Learning to approximate global shape priors for figure-ground segmentation

Daniel Kuettel
dkuettel@vision.ee.ethz.ch
Vittorio Ferrari
vferrari@staffmail.ed.ac.uk

ETH Zurich

University of Edinburgh

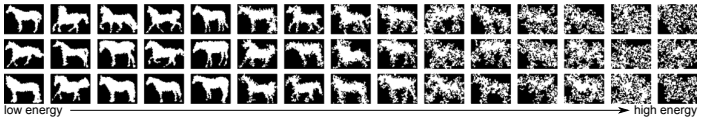

Figure 1: *Training samples drawn from a horse shape prior.*

Many problems in computer vision are formulated as minimizing the energy of a discrete graphical model [2]. This involves designing an energy function to model the problem, and then minimizing it to find the lowest energy labelling. The energy minimization framework allows to cleanly separate modelling and inference. However, it involves a trade-off between expressiveness and optimizability. On the one hand we want sophisticated energy functions which model the problem accurately. On the other hand, we want to find the global optimum of the energy over all possible labellings of its variables. Unfortunately, general efficient minimization algorithms only exist for restricted classes of energy functions. A very popular such class are binary *pairwise* functions [2, 5]. The optimization of higher-order functions is still an open, highly challenging problem [1, 6, 7, 8, 9, 10]. A recent trend is to design energies containing higher-order potentials of a particular form, along with an optimizer specialized for that form [1, 3, 4, 6, 8, 9, 10]. The downside of this approach is that different optimizers needs to be invented for each form of higher-order potential, which requires great skill and knowledge. In this paper, we propose an alternative approach which does not require inventing specialized minimization algorithms. The designer only provides an *arbitrary* energy function E that suits her problem.

Our key idea is to *automatically* learn a pairwise function $\tilde{E}$, which tries to approximate an arbitrary energy function E as well as possible. After learning the parameters of this *proxy* function, it can be efficiently minimized using standard algorithms for pairwise models. We make no assumptions about the form of E. It can be arbitrarily complex and could possibly contain a global potential defined over all variables. The approximation $\tilde{E}$ instead is limited to unary and pairwise potentials. It is learned to hold as well as possible over the entire domain $\tilde{E}(\mathbf{x}) \cong E(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}^N$. The key advantage of learning a pairwise approximation $\tilde{E}$ is that it can be minimized efficiently.

Our goal is to learn the parameters Φ of $\tilde{E}$ so that $\tilde{E}(\mathbf{x};\Phi) \cong E(\mathbf{x})$ for $\forall \mathbf{x} \in \mathcal{X}^N$. We simply treat $E(\mathbf{x})$ as a *black box* mapping a configuration $\mathbf{x}$ to an energy. In principle, with no assumption about the structure of $E(\mathbf{x})$, we would have to evaluate every possible labeling $\mathbf{x} \in \mathcal{X}^N$ to do this. Unfortunately this is not feasible, as there are 2^N different $\mathbf{x}$. We relax the problem by requiring $\tilde{E}$ to approximate E on a very large subset of sample labellings $\tilde{\mathcal{X}} = \{\mathbf{x}_1, \ldots, \mathbf{x}_K\} \subset \mathcal{X}^N$ (typically millions). The underlying assumption is that E varies smoothly over its domain $\mathcal{X}^N$, and therefore the approximation is good also for most other labelling. This is a weak assumption, as smoothness holds for most practical energy functions. Thus we require $\tilde{E}(\mathbf{x}_k;\Phi) \cong E(\mathbf{x}_k), 1 \leq k \leq K$

We cast learning Φ as a least squares minimization problem

$$\min_\Phi \sum_{k=1}^{K} \left(\tilde{E}(\mathbf{x}_k;\Phi) - E(\mathbf{x}_k)\right)^2 \qquad (1)$$

We show how to efficiently learn the parameters Φ on millions of samples from E.

We apply this idea to figure-ground segmentation, where variables x_i are pixels and the labels correspond to foreground ($x_i = 1$) and background ($x_i = 0$). In this context, $E(\mathbf{x})$ is a global shape prior giving lower energy to segmentations $\mathbf{x}$ that fit a shape class such as horses or mugs. In our experiments we use the smallest Chamfer distance to a set of exemplar shapes as E

$$E_c(\mathbf{x}) = \min_{\mathbf{s} \in \mathcal{S}} \left(\frac{1}{|\partial \mathbf{x}|} \sum_{x \in \partial \mathbf{x}} \min_{s \in \partial \mathbf{s}} d(x,s) + \frac{1}{|\partial \mathbf{s}|} \sum_{s \in \partial \mathbf{s}} \min_{x \in \partial \mathbf{x}} d(x,s) \right) \qquad (2)$$

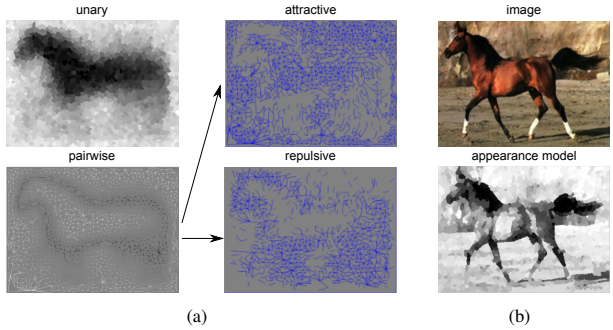

Figure 2: *(a) Learned proxy model for the horse shape prior. (b) Appearance model for a particular test image (darker = more foreground).*

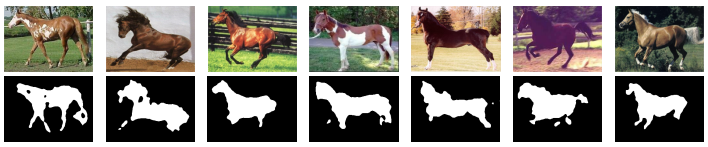

Figure 3: *Example segmentations using a combination of proxy and appearance model.*

where $\mathcal{S}$ is the set of exemplars, d is the Euclidean distance, and ∂ denotes the outline of a segmentation (i.e. the pixels at the boundary between foreground and background).

We conduct our experiments for horse figure-ground segmentation. We build a min-Chamfer prior for horse shapes based on eq. 2. Fig. 1 shows training samples from this shape prior. Fig. 2(a) shows the proxy model learned from E. Our experiments confirm that the proxy closely mimics the original shape prior and separates well between good and bad segmentations.

The shape prior is then combined with an appearance model for horses, see fig. 2(b) and applied to horse test images in fig. 3. When simplifying the min-Chamfer prior to be asymmetric and unnormalized it can be globally optimized with a slow algorithm. For this prior we observe that our learned proxy model achieves virtually the same segmentation accuracy (82.7% versus 83.0%) but at a much lower runtime cost. It also reaches an energy close to the global optimum. Finally, we experiment with the full min-Chamfer prior, that cannot be globally optimized. In this case, the accuracy of the learned proxy improves further to 86.6%. This shows that it is advantageous to use a more advanced shape prior and then solve it approximately.

[1] C. Arora, S. Banerjee, P. Kalra, and S. Maheshwari. Generic cuts: An efficient algorithm for optimal inference in higher order mrf-map. In *ECCV*, 2012.

[2] Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *ICCV*, 2001.

[3] T. Kim, K. Lee, and S. Lee. Nonparametric higher-order learning for interactive segmentation. In *CVPR*, 2010.

[4] P. Kohli, L. Ladicky, and P. Torr. Robust higher order potentials for enforcing label consistency. In *CVPR*, 2008.

[5] Vladimir Kolmogorov and Carsten Rother. Minimizing nonsubmodular functions with graph cuts – a review. *IEEE Trans. on PAMI*, 29(7):1274–1279, 2007.

[6] N. Komodakis and N. Paragios. Beyond pairwise energies: Efficient optimization for higher-order mrfs. In *CVPR*, 2009.

[7] L. Ladicky, C. Russel, P. Kohli, and P.H.S. Torr. Graph cut based inference with co-occurrence statistics. In *ECCV*, 2010.

[8] S. Ramalingam, P. Kohli, K. Alahari, and P. Torr. Exact inference in multi-label crfs with higher order cliques. In *CVPR*, 2008.

[9] C. Rother, P. Kohli, W. Feng, and J. Jia. Minimizing sparse higher order energy functions of discrete variables. In *CVPR*, 2009.

[10] V. Vineet, J. Warrell, and P. Torr. Filter-based mean-field inference for random fields with higher-order terms and product label-spaces. In *ECCV*, 2012.