Skin detection using the EM algorithm with spatial constraints *

A. Diplaros T. Gevers N. Vlassis
Informatics Institute, University of Amsterdam
The Netherlands
{diplaros,gevers,vlassis}@science.uva.nl

Abstract – *In this paper, we propose a color-based method for skin detection and segmentation, which also takes into account the spatial coherence of the skin pixels. We treat the problem of skin detection as an inference problem. We assume that each pixel in an image has a hidden binary label associated with it, that specifies if it is skin or not. In order to solve the inference problem ,we use a variational EM algorithm which incorporates the spatial constraints with just a small computational overhead in the E-step. Finally, we show that our method provides better results than the standard EM algorithm and a state-of-art skin-detection method from the literature [9].*

Keywords: skin color, skin detection, color models, EM algorithm, unsupervised learning, segmentation

1 Introduction

Skin color is considered to be a useful and discriminating image feature for face and people detection, localization, and tracking [1] [2] [3]. Like in almost any other computer vision research field, confounding imaging conditions (e.g., change of illumination, shadows, shading and highlights) complicate the skin detection problem. In addition, the color of skin may vary among people. Furthermore, for the same person the skin color differs significantly [10] both in place (i.e. face versus hands) and in time (i.e. after long sun exposure). Numerous techniques for skin color modelling and detection have been proposed [8].

There are three broad categories of methods for skin detection and segmentation. The fist category of methods uses explicit rules on the color values [14] [13]. In general, these methods are very simple to implement and computationally inexpensive. However, they are very rigid and cannot cope with the complexity of the problem.

The second category uses an nonparametric model for skin color distributions. These methods estimate the skin color distribution from the training data without deriving an explicit model of the skin color [9] [15]. This category includes methods that build and use the skin distribution map (SMP), which is the discrete probability distribution of observed colors are skin. These methods are fast, but, require significant storage space. Furthermore, their performance depends heavily on the selection of the training set.

The third category uses parametric models for the skin color distributions. These models usually consist of a Gaussian or a mixture of Gaussians and offer a more compact skin representation along with the ability to generalize and interpolate the training data. For example in this category falls [11] [12], where, the *"perceptually plausible hue-saturation chrominance space"* TSL (Tint, Saturation, Lightness) is introduced in order to model skin pixels samples compactly by only one gaussian.

Except for the methods in the first category, almost all other methods build an extra non-skin model. In this case, the image pixels are detected as skin by comparing (using the likelihood ratio) which of their color's probability, of being skin or non-skin, is higher.

All these methods use a number of images to build their models (or derive rules) in order to detect skin pixels in an image. For a particular image the model will not coincide with the actual distribution. An interesting research direction is the combined problem of skin detection and model learning for an image. This is the problem we address in this paper. In particular using an initial skin color model, we try to estimate the actual skin color distribution in an image and learn the non-skin distribution.

The rest of the paper is organized as follows. In Section 2 we present the motivation of our approach. Section 3 describes the algorithm. In Section 4, experimental results are presented. Finally, conclusion are drawn.

2 Skin detection and the EM algorithm

In this paper, we treat the skin detection problem as an inference problem with hidden variables, for which we use the EM algorithm [4] . In particular, we assume that each pixel i has an associated label with it. The associated label is a binary variable $s_i \in \{0, 1\}$ that specifies whether that pixel is skin ($s_i = 1$) or not ($s_i = 0$). Of course, when we are given

*0-7803-8566-7/04/$20.00

an image we do not know *a priori* those labels (i.e. they are hidden). Instead, we only receive some observation for each pixel i in the form of a color vector $c_i = [r, g, b]_i$. Additionally, we may have an observation model $p(c_i|s_i)$ that relates an observed pixel color with the corresponding pixel label (e.g., a multivariate Gaussian or a mixture of Gaussians), or we can try to learn such a model from the data. The problem of skin detection is how to infer the hidden pixel labels s_i from the observed pixel colors c_i, which may or may not involve learning the model parameters.

A simple approach to the problem would be to treat each pixel independently of the other and to apply the standard EM algorithm for independent and identically distributed (iid) data in order to infer the pixel labels and (potentially) learn the model parameters [9]. However, such an approach may give suboptimal results because it neglects the correlations between the labels of neighboring pixels. For instance, if there is a face in the image, these face pixels are spatially constrained to be skin pixels.

A more principled approach to incorporate such spatial constraints into the EM algorithm would be to model the pixel labels using a hidden Markov random field (HMRF). This approach assumes a joint prior distribution $p(\{s_i\}_{i=1}^n)$ over all n pixel labels (typically parametrized by some unknown parameter) that gives high probability to configurations where neighboring pixels have similar labels. However, a well-known problem with using a HMRF for modeling and solving an image segmentation problem is its computational complexity: typically, an iterative procedure is required for computing (approximate) posterior distributions for the pixel labels [5].

Our approach lies somewhere in the middle of these two extremes. It provides a simple way to take into account the correlations between the labels of neighboring pixels, while avoiding the heavy computation cost of the HMRF.

3 Incorporating spatial constraints with a variational EM algorithm

The idea behind our method is to treat the pixel labels as independent random variables from a common prior distribution $p(s_i)$ (which we are going to learn by the EM algorithm), but constrain their posterior distributions (computed in the E-step of the EM algorithm) according to the spatial dependencies between pixels. In particular, we define a log-likelihood function:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \log \sum_{s_i} p(c_i|s_i)p(s_i) \tag{1}$$

where the parameter θ summarizes all unknown parameters in the model. These unknown parameters are learned by the EM algorithm. For example, θ may include the prior probability of the skin model component (vs. the non-skin), as well as the parameters of the two (skin and non-skin) mixture components.

In order to capture the spatial constraints of the pixel labels into an EM algorithm, we employ a *variational* approximation in which we maximize in each step a lower bound of $\mathcal{L}(\theta)$ [6, 7]. This bound $\mathcal{F}(\theta, Q)$ is a function of the current mixture parameters θ and a factorized distribution $Q = \prod_{i=1}^{n} q_i(s_i)$, where each $q_i(s_i)$ corresponds to pixel i but defines an otherwise arbitrary discrete distribution over s_i. For a particular realization of s_i we will refer to $q_i(s_i)$ as the *responsibility* of label s_i for the pixel i.

This lower bound, analogous to the (negative) variational free energy in statistical physics, can be expressed by the following two (equivalent) decompositions

$$\mathcal{F}(\theta, Q) = \sum_{i=1}^{n} [\log p(c_i; \theta) - D(q_i(s_i) \parallel p(s_i|c_i; \theta))] \tag{2}$$

$$= \sum_{i=1}^{n} \sum_{s=1}^{k} q_i(s_i)[\log p(c_i, s_i; \theta) - \log q_i(s_i)] \tag{3}$$

where $D(\cdot \parallel \cdot)$ denotes the Kullback-Leibler divergence between two distributions, and $p(s_i|c_i)$ is the posterior distribution over components of a data point c_i computed by applying the Bayes' rule:

$$p(s_i|c_i) = \frac{p(c_i|s_i)p(s_i)}{p(c_i)}. \tag{4}$$

The dependence of p on θ is throughout assumed, although not always written explicitly.

Since the Kullback-Leibler divergence between two distributions is non-negative, the decomposition (2) defines indeed a lower bound on the log-likelihood. Moreover, the closer the responsibilities $q_i(s_i)$ are to the posteriors $p(s_i|c_i)$, the tighter the bound. In particular, maxima of $\mathcal{F}$ are also maxima of $\mathcal{L}$ [6]. In the original derivation of EM [4], each E step of the algorithm sets $q_i(s_i) = p(s_i|c_i)$ in which case, and for the current value θ^t of the parameter vector, holds $\mathcal{F}(\theta^t, Q) = \mathcal{L}(\theta^t)$. However, other (suboptimal) assignments to the individual $q_i(s_i)$ are also allowed provided that $\mathcal{F}$ increases in each step.

For particular values of the responsibilities $q_i(s)$, we can solve for the unknown parameters of the mixture by using the second decomposition (3) of $\mathcal{F}$. It is easy to see that maximizing $\mathcal{F}$ for the unknown parameters of a component s yields the following solutions:

$$p(s) = \frac{1}{n} \sum_{i=1}^{n} q_i(s_i), \tag{5}$$

$$m_s = \frac{1}{np(s)} \sum_{i=1}^{n} q_i(s_i)c_i, \tag{6}$$

$$C_s = \frac{1}{np(s)} \sum_{i=1}^{n} q_i(s_i)c_i c_i^\top - m_s m_s^\top. \tag{7}$$

An attractive property of the variational EM framework is that in each step of the algorithm we are allowed to assign *any* responsibility distribution $q_i(s_i)$ to individual pixels as long as this increases the energy $\mathcal{F}$.

Figure 1: Our test image which contains multiple objects, shadows, and specularities.

In summary, our variational EM algorithm is as follows:

1. (Initialization) Start with a random guess for the parameter vector θ.
2. (Standard E-step) Compute the Bayes posterior probabilities using Eq. 4 over pixel labels given the pixel colors (as if the pixels were iid samples) given the current estimate of θ.
3. Smooth the responsibilities of neighboring pixels by applying a local filter on the set of assigned posteriors (and then renormalize if needed). An efficient way to do this is to represent the set of assigned responsibilities as an image and apply a standard Gaussian smoothing filter.
4. (Standard M-step) Use the smoothed responsibilities in order to update the parameter θ as in standard EM [4]. If convergence stop else go to step 2.

The main difference of our approach with HMRF is that instead of imposing a composite prior over the hidden variables and infer their posteriors using some (typically mean-field like) approximation, here we impose a very easy (multinomial) common prior which we learn by the EM algorithm, and constrain the posteriors via the spatial dependencies of the pixels and through control of the energy $\mathcal{F}$.

4 Demonstration

In order to demonstrate the effectiveness of our method we have conducted a skin segmentation experiment. In this experiment we assume that each component of the mixture is a multivariate Gaussian, where the skin model component is known and fixed. Our task is to learn the prior distribution $p(s)$ of the two components, skin and non-skin, while at the same time infer the pixel labels $p(s_i)$.

We have employed the skin color model of [9] which is a mixture of 16 Gaussians in the RGB color space. We have

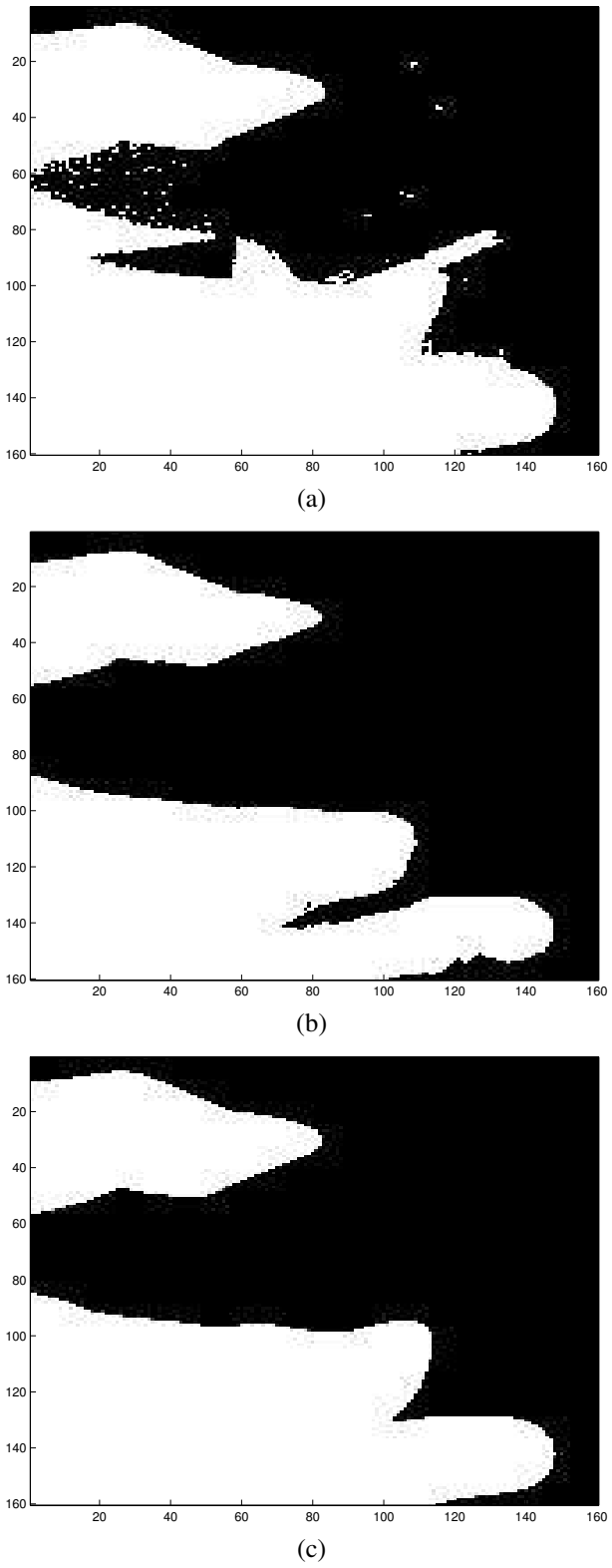

Figure 2: Skin segmentation example: (a) the segmentation result obtained with the standard EM algorithm, (b) the segmentation results obtained by [9] with parameter Θ set to 0.4 and (c) the segmentation result obtained by our algorithm.

mapped this model into the *rb* channels of the normalized-*rgb* color space (chromaticity) where the skin component model can be adequately described using only one two-dimensional Gaussian. For simplicity, we also used one Gaussian to describe the non-skin component (which we learn with the EM).

Fig. 1 shows our real-world test image, which contains multiple objects, shadows, and specularities. Fig. 2(a) shows the skin segmentation results using the standard EM algorithm (without the smoothing step 3). Note that the specularities and shadows are incorrectly classified as skin. Fig. 2(b) shows the skin segmentation results obtained by [9]. This method takes a user-defined parameter Θ which we set to 0.4 (same as used in their examples). As we can see from the image and is also stated in [9] this method *"tends to fail on highly saturated or shadowed skin"*. Fig. 2(c) shows the results of our algorithm. The results clearly show that our algorithm has correctly classified the majority of the skin pixels in the image, with the exception of a tiny, deeply shadowed skin region at the tip of the pointing finger.

5 Conclusions

We have proposed a method for skin detection and segmentation with concurrent model learning. The method incorporates the spatial constraints into the EM algorithm with just a small computational overhead in the E-step of the standard EM algorithm. Our experimental results shows that our method provides better detections results than two standard methods.

References

[1] L. Sigal, S. Sclaroff and V. Athitsos,"Skin Color-Based Video Segmentation under Time-Varying Illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*,vol. 26(6), 2004.

[2] W. Hafner and O. Munkelt, "Using Color for Detecting Persons in Image Sequences," *Pattern Recognition and Image Analysis*, vol. 7(1), pp. 47–52, 1997.

[3] Y. Raja, S.J. McKenna and S. Gong, "Tracking and Segmenting People in Varying Lighting Conditions Using Colour," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 228–233, 1998.

[4] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B*, vol. 39, pp. 1–38, 1977.

[5] G. Celeux, F. Forbes and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognition*, vol. 36, pp. 131–144, 2003.

[6] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," *Learning in graphical models*, M. I. Jordan, Ed. Kluwer Academic Publishers, pp. 355–368, 1998.

[7] J. R. J. Nunnink, J. J. Verbeek and N. Vlassis, "Accelerated greedy mixture learning," *Proc. Belgian-Dutch Conference on Machine Learning*, Jan. 2004.

[8] V. Vezhnevets, V. Sazonov and A. Andreeva, "A Survey on Pixel-Based Skin Color Detection Techniques," *Proc. Graphicon-2003*, pp. 85-92, 2003.

[9] M. J. Jones and J. M. Rehg, "Statistical Color Models with Application to Skin Detection," *Int. Journal of Computer Vision*, vol. 46(1), pp. 81–96, 2002.

[10] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori and Y. Miyake, "Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin," *ACM Transactions on Graphics*, vol. 22(3), pp. 770–779, 2003. (Proceedings of ACM SIGGRAPH 2003).

[11] J-C. Terrillon, M. David and S. Akamatsu, "Automatic Detection of Human Faces in Natural Scene Images by Use of a Skin Color Model and of Invariant Moments," *Proc. of the Third International Conference on Automatic Face and Gesture Recognition*, pp. 112–117, 1998.

[12] J-C.Terrillon, M. Shirazi, H. Fukamachi and S. Akamatsu, "Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images," *Proc. of Forth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 54–61, 2000.

[13] M. Fleck, D. Forsyth and C. Bregler, "Finding Naked People," *Proc. of the ECCV'96, 4th European Conference on Computer Vision*, pp. 593–602, 1996.

[14] G. Gomez and E. Morales, "Automatic feature construction and a simple rule induction algorithm for skin detection," *Proc. of the ICML Workshop on Machine Learning in Computer Vision*, pp. 31-38, 2002.

[15] G. Gomez, "On selecting colour components for skin detection," *Proc. of the ICPR*, vol. 2, pp. 961–964, 2002.