Markov Random Field Models for Hair and Face Segmentation

Kuang-chih Lee[*], Dragomir Anguelov[†], Baris Sumengen, Salih Burak Gokturk
Riya Inc.
3 Waters Park Drive, Suite 120, San Mateo CA 94305
{kclee, drago, baris, burak}@riya.com

Abstract

This paper presents an algorithm for measuring hair and face appearance in 2D images. Our approach starts by using learned mixture models of color and location information to suggest the hypotheses of the face, hair, and background regions. In turn, the image gradient information is used to generate the likely suggestions in the neighboring image regions. Either Graph-Cut or Loopy Belief Propagation algorithm is then applied to optimize the resulting Markov network in order to obtain the most likely hair and face segmentation from the background. We demonstrate that our algorithm can precisely identify the hair and face regions from a large dataset of face images automatically detected by the state-of-the-art face detector.

1. Introduction

In the past few decades, there has been intensive research and great strides in designing and developing algorithms for face recognition. A more complete recent literature survey of face recognition can be found in [4, 8, 22]. All those previous approaches perform recognition using face appearance alone, and only focus on analyzing the information of a small cropped face area, which is usually composed of one or more cropped rectangles around eyes, nose and lip.

However, literatures in psychology has claimed that there are also other important facial features for human visual system to recognize the face identity, such as features extracted from forehead, chin, and hair areas [5, 16]. Unfortunately, those features are not well studied to date in the computer vision communities since they are generally hard to be detected and extracted automatically. For example, detection of human hair is challenging because human hair usually contains non-rigid shape and large color variations.

The work of Yacoob and Davis [19] is the only prior work we can find on hair detection. Their approach constructs a simple color model and uses it to recognize the hair pixel. However, their detection can only work under controlled background environment and very less hair color variation. On the other hand, hair modelling, synthesis, and animation have already become active research topics in computer graphics [6, 9, 10, 12, 18].

In this paper, we present a probabilistic graphical model to perform the segmentation task of the human hair and face regions from the background. Our approach extends the traditional segmentation algorithms, such as Graph-Cut [3, 15] and Loopy Belief Propagation [20], by incorporating the color and location model. Therefore, our proposed hair and face segmentation algorithm is fully automatic without any human intervention commonly used in [2, 7, 15].

Recently, Song et al. [21] and Anguelov et al. [1] addressed new approaches to improve the performance of the standard clustering and recognition algorithms for human faces by integrating the temporal (the time photos were taken) and clothing information since people tend not to change clothes in a short period. We believe that our proposed algorithm can provide additional human hair information as a short term cue in these framework to further improve the recognition performance as well.

2. Mathematical Framework

Let $I(\mathbf{x}) = \mathbf{z}$ denote an image mapping each pixel $\mathbf{x} = [x, y]^T$ to its color value $\mathbf{z}$. The color value is usually quantized in some color space, such as the regular RGB color space or the LAB perceptual color space. Let $S(\mathbf{x}) = f_{\mathbf{x}}$ denote the segmentation, where $f_{\mathbf{x}} \in \{0, 1, 2\}$ represents the labelling, with 0 for background, 1 for the hair region, and 2 for the face region. Our goal is to solve the segmentation problem that assigns each pixel $\mathbf{x}$ a label $f_{\mathbf{x}}$, where $f_{\mathbf{x}}$ is locally smooth and consistent with the observed data. We further assume that the segmentation S is also guided by location $P(\mathbf{x}|f)$ and the color $P(\mathbf{z}|f)$ prior distributions of the labelling in the image.

In the next subsection, we first introduce how to learn the color and location distribution in our framework.

[*] Now at DigitalPersona Inc.
[†] Now at Google Inc.

978-1-4244-2154-1/08/$25.00 ©2008 IEEE

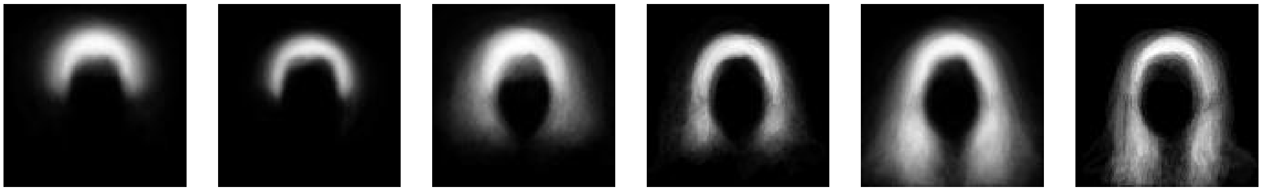

Figure 1. Probabilistic Hair Masks. Each mask can be treated as a different hair style, such as short, long, thick, and thin hairs.

2.1. Modelling of Color and Location Distribution

In the training process, the location prior probabilities $P(\mathbf{x}|f)$ are computed by counting the actual number of each labelling at pixel position $\mathbf{x}$ from the face images in the training dataset. Since the hair location is strongly dependent on the hair style, we manually cluster the ground truth images into six different hair styles, and then compute the location prior probabilities $P(\mathbf{x}|f)$ for each hair style. Figure 1 shows the hair probability mask in each hair style by stacking all the location probabilities into an image mask.

The computation of the color prior distribution contains two training steps, the offline training step and the online model update step. The offline training step is described in the following first. The natural hair color prior distribution $P(\mathbf{z}|f=1)$ generally depends on gender, ethnicity, age, and illumination conditions. We first manually cluster labelled hair pixels into five different color classes: black, brown, blond, red, and gray. The hair color distribution in each hair color class is then computed by Gaussian Mixture Model (GMM) from the labelled hair pixels. The prior face color distribution $P(\mathbf{z}|f=2)$ is also computed by GMM from labelled face pixels. In our current implementation, each GMM has 5 modes and learned by Expectation Maximization (EM) algorithm. Finally, the prior background color distribution $P(\mathbf{z}|f=0)$ is initialized as a uniform distribution.

When the test image is given, pixels are sampled from high probability areas based on location prior model $P(\mathbf{x}|f)$ for each face, hair, and background region. The background color model $P(\mathbf{z}|f=0)$ is updated by learning a GMM from the background pixels. Face pixels which disagree with the prior face color GMM model are discarded, and then $P(\mathbf{z}|f=2)$ is updated learning a GMM from the remaining face pixels. For pixels belonging to hair regions, first maximum likelihood strategy are applied to decide the hair color class based on the offline learning results, and then discard pixels which disagree with this particular hair color class GMM. Finally $P(\mathbf{z}|f=1)$ is again updated by learning a GMM from the remaining hair pixels.

Next we will describe how to apply Graph-Cut and alternative Loopy Belief Propagation by integrating the color and location prior to perform the hair and face segmentation.

2.2. Segmentation by Graph-Cut

Boykov et al. [3] address the image segmentation problem as the energy minimization problem. The energy function $E(S)$ defines the goodness measurement of the segmentation S. Typically, E can be specified as follows:

$$E(S) = R(S) + \alpha B(S). \tag{1}$$

R specifies how well pixels fit into the color and location probabilistic models under segmentation labelling S, which can be defined as:

$$R(S) = \sum_{\mathbf{x}} -\log P(I(\mathbf{x})|f_{\mathbf{x}}) - \beta \log P(\mathbf{x}|f_{\mathbf{x}}), \tag{2}$$

where $P(I(\mathbf{x})|f_{\mathbf{x}})$ denotes the color likelihood probability, and $P(\mathbf{x}|f_{\mathbf{x}})$ denotes the location likelihood probability described in Subsection 2.1. β denotes the relative importance of the color and location probability terms. We found that good results are obtained with the setting $\beta = 1$.

The second term B in Equation 1 incorporates the smoothness of the nearby pixels into segmentation S, and can be denoted by the image gradient in the RGB or LAB color space around the segmentation boundary:

$$B(S) = \sum_{\mathbf{x}_i,\mathbf{x}_j \in \Omega} \delta(S(\mathbf{x}_i) \neq S(\mathbf{x}_j)) \exp(-\gamma \|I(\mathbf{x}_i) - I(\mathbf{x}_j)\|^2), \tag{3}$$

where $\delta(S(\mathbf{x}_i) \neq S(\mathbf{x}_j)) = 1$ if $\mathbf{x}_i$ and $\mathbf{x}_j$ have different labelling, and otherwise it is 0. Ω denotes the set of neighboring pixels, and γ denotes a normalized constant and can be defined by the inverse of the average contrast between neighboring pixels.

Note that the constant term α in Equation 1 represents the relative importance between the two energy terms R and B, and it can be picked by a regression test over a large training dataset. We found that good results are obtained with the setting $\alpha = 0.5$.

Given an input face image, the goal is to seek the optimal segmentation S^* by minimizing the energy function E:

$$S^* = \arg\min_{S} \; E(S), \tag{4}$$

Boykov et al. [3] have shown how to solve this minimization problem by constructing a graph such that the minimum cut is the optimal segmentation labelling. In addition, the minimum cut can be computed efficiently with an iterative procedure called *alpha-expansion*.

Carsten Rother et al. [15] have also shown that this energy minimization procedure can be applied iteratively to improve the segmentation performance, because the segmentation result in the previous iteration can be used to refine the color GMM parameters for the hair, face, and background regions.

2.3. Segmentation by Loopy Belief Propagation

In this subsection, we reformulate the segmentation problem as the Markov network inference, and present another alternative algorithm, Loopy Belief Propagation (LBP), to compute optimal hair and face segmentation based on the color and location prior model.

We build the pairwise Markov network with the likelihood given in the following form:

$$P_\phi(S) = \frac{1}{Z}\prod_{i=1}^{N}\phi_i(S(\mathbf{x}_i))\prod_{\mathbf{x}_i,\mathbf{x}_j\in\Omega}\phi_{ij}(S(\mathbf{x}_i),S(\mathbf{x}_j)), \quad (5)$$

where Z is a normalization constant, also called partition function, given by $Z = \sum_{S'}\prod_{i=1}^{N}\phi_i(S'(\mathbf{x}_i))\prod_{\mathbf{x}_i,\mathbf{x}_j\in\Omega}\phi_{ij}(S'(\mathbf{x}_i),S'(\mathbf{x}_j))$. The singleton potential function $\phi(S(\mathbf{x}))$ represents the likelihood of the labelling assignment, and is denoted as following product of two scores:

$$\phi(S(\mathbf{x})=f) = \phi_{color}(I(\mathbf{x}))\phi_{location}(I(\mathbf{x})), \quad (6)$$

where $\phi_{color}(I(\mathbf{x}))$ and $\phi_{location}(I(\mathbf{x}))$ quantify how well the color and location prior model match the segmentation hypothesis. We simply define $\phi_{color}(I(\mathbf{x})) = P(I(\mathbf{x})|f)$, and $\phi_{location}(I(\mathbf{x})) = P(\mathbf{x}|f)$, where $P(I(\mathbf{x})|f)$ and $P(\mathbf{x}|f)$ denote the location likelihood probability described in previous Subsection 2.1.

The pairwise potential function $\phi_{ij}(S(\mathbf{x}_i),S(\mathbf{x}_j))$ represents a joint configuration where pixel $\mathbf{x}_i$ has the labelling $S(\mathbf{x}_i)$ and pixel $\mathbf{x}_j$ has the labelling $S(\mathbf{x}_j)$:

$$\phi_{ij}(S(\mathbf{x}_i),S(\mathbf{x}_j)) = \begin{cases} \max(\exp(-\gamma\|I(\mathbf{x}_i)-I(\mathbf{x}_j)\|^2),\epsilon), \\ \qquad\qquad \text{if } S(\mathbf{x}_i)=S(\mathbf{x}_j) \\ \epsilon, \qquad \text{if } S(\mathbf{x}_i)\neq S(\mathbf{x}_j) \end{cases} \quad (7)$$

where ϵ denotes a small constant to ensure there is a uniform probability for all the possible labelling in the neighborhood pixels. This is particularly useful when the probabilistic term of the image gradient $\exp(-\gamma\|I(\mathbf{x}_i)-I(\mathbf{x}_j)\|^2)$ is noisy.

The segmentation problem now can be treated as the *maximum a posteriori* (MAP) inference problem in a

---

Probabilistic Hair and Face Segmentation Algorithm:

Input: (I)
I: input registered face image
Output: (S, H_s)
S: Hair and face segmentation result of I
H_s: Hair style
Model Parameters: ($M_{H_s}, \theta_{H_c}, \theta_f$)
$\{M_{H_s}|H_s = 0,\ldots,5\}$: The probabilistic mask for the location prior in Figure 1
$\{\theta_{H_c}|H_c = 0,\ldots,4\}$: GMMs for each trained hair color class H_c
θ_{F_c}: GMM for the trained face color model
$\{\theta_f|f = 0,\ldots,2\}$: GMMs for online hair, face, and background models
Offline Training:
A set of 150 manually labelled groundtruth face images are used to train the location prior model, $P(\mathbf{x}|f)$ by computing the probabilistic mask M_{H_s}, as well as face and hair color GMM model parameters, θ_{H_c} and θ_{F_c}.
Begin

1. **Initialization**: Sample pixels from high probability areas of background, face, and hair regions based on $\{M_{H_s}\}$ in the input image I. Discard pixels in the face regions which disagree with face color model GMM θ_{F_c}. For pixels belonging to hair regions, apply maximum likelihood strategy to decide the hair color class H_c^* based on GMMs $\{\theta_{H_c}\}$. Discard hair pixels again which disagree with this hair color class H_c^*.
2. **Online Updating Models**: For pixels assign to face, hair, and background regions, use the Expectation Maximization algorithm to learn the GMMs $\{\theta_f\}$ for face, hair, and background regions with five random seeds initialization.
3. **Estimating Segmentation** S: Use Graph-Cut or Loopy Belief Propagation algorithm to seek the optimal segmentation S by evaluating Equation 4 or Equation 8.
4. **Looping back to Step 2 till Convergence**
5. **Estimating Hair Style (Optional)** : Use maximum likelihood strategy to perform the hair style estimation based on the distance between the segmentation result S and hair location mask M_{H_s}.

End

Figure 2. Summary of the proposed segmentation algorithm.

Markov network:

$$S^* = \arg\max_S \; P_\phi(S). \tag{8}$$

The Loopy Belief Propagation can be directly applied to solve the MAP problem in Equation 8. However, when the size of image becomes larger, the large amount of nodes and edges in the Markov network make the LBP inference process quite expensive, and also increase the chance to non-convergency. Therefore, in our current implementation, we first pre-process all the images by merging pixels with their similar neighborhoods to M superpixels [11, 14] ($M = 60$ in our current implementation), then construct the Markov network on top of superpixels, and finally solve the inference problem by the standard LBP algorithm.

After we obtain the segmentation result, we can also estimate the hair style, such as long or short hair, by applying maximum likelihood strategy based on the distance between the segmentation result and hair masks shown in Figure 1. The detailed segmentation algorithm is summarized in Figure 2.

3. Experiments and Results

3.1. Ground Truth Collection and Data Preparation

In order to train the color and location prior model and evaluate the performance of our framework, we manually labelled hair, face, and background regions of 150 face images as a ground truth dataset. These face images plus another 5000 testing images are detected by the face and facial features (eye, nose, and lip corners) detectors based on the enhanced state-of-the-art ada-boost detector [13] and Bayesian inference registration process [17], and then geometrically aligned and cropped to our standard formats. Our standard formats include high resolution 300×300 and low resolution 128×128. There are 1000 labelled test images with high resolution 300×300 for qualitative and quantitative evaluation, and the rest 4000 images with low resolution 128×128 for qualitative evaluation only.

3.2. Hair and Face Segmentation Results

Figure 3 displays the qualitative comparisons of the segmentation results with the simple color clustering algorithm proposed by Yacoob and Davis [19]. The simple color clustering algorithm fails to generate accurate hair segmentation because the color variation of the blond hair is large in the first example, and the hair and cloth have similar color in the second example. Our proposed model considers both location and color information in a principled way, and thereby leads to better segmentation results.

The error rates based on the number of pixels that were misclassified for our proposed Graph-Cut algorithm, LBP algorithm, and the simple color clustering algorithm [19] are 8.8%, 9.4%, and 38.4%, respectively. We also test our hair classification algorithm by grouping hair length into short and long two hair classes based on the human judged fact if the hair touches the shoulder or not. The error rate of the hair length classification based on Graph-Cut and LBP segmentation results are 3.1% and 3.8%. Figure 5 and Figure 6 show more segmentation results in the low resolution dataset.

Both of our proposed algorithms demonstrated good results on the challenging conditions: less color contract between hair and background regions, hair with smooth color variations, and hair color similar to facial skin color. Our proposed Graph-Cut algorithm performed slightly better than the LBP algorithm in terms of the quality and speed. Our explanation is that the performance of current LBP implementation really depends on the superpixel initialization. The merging process of superpixels currently is only based on the color gradient and therefore it is easier to accumulate errors.

Most of the failure cases happened when the images are dark. Figure 4 shows one failure example. The color model failed to discriminate the hair and background in this case, but the location model still captured the hair shape so that all the dark background pixels were not assigned to hair. Other failure cases we have seen include images with a bold head, and images with dramatic changes of multiple hair colors.

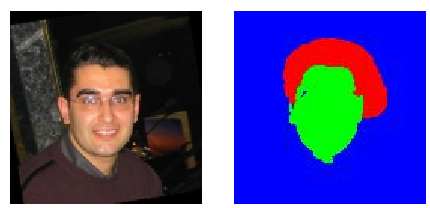

Figure 4. One failure example generated by our proposed Graph Cut algorithm. The hair looks thicker since the background color is as black as the hair color.

4. Summary and Conclusions

We have proposed a probabilistic hair and face segmentation algorithm to fully automatically segment face and hair regions from the background. We have demonstrated that our segmentation algorithm is effective for a large testing face dataset under challenging conditions. Future steps would be to generalize this algorithm for profile views and integrate our hair and face segmentation algorithm with the proper feature extractor in the face recognition system.

Acknowledgement

We would like to thank Vincent Vanhoucke, Lorenzo Torresani, Danny Yang, and Diem Vu for the help of discussion and data processing.

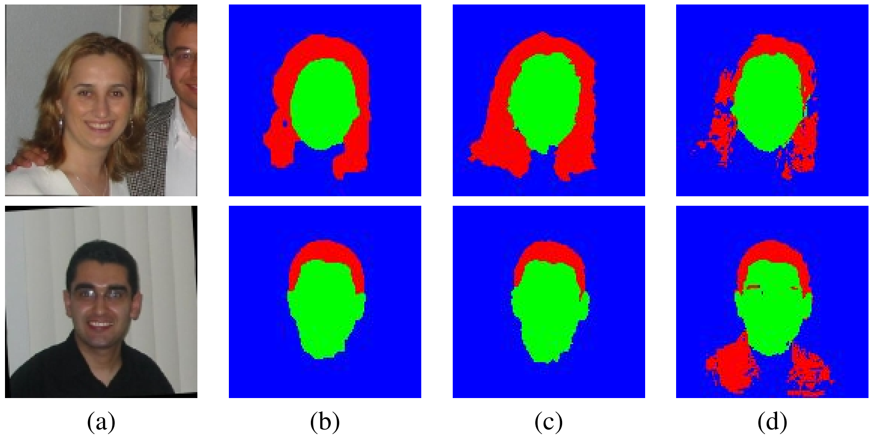

Figure 3. Comparison of hair and face segmentation results. **(a)**: The input image. **(b)**: Hair and face segmentation results by our proposed Graph-Cut Algorithm. **(c)**: Hair and face segmentation results by our proposed LBP Algorithm.**(d)**: Hair and face segmentation results by the simple color clustering algorithm proposed by [19].

References

[1] D. Anguelov, K. Lee, S. B. Gokturk, and B. Sumengen. Contextual identity recognition in personal photo albums. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2007.

[2] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. *Proc. Int'l. Conf. on Computer Vision*, 2001.

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approaximate energy minimization via graph cuts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11), 2001.

[4] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740, 1995.

[5] G. Davies, H. Ellis, and J. Shepherd. Perceiving and remembering faces. *London: Academic Press, Academic Press series in cognition and perception*, 1981.

[6] J. Kajiya and T. Kay. Rendering fur with three dimensional textures. *Computer Graphics*, 23(3):271–280, 1989.

[7] A. Levin, D. Lischinski, and Y. Weiss. A closed form solution to natural image matting. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2006.

[8] S. Z. Li and A. K. Jain. Handbook of face recognition. *Springer*, 2004.

[9] S. Marschner, H.W.Jensen, M.Cammarano, S. Worley, and P. Hanrahan. Light scattering from human hair fibers. *Proc. SIGGRAPH*, pages 780–791, 2003.

[10] J. T. Moon and S. R. Marschner. Simulating multiple scattering in hair using a photon mapping approach. *Proc. SIGGRAPH*, 2006.

[11] G. Mori, X. Ren, A. Efros, and J. Malik. Recovering human body configurations: Combining segmentation and recognition. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2004.

[12] S. Paris, H. M. Briceno, and F. X. Sillion. Capture of hair geometry from multiple images. *Proc. SIGGRAPH*, 2004.

[13] P.Viola and M.Jones. Robust real-time object detection. In *Int'l. J. Computer Vision*, 2001.

[14] X. Ren and J. Malik. Learning a classification model for segmentation. *Proc. Int'l. Conf. on Computer Vision*, 2003.

[15] C. Rother, V. Kolmogorov, and A. Blake. Grabcut - interactive foreground extraction using iterated graph cuts. *Proc. SIGGRAPH*, 2004.

[16] P. Sinha and T. Poggio. United we stand: The role of head-structure in face recognition. *Perception*, 31(1), 2002.

[17] E. Sudderth, A. Ihler, W. Freeman, and A. Willsky. Nonparametric belief propagation, 2002.

[18] Y. Wei, E. Ofek, L. Quan, and H.-Y. Shum. Modelling hair from multiple views. *Proc. SIGGRAPH*, 2005.

[19] Y. Yacoob and L. S. Davis. Detection and analysis of hair. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(7), 2006.

[20] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. *Int'l Joint Conf. on Artificial Intelligence*, 2001.

[21] Y.Song and T.Leung. Context-aided human recognition - clustering. *In Proc. ECCV*, 3:382–395, 2006.

[22] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.

Figure 5. Hair and face segmentation results on low resolution images by our proposed Graph-Cut Algorithm.

Figure 6. Hair and face segmentation results on low resolution images by our proposed LBP Algorithm.