownership structures: T-junctions and parallel lines. For mid-level features, we detect four specific Gestalt-like groupings patterns in the image: closure, radial, spiral and hyperbolic by a reformulation of the image *torque* closure operator [5]. We then derive Gestalt-like features from the maximum scale-space response of these operators over the input image.

We use a SRF classifier, trained with HoG, spectral and Gestalt-like features, for border ownership prediction (Fig. 1 (A)–(C)). We first extract features, $x_f \in \mathcal{X}_f$, from random 16×16 input patches, $\mathcal{X}_f$, centered along annotated ownership regions (as positives) and negatives in non-boundary regions. 200,000 patches are sampled per dataset. Features from each patch are then paired with an *orientation coded* annotation structure, $\mathcal{Y}$, of the border ownership. As the ownership annotation essentially encodes a directed edge, by using a 8-way local neighborhood system, we discretize the ownership labels into an orientation code with 8 discrete values. We then proceed to train a SRF in a similar way as [2] where we impose a mapping function prior to computing the decision thresholds θ_i for the i^{th} split function $h_i(x_f, \theta_i)$ of the decision tree. We train t =16 trees with a maximum tree depth of 64 and we sample patches from three (original, half and double) different scales. During inference, we sample test patches densely (at the original resolution) over the entire image and classify them using all 16 decision trees. The final ownership label at each pixel is determined by averaging the predicted orientation labels across all trees, producing an orientation code that we convert directly into an ownership prediction.

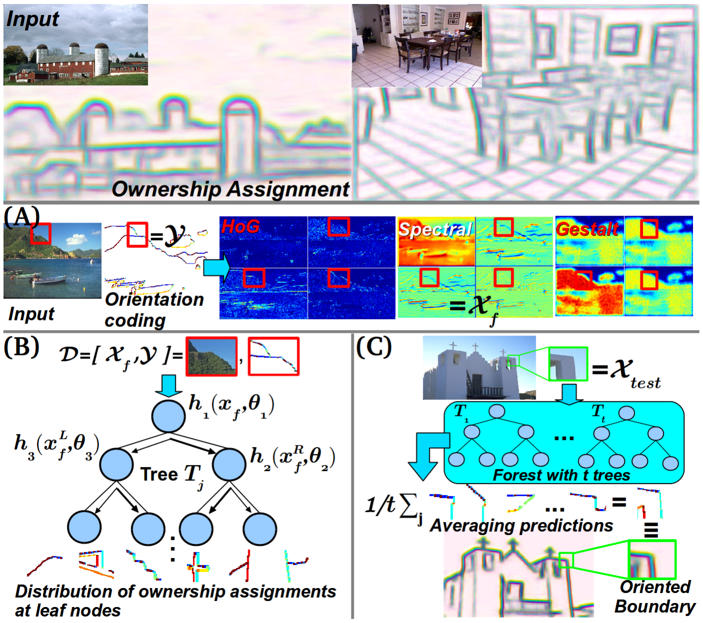

Figure 1: Overview of the approach. (Top) Example results for two images (red: foreground, yellow: background, blue: boundary). (A) Features x_f are extracted from input patches $\mathcal{X}_f$ together with its orientation coded annotation $\mathcal{Y}$. (B) Learning split thresholds θ associated with each split function $h(x_f, \theta)$ from $\{\mathcal{X}_f, \mathcal{Y}\}$ in the SRF. (C) Ownership assignment per test patch is produced by averaging predictions over all decision trees.

We provide more details of the approach, datasets, features and experimental results in the full paper. In conclusion, this work has demonstrated a fast and effective approach that exploits both local and mid-level cues for border ownership assignment. Finally, by making our code available, we hope that this work further motivates the Computer Vision community to exploit border ownership as an important cue for higher-level tasks such as depth prediction, object detection, segmentation and recognition.

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *PAMI*, 33(5):898–916, 2011.

[2] P. Dollár and C. L. Zitnick. Fast edge detection using structured forests. *PAMI*, 2015.

[3] P. Kontschieder, S. R. Bulo, H. Bischof, and M. Pelillo. Structured class-labels in random forests for semantic image labelling. In *ICCV*, pages 2190–2197, 2011.

[4] I. Leichter and M. Lindenbaum. Boundary ownership by lifting to 2.1 d. In *ICCV*, pages 9–16, 2009.

[5] M. Nishigaki, C. Fermüller, and D. DeMenthon. The image torque operator: A new tool for mid-level vision. In *CVPR*, pages 502–509, 2012.

[6] S. E. Palmer and T. Ghose. Extremal edge: A powerful cue to depth perception and figure-ground organization. *Psychological Science*, 19 (1):77–83, 2008.

[7] X. Ren, C. C. Fowlkes, and J. Malik. Figure/ground assignment in natural images. In *ECCV*, pages 614–627. 2006.

[8] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, pages 746–760, 2012.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.