Vistas: Hierarchial boundary priors using multiscale conditional random fields.

Jonathan Warrell
j.warrell@cs.ucl.ac.uk
Alastair P. Moore
a.moore@cs.ucl.ac.uk
Simon J. D. Prince
s.prince@cs.ucl.ac.uk

Department of Computer Science
University College London
Gower Street
London
WC1E 6BT, UK

http://pvl.cs.ucl.ac.uk

Detection of natural boundaries is a fundamental problem in computer vision but evaluation of boundary detection performance has tended to concentrate on images with low scene complexity. Importantly, recent boundary detection analysis [7] shows that performance on scenes with higher scene complexity is low.

However, work in [6] has shown that for datasets with some scene consistency [1] it is possible to learn a distribution over the density of boundaries in the image. The work exploits the fact that when a 2D image is a projection of a 3D scene, perspective effects result in an uneven distribution of the sizes of object classes, and therefore an uneven distribution in the density of object boundaries across the scene. This observation is illustrated in Figure 1a-b. This suggests learning a non-stationary model for boundary priors. In the paper we achieve this by combining two methods that have proved effective for image labeling problems [3, 4] and learn a mixture of multiscale conditional random fields.

The basis of our CRF is a generative model over observed boundaries $\mathbf{x} = [x_1 \ldots x_P]^T$ at the p pixels of an image patch. x_p is binary and is 1 when a boundary is present and 0 when it is absent. We use a clustered sub-space model introduced in [6] which models the probability of $\mathbf{x}$ with an activation $\mathbf{a} = \mu_c + \mathbf{F}_c\mathbf{h}$ comprising a mean, μ, factor matrix, $\mathbf{F}$, and hidden variable $\mathbf{h}$. This model is learned using the EM algorithm. We can summarize the model concisely (see Figure 1c) as:

$$
\begin{aligned}
Pr(c=k) &= \pi_k & (1)\\
Pr(\mathbf{h}) &= \mathcal{G}_{\mathbf{h}}[\mathbf{0},\mathbf{I}] & (2)\\
Pr(\mathbf{a}|\mathbf{h}) &= \delta_{\mathbf{a}}(\mu_c + \mathbf{F}_c\mathbf{h}) & (3)\\
Pr(\mathbf{x}|\mathbf{a}) &= \prod_{p=1}^{P} \mathrm{Bin}_{\mathbf{x}_p}[\sigma(a_p)] & (4)
\end{aligned}
$$

where $\mathcal{G}_\alpha[\beta,\Gamma]$ represents a Gaussian in variable α with mean β and covariance Γ, the function $\delta_\alpha(\beta)$ denotes a deterministic relationship, and the function $\mathrm{Bin}_\alpha[\beta]$ denotes the binomial likelihood of observing value α given binomial parameter β. The term π_k represents the prior probability of choosing the k'th cluster and we also have a prior over $\mathbf{h}$.

In the paper, we apply this model at three distinct scales to create a detailed boundary prior. We apply it at the image level to model overall scene shape clusters e. Each image is additionally broken into a 4×4 grid of large, non-overlapping regions, which we call 'vistas', and whose characteristic boundary distribution clusters d are modeled independently. Finally, the model is applied locally to learn characteristic boundary patch clusters c, which overlap at several scales. The models learned are treated as a prior, and combined with unary boundary estimates from observations $\mathbf{z}$ in a CRF, as shown in Figure 1d. The full model allows the distribution of boundary energy at the scene and vista levels to directly influence the dictionary of boundary patches used at the local level. We go on to demonstrate that our prior improves boundary detection performance on a challenging public road scenes database [1], as shown in Figure 2d.

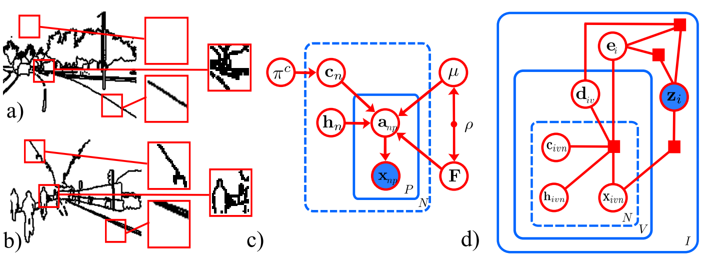

Figure 1: a)-b) Non-stationary boundary density. Patches from different regions of images of road scenes. c) A Clustered Latent Trait (CLT) model for patches. The plates denote a set of N discrete image patches each with P pixels. d) Hierarchical CRF model. The plates denote a set of I images each with a set of V vistas (a grid of non-overlapping large image regions), each with a set of N discrete image patches. The CLT model for image patches can be seen marked with a dashed plate in each model.

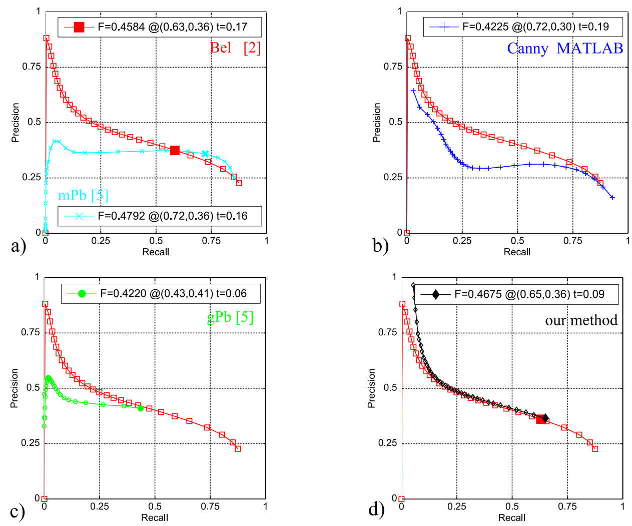

Figure 2: Precision Recall curves. a)-c) Precision/Recall curves for 3 competing boundary detection methods with BEL [2] for reference. d) Precision recall curve for our method using the BEL classifier as the unary term. Note a modest improvement along the length of the PR curve with our CLT prior having greater effect in the regions of lower recall. Please refer to paper for full references.

We conclude the paper by benchmarking six other boundary detection algorithms on the dataset. Some of these can be seen in Figure 2a-c. We discover that there is a difference in the performance of competing algorithms compared to other datasets. For instance, the current best performance on the Berkeley Segmentation Database is the gPb [5] algorithm. However it performs poorly here suggesting that there is less useful information in these street scenes to be found in the spectral components of the image. Overall the performance of all competing methods is low which suggests there is room for significant improvement when considering specific complex scenes. Our generic prior represents a first step in improving on these results (see figure 2d), and we believe that including object class information is the most promising direction for research intent on further improving results on this dataset.

Acknowledgements The first two authors contributed equally to the work in this paper and appear in alphabetical order. We thank Gabriel Brostow for help with the CamVid dataset [1].

[1] G. Brostow, J. Fauqueur, and R. Cipolla. Semantic obje classes in video: A high-definition ground truth databse. *Pattern Recognition Letters*, 2007.

[2] Piotr Dollár, Zhuowen Tu, and Serge Belongie. Supervised learning of edges and object boundaries. *CVPR*, 2:1964–1971, 2006.

[3] Xuming He, Richard D. Zemel, and Miguel A. Carreira-Perpinan. Multiscale Conditional Random Fields for Image Labeling. *CVPR*, 2:695–702, 2004.

[4] Xuming He, Richard S. Zemel, and Debajyoti Ray. Learning and incorporating top-down cues in image segmentation. *ECCV*, 1:338–351, 2006.

[5] Michael Maire, Pablo Arbelaez, Charless Fowlkes, and Jitendra Malik. Using contours to detect and localize junctions in natural images. *CVPR*, 2008.

[6] Alastair P. Moore, Simon J. D. Prince, Jonathan Warrell, Umar Mohammed, and Graham Jones. Scene shape priors for superpixel segmentation. *ICCV*, 2009.

[7] Xiaofeng Ren. Multi-scale improves boundary detection in natural images. *ECCV*, 2008.