

# Segmentation using Deformable Spatial Priors with Application to Clothing

Basela S. Hasan  
scbsh@leeds.ac.uk  
David. C. Hogg  
d.c.hogg@leeds.ac.uk

School of Computing  
University of Leeds  
Leeds, UK

The formulation of image segmentation as maximum a posteriori probability (MAP) inference over a Markov Random Field (MRF) is both elegant and effective. Typically, the MRF is configured to favour contiguous regions with the same labelling, and consistency between the label at each pixel and prior intensity distributions for foreground and background regions. Boykov and Jolly show how to solve this labelling problem efficiently by reformulating as finding a minimum graph-cut [1]. In an extension to this method, the colour distribution is treated as a latent property [5] and in a further extension (OBJ CUT) [3], the foreground object is assumed to be an instance of a known object category, so that prior shape information can be exploited as a top-down influence. A 2D *spatial prior* on the foreground/background probabilities for each pixel is used as shape model in [4] and with smooth deformations in [6].

Our novel contribution is threefold: (1) we use a spatial prior with category specific deformation function, ranging over multiple labels corresponding to the different parts of an object; (2) we deal jointly with multiple overlapping object instances within the same image, integrating this into a global optimisation within the same MRF framework; (3) we demonstrate an improvement using this approach on the state of the art for the problem of clothing segmentation from images of groups of people. The use of a deformable prior is motivated by the Active Appearance Model (AAM) [2], except that we deform a map of prior probabilities for the labels at each location, relative to an object-centred frame of reference, rather than textures.

The input to our method is an image  $D = \{d_1, d_2, \dots, d_N\}$ , where  $d_i$  is the observed RGB colour at pixel  $i$ , and  $J$  instance hypotheses  $\{o_1, o_2, \dots, o_J\}$  for a known object category, specifying their position and scale in the image. A candidate segmentation is an assignment of one of  $K + 1$  labels to each pixel in the input image  $L = \{l_1, l_2, \dots, l_N\}$ , meaning that a pixel belongs to one of  $K$  object parts or background. The segmentation task is posed as inference in a MRF model. Given prior information about the shape of each object  $S = \{s_1, s_2, \dots, s_J\}$  and the colour model for each part of each object denoted by  $\Theta = \{\theta_{11}, \dots, \theta_{1K}, \theta_{21}, \dots, \theta_{2k}, \dots, \theta_{J1}, \dots, \theta_{JK}\}$ , we seek the solution  $\hat{L}$  which maximises the posterior probability for  $L$ :

$$\hat{L} = \underset{L}{\operatorname{argmax}} (P(L|D, S, \Theta)) \quad (1)$$

Using Bayes' theorem and the usual spatial Markov assumption, the posterior probability for  $L$  can be written as:

$$\begin{aligned} P(L|D, S, \Theta) &\propto P(D|L, S, \Theta) \times P(L|S, \Theta) \\ &= \prod_i P(d_i|l_i, S, \Theta) \times \prod_i P(l_i|S) \times \prod_{i' \in \mathcal{N}_i} P(l_i|l_{i'}) \end{aligned} \quad (2)$$

where  $\mathcal{N}$  denotes an 8-neighbourhood.

Our shape prior  $P(l_i|S)$  consists of a spatial array of prior probabilities for part labels, coupled with a parameterised deformation function mapping this array into an object-centred frame of reference. The deformation function is defined in terms of linear offsets of a set of landmark points in canonical position and has one or more latent parameters:

$$\hat{m} = \bar{m} + \Phi b \quad (4)$$

where  $\bar{m}$  is the set of canonical positions,  $\Phi$  is a basis for a linear deformation subspace, and  $b$  is a parameter vector giving different deformations. We learn  $\hat{m}$  and  $\Phi$  from training data as in [2].

The conditional probability for a label  $l$  at location  $i$  given a shape model  $s_j$  is given by:

$$P(l_i|s_j) = Q(l_i|T_j(x_i) + \bar{m} + \Phi s_j) \quad (5)$$

where  $T_j(x_i)$  is the transformation giving the position and scale of the instance. In our experiments, these are obtained from face detection. We optimize using Branch and Bound on the spatial deformation coupled with EM to estimate the colour models.

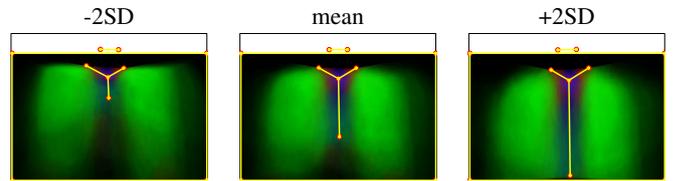


Figure 1: Deformation of spatial prior along principal axis. Ten canonical positions are shown in yellow sticks (including rectangle vertices). Colour code: R = P(shirt), G = P(jacket), B = P(tie).

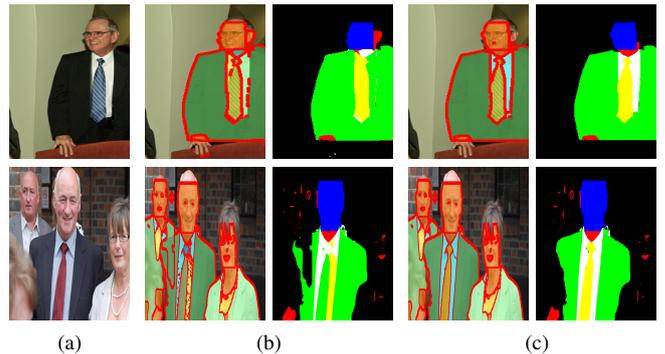


Figure 2: For each input image in (a), Results using fixed prior in (b) are improved in (c) using the deformable prior.

We have tested the method on a dataset of images depicting people wearing suits - the task being to segment into jacket, shirt and tie. Figure 2 compares performance against a baseline method using a fixed spatial prior.

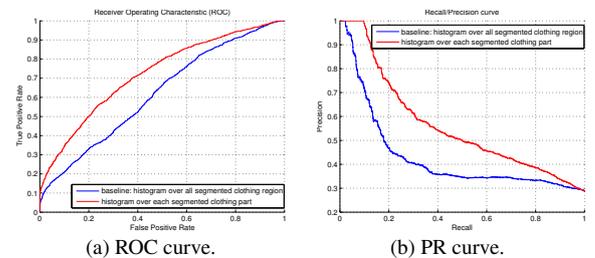


Figure 3: Spatial information about people's clothing is of significant importance for the clothing recognition.

- [1] Y. Y. Boykov and M. Jolly. Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images. *Computer Vision, IEEE International Conference on*, 1:105–112, 2001.
- [2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:681–685, 2001. ISSN 0162-8828.
- [3] M. P. Kumar, P. H. S. Torr, and A. Zisserman. OBJ CUT. In *CVPR (1)*, pages 18–25, 2005.
- [4] K.C. Lee, D. Anguelov, B. Sumengen, and S.B. Gokturk. Markov Random Field Models for Hair and Face Segmentation. In *Proceedings of IEEE Conference On Automatic Face and Gesture Recognition*, pages 1–6, 2008.
- [5] C. Rother, V. Kolmogorov, and A. Blake. “GrabCut”: Interactive Foreground Extraction using Iterated Graph Cuts. *ACM Transactions on Graphics*, 23:309–314, 2004.
- [6] J. Winn and N. Jovic. LOCUS: Learning Object Classes with Unsupervised Segmentation. *Computer Vision, IEEE International Conference on*, 1:756–763, 2005. ISSN 1550-5499.