
There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.

http://eprints.gla.ac.uk/208868/

Deposited on: 28 January 2020

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk
FIGARO, HAIR DETECTION AND SEGMENTATION IN THE WILD

Michele Svanera, Umar Riaz Muhammad, Riccardo Leonardi, and Sergio Benini

Department of Information Engineering
University of Brescia, Italy
Email: {m.svanera005, riccardo.leonardi, sergio.benini}@unibs.it

ABSTRACT

Hair is one of the elements that mostly characterize people appearance. Being able to detect hair in images can be useful in many applications, such as face recognition, gender classification, and video surveillance. To this purpose we propose a novel multi-class image database for hair detection in the wild, called Figaro. We tackle the problem of hair detection without relying on a-priori information related to head shape and location. Without using any human-body part classifier, we first classify image patches into hair vs. non-hair by relying on Histogram of Gradients (HOG) and Linear Ternary Pattern (LTP) texture features in a random forest scheme. Then we obtain results at pixel level by refining classified patches by a graph-based multiple segmentation method. Achieved segmentation accuracy (85%) is comparable to state-of-the-art on less challenging databases.

Index Terms— image texture analysis, image segmentation, hair detection, hair database

1. INTRODUCTION

Hair is one of the defining characteristics of humans. Even if attitudes towards hair, such as hairstyles and hair removal, vary widely across different cultures and historical periods, hairstyle is often related to person’s social position, such as age, gender, or religion [1]. Hair detection in images is useful for face recognition and retrieval [2], gender classification [3], head detection [4], surveillance applications [5], and the analysis of shot types in videos [6].

Despite being such an important feature of human appearance, hair detection and modelling have not been widely studied in computer vision, probably because it is a challenging task due to the variety of hair patterns and the confusion between hair and the background. To ease the problem, most of previous approaches on hair detection assume faces in frontal view, and infer hair location with the help of a-priori known spatial information. A typical detection chain includes a face and/or eye zone detection [7, 8, 2] followed by a color modelling scheme for discriminating skin vs. hair. A variant to this procedure introduces the use of probability maps based on head shape and position, such as in [9, 10, 11].

Only recently few approaches remove the restriction of analysis to frontal view. Wan et al. in [12] propose a coarse-to-fine hair segmentation method on near-frontal and multi-pose head-shoulder database by using a combination of probability maps, fixation points, and graph-based segmentation, while authors in [5] combine fast head detection and background modelling working also from back views.

An obstacle which probably limits research on this topic is the lack of a shared database annotated for hair detection from unconstrained view. Most repositories available for facial features [13, 7] contain only frontal or near-frontal views and are normalized according to face or head size. Databases for texture analysis do not usually include hair texture, except rare cases, such as the class braided in [14]. Finally, with the exception of OpenSurfaces [15], which is however lacking further subdivision into hairstyles, no database provides properly segmented hair images from unconstrained view.

As a first contribution, in this paper we make available Figaro, the first multi-class image database for hair detection in the wild1. It contains 840 unconstrained view images with persons, subdivided into seven different hairstyles classes (straight, wavy, curly, kinky, braids, dreadlocks, short-men), where each image is provided with the related manually segmented hair mask. Examples of database images and related ground-truth masks are given in Figure 1, while a complete description of the database is given in Section 2.

As a second contribution, in this paper we tackle hair detection and segmentation on unconstrained view images contained in Figaro without using any human-body part classifier (face, head-and-shoulder, etc.) nor any a-priori information about head shape and location. A first step classifies image patches into hair vs. non-hair by relying only on texture characteristics. In particular Section 3 describes how Histogram of Gradients (HOG) [16] and Linear Ternary Pattern (LTP) [17] features are exploited to feed a Random Forest classifier [18]. A second stage performs hair segmentation at pixel level (as described in Section 4) by combining results obtained at patch level with those obtained by a graph-based segmentation [19]. After presenting results in Section 5, conclusion and future work are drawn in Section 6.

1http://projects.1-ctm.eu/en/project/figaro
Fig. 1. Example from Figaro, a multi-class hair image database. First row: examples from the seven hair-style classes (straight, wavy, curly, kinky, braids, dreadlocks, short-men). Second row: the related ground-truth masks.

2. DATABASE DESCRIPTION

To perform hair detection in the wild we build - and now make publicly available - a database using Google Images [20] for the retrieval process. After a manual filtering for removing errors and not suitable samples, Figaro eventually contains 840 head-shoulder images, covering diverse hair color, with no constrains on the angle of shooting, complex indoor and outdoor backgrounds, and trying to keep an equal balance between female and male images.

Since hair exists in a variety of textures, depending on curl pattern, volume, and consistency, Figaro is the first multi-class hair database where images are organized in the following classes: straight, wavy, curly, kinky, braids, dreadlocks and short-men, each containing 120 images. This categorization extends the four types of Andre Walker’s taxonomy (straight, wavy, curly, kinky) to three extra categories (braids, dreadlocks and short-men), to broaden the scope of the analysis.

The ground-truth of hair regions is manually labeled using a commercial image editor to select a binary mask associated to hair location in the image (white for hair, black for non-hair), as shown in the examples of Figure 1 (second row) for all class examples.

Being images different in size and aspect ratio, a normalization procedure is applied. Since individuals are shot from all angles, it is not possible to carry out typical normalization procedures used for frontal views, such as using the head rectangle or applying a fixed distance between the eyes. Here the employed normalization factor is chosen so as to reduce the size of the maximum squared area contained in the white region of the ground-truth mask to $250 \times 250$ pixels. For this reason images containing only hair regions smaller than $250 \times 250$ pixels are a-priori discarded from the database.

3. HAIR PATCH CLASSIFICATION

The first processing step aims at discriminating patches of hair from those which are not. We tackle such a binary classification task by a pipeline which involves a feature extraction phase, followed by a machine learning method to train a classifier, which eventually returns a binary image where patches are classified as hair or non-hair, as shown in Figure 2.

Fig. 2. System workflow for hair patch classification.

Detecting the presence of hair is accomplished without relying on any a-priori information about head location, but only making use of texture descriptors. In literature textures are used for effective classification of material and visual traits: filter banks in [21], Gabor wavelets in [2], convolution kernels in [22], and a combination of features including multi-orientation Gabor filters in [23]. Recent developments in texture classification [14] use Convolutional neural networks (CNN) at early stages on fixed $227 \times 227$ patches, size which is hardly compatible with images in Figaro.

In this work we explore two classic texture features: Histogram of Oriented Gradient (HOG) and Local Ternary Pattern (LTP), both working on patches of small dimensions. Originally developed for pedestrian detection [16], HOG technique counts occurrences of gradient orientation in image local regions, and has become a feature widely used in object detection. LTP, originally proposed in [17] as a generalization
of the Local Binary Pattern (LBP) [24], can be considered as a way of summarizing local gray-level structure. Given a local neighborhood around each pixel, LTP thresholds the neighborhood at the value of the central pixel ($\pm$ a threshold $thr$) and uses the resulting binary codes as local descriptors. To reduce the number of bins when histogramming the feature, we adopt the uniform pattern extension which assigns all non-uniform patterns to a single bin.

Texture descriptors in different combinations, HOG, LTP, and a concatenation HOG+LTP, feed a Random Forest (RF) classifier [18], a discriminative model made by an ensemble of randomly trained decision trees. Training is performed on patches extracted on 700 randomly sampled images (100 per class), leaving 140 for testing, and averaging on ten iterations. Tests are conducted on all combinations of the following parameters: HOG ($block_{dim}$ = \{5, 15, 25, 35, 50, 75, 100\}, $bins$ = \{4, 8, 12\}, $cell_{dim}$ = \{5, $block_{dim}$\}); LTP ($block_{dim}$ = \{15, 25, 35, 50, 75, 100\}, $thr$ = \{0.02, 0.05, 0.07\}); HOG+LTP ($block_{dim}$ = \{25, 35\}, $cell_{dim}$ = $block_{dim}$, $bins$ = 12, threshold $t$ = 0.02); RF ($trees$ = \{100, 200\}).

Performance on patch classification are evaluated by comparing classification results with ground-truth masks. Two precision-recall curves which return best scores of F-measure, are presented in Figure 3 for each different combination of the features. All best curves are obtained with blocks of 25 and 35 pixels, same cell dimension, and 100 trees. Two best HOG curves are obtained using 12 bins. Best LTP results ($thr$ = \{0.02\}) perform better than HOG configurations. Best results obtained by combining HOG+LTP ($bins$ = \{12\}, $thr$ = \{0.02\}) show no substantial increase in performance with respect to LTP best curves. Results in Figure 3 also account for improvements achieved by a morphological opening ($3 \times 3$ cross kernel) which removes isolated non-hair patches, while a closing fills small holes found in the hair region.

4. HAIR SEGMENTATION

The second processing step aims at segmenting the hair region at pixel level. Exploiting the coarse output of the previous step, which provides classified images at patch level, a graph-based segmentation method is applied to the candidate hair regions to refine results by incorporating hair color information. The advantage of this coarse-to-fine strategy relies in the fact that patch classification removes most of the background areas; then the segmentation step is performed on patch based candidate regions avoiding interference by unknown zones far from hair.

The proposed processing chain is described in Figure 4. The input image $I$ is processed by the Felzenszwalb and Huttenlocher algorithm [19], an efficient graph-based segmentation method. This technique is ruled by three different parameters: the smooth value $\sigma$, a pixel difference threshold $k$, and the minimum pixel number in a segment $m$. Without any a-priori information about the target to segment, this algorithm has a relevant problem of parameter selection that reduces its generality: the resulting segmented image could be in fact over-segmented or under-segmented, depending on the selected parameter set. Turning the algorithm weakness into strength, we exploit different combinations of parameters $P = \{\sigma, k, m, s\}$, where $s$ is a scale factor to downscale the image before segmentation, to obtain multiple segmentations of the input image, as shown in Figure 4-a.

Considering each single segment of each segmented image obtained with parameter $P$ as a binary image itself $I_{segm}^{P}$ (as in Figure 4-b), we match all segment images $I_{segm}^{P}$ with the patch based hair classified image $I_{patch}$, i.e. the output of the first processing step, in order to estimate the best parameter set $\hat{P} = \{\hat{\sigma}, \hat{k}, \hat{m}, \hat{s}\}$ so as:

$$\hat{P} = \arg \max_{P_i} (\text{accuracy}(I_{patch}, I_{segm}^{P}))$$

i.e. this maximisation step (in Figure 4-c) selects, from all possible segments obtained from multiple segmentations, the one that best covers, in terms of accuracy, the patch based hair candidate.

![Fig. 3. Hair classification performance on image patches.](image)

![Fig. 4. System workflow for pixel segmentation.](image)
5. EXPERIMENTAL RESULTS

Experiments evaluate segmentation results from best six patch based classifications shown in Figure 3. Multiple segmentations on original images are obtained by all combinations of the following parameters: $\sigma = \{0.3, 0.5, 0.7\}$, $k = \{300, 500, 700\}$, $m = \{500, 1000, 1500, 2000\}$, $s = \{1, 0.8, 0.6\}$. As in Section 3, tests are conducted on 140 test images randomly sampled from the database, and averaging on ten iterations. Performance assess the consistency of hair segmentation and the ground-truth masks in terms of Precision, Recall, F-measure and Accuracy, as reported in Table 2.

<table>
<thead>
<tr>
<th>Method</th>
<th>Prec. (%)</th>
<th>Rec. (%)</th>
<th>$F_1$ (%)</th>
<th>Acc. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{HOG}^{25}$</td>
<td>74.2</td>
<td>77.4</td>
<td>75.8</td>
<td>84.2</td>
</tr>
<tr>
<td>$\text{HOG}^{35}$</td>
<td>74.1</td>
<td>76.8</td>
<td>75.4</td>
<td>84.0</td>
</tr>
<tr>
<td>$\text{LTP}^{25}$</td>
<td>80.1</td>
<td>74.7</td>
<td>77.3</td>
<td>86.0</td>
</tr>
<tr>
<td>$\text{LTP}^{35}$</td>
<td>80.8</td>
<td>74.4</td>
<td>77.5</td>
<td>86.2</td>
</tr>
<tr>
<td>$\text{HOG} + \text{LTP}^{25}$</td>
<td>82.1</td>
<td>70.2</td>
<td>75.7</td>
<td>86.2</td>
</tr>
<tr>
<td>$\text{HOG} + \text{LTP}^{35}$</td>
<td>80.2</td>
<td>73.9</td>
<td>76.9</td>
<td>85.6</td>
</tr>
</tbody>
</table>

Again $\text{LTP}^{35}$ achieves the best performance, while the concatenation of both texture features show no substantial increase with respect to single LTP cases. As a support to the choice of having multiple segmentations to select the best matching segment, we report the marginal distribution of best parameters for $\text{LTP}^{35}$ (in Figure 5), showing that, due to the database variety, a high variation of parameters is needed. Measuring segmentation performance obtained using the ground-truth masks for best segment matching (instead of the patch based classified images) we obtain 92.3% of Accuracy and 87.8% of $F_1$, implying that segmentation errors are mainly due to previous misclassification of hair patches. Some examples of input images (with ground-truth), patch based classifications, and final outputs are shown in Figure 6.

Since a direct comparison with other methods is not possible due to the unavailability of databases shared by other research groups, we report performance declared in [12] for indirect comparison. In this work authors compare more methods on a conventional Near-frontal Head-shoulder Database of 1000 images (NDH1000), and a more challenging Multipose Head-shoulder Database of 3820 images (MHD3820). By inspecting images in [12] we can estimate that the challenge level provided by MHD3820, at least on the 920 non-frontal views included, is comparable with images in Figaro. Considering that [12] to obtain the best score on MHD3820 ($F_1 = 77.3\%$, as reported in Table 2) exploits a-priori information on hair location, and that only a portion of the images are in non-frontal views, we expect our algorithm, which scores $F_1 = 77.5\%$ on Figaro, to be at least comparable with reported state-of-the-art, if not superior.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>NHD1000</th>
<th>MHD3820</th>
<th>FIGARO840</th>
</tr>
</thead>
<tbody>
<tr>
<td>[8]</td>
<td>65.6</td>
<td>64.0</td>
<td>-</td>
</tr>
<tr>
<td>[9]</td>
<td>81.3</td>
<td>66.9</td>
<td>-</td>
</tr>
<tr>
<td>[12]</td>
<td>85.0</td>
<td>77.3</td>
<td>-</td>
</tr>
<tr>
<td>Our method</td>
<td>n.a.</td>
<td>n.a.</td>
<td>77.5</td>
</tr>
</tbody>
</table>

6. CONCLUSION

In this paper we presented a method for hair detection and segmentation. Detection is first achieved by a patch based approach which exploits texture features for binary classification. Hair segmentation is then accomplished by refining results obtained at patch level by multiple segmentations obtained by a graph-based algorithm. As a first advance, hair segmentation is performed on unconstrained view images without using any a-priori information about head and hair location. Second, we make available Figaro, a database of 840 images with human hair taken from unconstrained views, for allowing further comparison among diverse hair segmentation algorithms. Results are comparable to other recent methods exploiting a-priori information on constrained databases.
7. REFERENCES


