Automatic Hair Detection in the Wild

P. Julian[1,2] C. Dehais[2] F. Lauze[3] V. Charvillat[1] A. Bartoli[4] A. Choukroun[2]

[1] *IRIT, Toulouse* [2] *FittingBox, Toulouse* [3] *DIKU, Copenhagen* [4] *UdA, Clermont-Ferrand*

(a) Active contour as in [2] (b) Active shape as in [4] (c) Our method

Figure 1. Robust hair detection by combining active contours and active shapes.

Abstract

This paper presents an algorithm for segmenting the hair region in uncontrolled, real life conditions images. Our method is based on a simple statistical hair shape model representing the upper hair part. We detect this region by minimizing an energy which uses active shape and active contour. The upper hair region then allows us to learn the hair appearance parameters (color and texture) for the image considered. Finally, those parameters drive a pixel-wise segmentation technique that yields the desired (complete) hair region. We demonstrate the applicability of our method on several real images.

1 Introduction

Several problems related to face analysis can benefit from automatic hair detection and segmentation, *e.g.* face detection and face recognition [13]. Our own application is a virtual glasses trying on solution[1], currently used by a number of online glasses retailers. Hair detection allows for a more realistic trying on, in which occluding hair parts (for example a lock of hair) can be rendered on top of the virtual glasses. As our application is aimed at the general public, users can upload any kind of pictures meaning that we cannot control aspects of the image quality such as pose, illumination conditions and resolution. Our approach should therefore be automatic and robust to these varying conditions.

Hair detection and analysis has attracted several researchers in the past few years. Yacoob and Davis [13] have used a hair color model that accounts for highlights and shadows and proposed an *ad-hoc* method improving in person identification tasks. The disambiguation of hair and skin regions is greatly improved by considering texture in addition to color, as shown in e.g. [6, 10, 11]. Including a prior on the shape of the region of interest is also known to help segmentation. In [8] Lee *et al.* use learnt probability maps for the hair location in a Markov Random Field framework.

A general appearance model for hair seems however beyond reach as hair exhibits huge variability, both in texture, color, shape, to name just a few. Instead, we may learn hair appearance of one given image, by extracting a small amount of hair patches, and use it for segmenting the hair in this image. This is the approach we took in [6], producing high quality segmentation. Thus the problem becomes to reliably find such patches. In this paper we describe a way to achieve it by segmenting automatically a specific region containing hair patches, thus "filling in the hole" , obtaining a complete hair segmentation approach. This is done by building and fitting a statistical model of the region lying at the top of the skull, which we call the Upper Hair Shape Model (UHSM). Patches of hair will thus be found *inside* the Upper Hair Shape (UHS) and will be used to initialize the segmentation described in [6].

Fitting the UHSM to a given image is performed through a variational method close to the approach of Leventon *et al.* [9] which aims to combine active contours [7, 1] with active shapes [4]. Leventon's approach relies on boundary terms but these terms between hair and background / skin can be difficult to define. We instead use region homogeneity terms à la Chan and Vese [2].

This paper is organized as follows. In the next sec-

[1] `http://www.fittingbox.fr/demo/`

tion we describe our statistical model and fitting procedure. In Section 3 we demonstrate its validity and we conclude in the last section.

2 Hair Shape Model and Active Contours

2.1 UHSM Learning

We follow [4] and [9] in order to learn our UHSM. A number of example upper hair shapes have been manually annotated from a collection of $N = 60$ images and their mirrored versions, and $n = 22$ points were used to annotate each individual shape.

Shapes are aligned and scaled with respect to the two outer eye corners. We then perform Principal Component Analysis on these 60 normalized vectors. The first four eigenvectors (components) are kept. Their associated eigenvalues represent 87% of the total variance and this showed to be sufficient in our application. The corresponding four parameters happen to model well the head orientation up/down (α_1), right/left (α_2), the hair thickness (α_3), face morphology (α_4), as illustrated in Figure 3. A choice of these parameters generates the

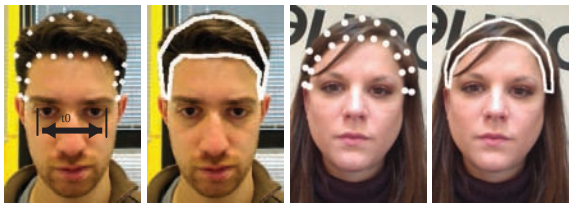

Figure 2. Examples of our UHSM.

vertices of a polygon. On a common image domain Ω a signed distance function to that polygon is built, positive inside the polygon, negative outside. Such functions are depicted in the last row of Figure 3 .This family of distance functions is our UHSM

$$u^* : \mathbb{R}^4 \times \Omega \to \mathbb{R}, (\alpha, x) \mapsto u^*(\alpha, x)$$

where $\alpha = (\alpha_1, \ldots, \alpha_4)$ is the vector of values for the four principal components (cf Figure 3).

2.2 Shape Fitting Energy

We want to extract a region from the image I on domain Ω which should be as large as possible while having an homogeneous content and being close enough to an UHSM. Leventon *et al.* proposed to adapt the descent direction of the geodesic active contours energy to the shape and pose parameters of their active shape

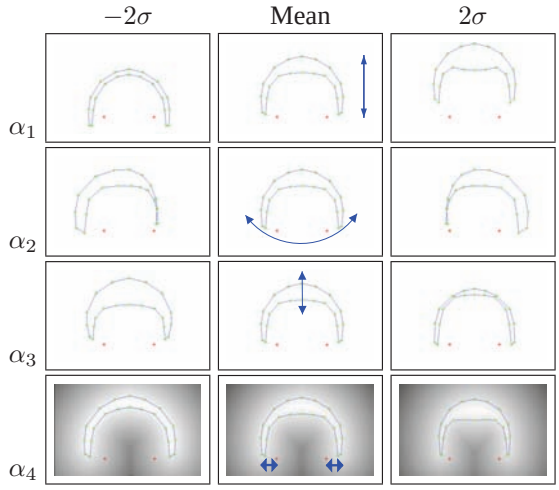

Figure 3. Upper hair eigenshapes.

model. We instead build a four terms energy that incorporates active contours and active shapes. We denote by u a function whose 0-level set is the sought contour while u^* will always denote the distance function associated with the UHSM with parameter α.

The proposed energy is :

$$\begin{aligned} E(u, \alpha, c) &= E_b(u) && (1)\\ &+ E_d^1(u, c) && (2)\\ &+ E_a(u, \alpha) && (3)\\ &+ E_d^2(\alpha). && (4) \end{aligned}$$

We briefly describe the tree first terms. Term (1) is a balloon force term and encourages the region $u > 0$ to grow. It takes the form

$$E_b(u) = -\lambda_b \int_\Omega H_\varepsilon(u(q))\, dq,$$

where H_ε is the regularized Heaviside function proposed in [2]. Term (2) encourages homogeneity of the region $u > 0$ as in [2] and takes the form

$$E_d^1(u, c) = \lambda_{d1} \int_\Omega \|I(q) - c\|^2 H_\varepsilon(u(q))\, dq,$$

where I is the image to be segmented. It is well known that the optimal c is thus the average value (color) of the region $u > 0$. Term (3) penalizes the deviation between u and u^*, thus coupling active contours and active shapes. We choose a simple one,

$$E_a(u, \alpha) = \lambda_a \int_\Omega (u(q) - u^*(q))^2\, dq.$$

The weights λ_b, λ_{d1} and λ_a are strictly positive.

The last term encodes the dependency between the UHSM and the observed image. Having the parameters of the UHSM and the outer eye corner locations allows us to extract two other regions where we expect to find respectively skin from the face and background (non-face region). This is illustrated in Figure 4. The background region B is a 30×110 pixels rectangle located 5 pixels above the UHSM region, denoted by H in the following. The skin region S is a 11×80 pixels rectangle and depends only on the eyes corners position. It is located below them, incorporating pixels from the upper cheeks and nose.

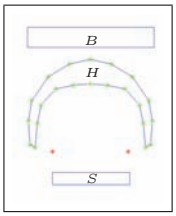

Figure 4. Hair, background and skin regions.

These three regions are expected to have a significantly different content. On each of the three regions H, S and B, we compute color histograms extracted in the $YCbCr$ color space, h^R_{CbCr} and h^R_{YCb}, $R = H, S, B$. The Bhattacharya distance d_B is then used to define a dissimilarity measure between pairs of histograms, as follows :

$$E^2_d(\alpha) = -\lambda_{d2} \sum_{R=B(\alpha)S} d_B(h^{H(\alpha)}_{CbCr}, h^R_{CbCr}) + d_B(h^{H(\alpha)}_{YCb}, h^R_{YCb}),$$

in order to maximize the dissimilarity between H and B and between H and S(λ_{d2} is positive).

2.3 Energy Minimization

The dependency of $E(u, \alpha, c)$ in α is somehow complicated. In order to approximate a minimizer of it, we proceed iteratively as follows. Given (u^n, c^n, α^n) at iteration n,

1. Set c^{n+1} to the average of I in region $u > 0$.
2. With c^{n+1} and u^n fixed, set α^{n+1} as the minimizer of $E(u^n, \alpha, c^{n+1})$ (we use the simplex method).
3. Set $u^{n+1} = u^n - dt\nabla_u E$ where $\nabla_u E$ is the gradient of E with respect to u and is given by

$$\nabla_u E = \delta_\varepsilon(u)\left(-\lambda_b + \lambda_{d1}\|I - c\|^2\right) + 2\lambda_a(u - u^*).$$

We start with the $\alpha^0 = \alpha_{init}$, defined in section 3.1 and $u^0 := u^*(\alpha^0)$. We iterate until convergence of the sequence (α^n).

3 Implementation and Results

We implemented the method described in section 2 in a complete system which also includes as a final step the learning-based approach described in [6]. The following section summarizes the system and sections 3.2 and 3.3 present some results obtained on uncontrolled images which are similar to those sent by actual users of our services.

3.1 Overview of the Complete System

Face features detection: First, we detect the face and eyes in the image using cascade classifiers [12] that we trained on faces and eyes samples. Then, we fit a Constrained Local Model [5] learned from faces.This training data overlaps the one of our UHSM and the faces are geometrically aligned on outer eye corners. The HUSM is then initialized for position and scale, we use eyes positions found with the classifiers and refined with CLM fitting, for an initial guess of shape parameters α, we use the CLM shape found at convergence. As the CLM model is made to partially overlap the UHSM at two points, t1 and t2, located on the face temples, an initial shape α_{init} is easily obtained from the fitted CLM as :

$$\alpha_{init} = \underset{\alpha,\ \alpha_3=0}{\operatorname{argmin}} \left(\|t_1{}^{\text{UHSM}}(\alpha) - t_1{}^{\text{CLM}}\|^2 + \|t_2{}^{\text{UHSM}}(\alpha) - t_2{}^{\text{CLM}}\|^2\right).$$

UHSM fitting: The USHM is fitted to the image as described in section 2.

Pixel-wise hair segmentation: At this point, we have a hair region where small patches of hair pixels can be reliably grabbed for use in the method described in [6]. In this method, those patches are used to learn an image-adapted appearance model of the hair color and texture. The final hair segmentation is then obtained by a Bayesian classification method.

3.2 UHS Detection

Several results of the UHS detection are presented in figure 5, where the initial UHS is in red and the final UHS is in blue. Those results show that the method is robust to varying hair shapes and background settings.

Constraining the model to the top of the head prevents the detection from slipping into long hair strands as it was the case in Figure 1 (b) for instance and (then) into the background 1(a), and at the same time, the shape boundaries are flexible enough to fit characteristic head features like an hair split, as in the bottom right image in Figure 5, and allow for a good separation of hair and skin with similar colors as in the top right image of 5.

Figure 5. Some UHSM fitting results.

3.3 Segmentation Results

Figure 6 shows our pixel-wise hair segmentation results which use color and texture descriptors learnt on the current image [6]. Notice first that the application of this step allows to grow the hair region to include long hair, as shown in top and bottom middle images. In the top right image, we correctly detect the hair region without labelling as "hair" the person's clothes. Remaining that the segmentation errors could be cleaned up by reintegrating spatial context (e.g. hair compactness in top right image and eyes location in bottom right image). Also note that, in the bottom left image, the head cast shadow is correctly not labeled as "hair".

Figure 6. Pixel wise segmentation results

4 Conclusion and Perspective

In this paper, we have presented a complete system for hair detection in uncontrolled, real-life conditions images. The system builds upon the method of [6], which produces high quality segmentations but requires manually provided hair patches. Our main contribution is therefore a method for robustly detecting a region containing such patches, by integrating a statistical shape model fitting in an active contour framework. We showed that our method provides a robust hair labelling. As a perspective of this work, we would like to experiment with enhancing this labelling using digital matting techniques like [3].

References

[1] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision (IJCV 97)*, 22:61–79, 1997.

[2] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10:266–277, 2001.

[3] Y.-Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski. A bayesian approach to digital matting. In *Computer Vision and Pattern Recognition (CVPR 01)*, volume 2, pages 264–271, 2001.

[4] T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[5] D. Cristinacce and T. Cootes. Feature detection and tracking with constrained local models. In *British Machine Vision Conference (BMVC 06)*, pages 929–938, Edinburgh, UK, 2006.

[6] P. Julian, V. Charvillat, C. Dehais, and F. Lauze. On the interest of texture for face segmentation. In *Orasis*, 2009.

[7] M. Kass, A. Witkin, and D. Terzopoulos. Snakes : Active contour models. *International journal of computer Vision*, pages 321–331, 1988.

[8] K. Lee, D. Anguelov, B. Sumengen, and S. Gokturk. Markov random field models for hair and face segmentation. In *Automatic Face and Gesture Recognition (FG08)*, pages 1–6, 2008.

[9] M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Computer Vision and Pattern Recognition (CVPR 00)*, volume 1, page 1316, Los Alamitos, CA, USA, 2000. IEEE Computer Society.

[10] U. Lipowezky, O. Mamo, and A. Cohen. Using integrated color and texture features for automatic hair detection. In *Convention of Electrical and Electronics Engineers in Israel*, 2008.

[11] C. Rousset and P. Y. Coulon. Frequential and color analysis for hair mask segmentation. In *International Conference on Image Processing (ICIP 08)*, pages 2276–2279. IEEE, 2008.

[12] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision (IJCV 02)*, 2002.

[13] Y. Yacoob and L. Davis. Detection, analysis and matching of hair. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, pages 741–748, Washington, DC, USA, 2005. IEEE Computer Society.