Joint Adaptive Colour Modelling and Skin, Hair and Clothing Segmentation Using Coherent Probabilistic Index Maps

Carl Scheffler
carl.scheffler@gmail.com
Jean-Marc Odobez
odobez@idiap.ch

Idiap Research Institute
Centre du Parc
Rue Marconi 19
Martigny, Switzerland

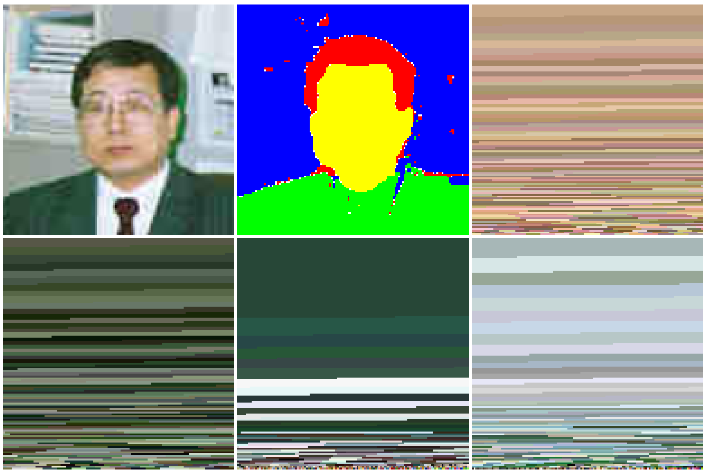

Figure 1: From top–left to bottom–right, a sample image; its segmentation into skin, hair, clothing and background regions; the posterior modes of the beliefs over skin, hair, clothing and background palettes, respectively.

Image segmentation and colour modelling are related problems, as argued by Jojic and Caspi [1]. We address the joint segmentation of regions around faces into four classes (skin, hair, clothing, and background) and the learning of their respective colour models. All inference is done within the Bayesian framework. We start from a prior belief over the palette of each class and the class of each pixel. A *palette* is simple a probability distribution over colours, representing how often we expect to observe each possible colour. Given observed pixel values, approximate posteriors over classes and palettes are computed using a variational inference [3] algorithm.

The model contains two types of palettes: discrete and continuous. Discrete palettes are discrete probability vectors (also known as normalised histograms) over RGB colours, $\vec{\pi}$, with Dirichlet prior and posterior beliefs. RGB colours are distributed uniformly into $16^3 = 4096$ histogram bins. Continuous palettes are normal distributions over colours, $\vec{c}$, in the $\mathrm{YP_bP_r}$ colour space, $\prod_{j=1}^{3} N(c_j \,|\, \mu_j, \lambda_j^{-1})$ with normal–gamma beliefs over μ_j and λ_j. We argue that discrete palettes are more appropriate for the background and clothing classes since we have only very broad prior knowledge about which colours will be observed in these classes, and since the colour distribution may well be multi-modal. We further argue that continuous palettes are appropriate for the skin and hair classes since their colours are known to be localised in a small chroma region (*i.e.* the $\mathrm{P_b}$ and $\mathrm{P_r}$ colour components) [4].

Figure 2 shows plots of the prior belief over skin palettes. Note that this is a distribution over distributions. Each palette is a distribution over colours; while the Bayesian belief over palettes is a distribution over distributions over colours.

The spatial prior of the model — *i.e.* the prior over the class of each pixel in an image — is an extension of the probabilistic index map of [1]. The extended model incorporates a Markov random field between neighbouring pixels, which encourages them to belong to the same class.

$$P\left(\langle z_{x,y}\rangle \,|\, \langle \vec{\pi}_{x,y}\rangle\right) = \left[\prod_{x,y} \pi_{x,y,z_{x,y}}\right] \exp\left(k_{\mathrm{MRF}} \sum_{(xy,uv)\in\mathscr{C}} \mathbb{I}\left[z_{x,y} = z_{u,v}\right]\right) \quad (1)$$

where $\mathscr{C}$ denotes the set of pairwise neighbours. The four neighbours of a pixel are directly to its north, south, east and west. The constant k_{MRF} encodes how important it is that adjacent pixels belong to the same class.

The intuition behind the spatial prior is that knowing the approximate location of, for instance, a person's hair and clothing provides a lot of information about where his or her face is located (and *vice versa*). In this work, we combine spatial and colour priors in one generative model of face images.

The prior belief for each component of the overall model is learned from labelled images from the Compaq skin database [2].

Image segmentation comes down to computing the posterior belief over the class of each pixel based on the colour model of each class. Colour modelling comes down to computing the posterior belief over the palette of each class. Figure 1 shows an example of the output of the variational inference procedure. This procedure is iterative: given a test image and starting from the prior belief over palettes, an approximate belief is computed over the class of each pixel. Then, given a belief over the class of each pixel, a new approximate belief over each palette is computed. These two steps are repeated until convergence.

In the main article we provide experimental results demonstrating the importance incorporating a Markov random field into the spatial prior and highlighting the differences between discrete and continuous palettes.

[1] N. Jojic and Y. Caspi. Capturing image structure with probabilistic index maps. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[2] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46:81–96, 2002.

[3] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37:183–233, 1999.

[4] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen. Adaptive skin color modeling using the skin locus for selecting training pixels. *Pattern Recognition*, 36:681–690, 2003.

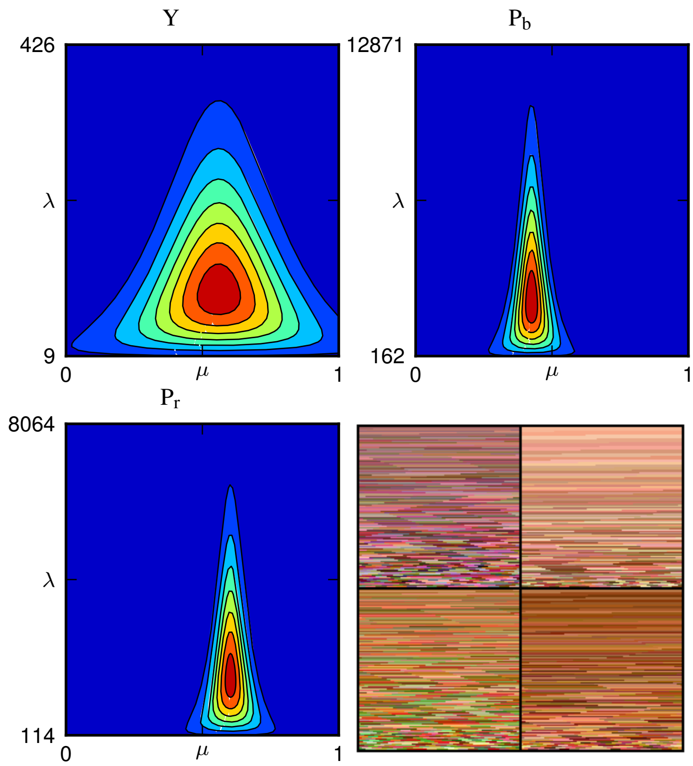

Figure 2: The prior skin colour model. The three contour plots show the prior belief over the mean and inverse variance of the normal distribution over each colour component. Also shown are four samples (*i.e.* four distributions over colours) from the prior model.