

Gender Recognition from Body

Liangliang Cao, Mert Dikmen, Yun Fu and Thomas S. Huang
Beckman Institute and Coordinated Science Lab
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
{cao4,mdikmen,yunfu2,huang}@ifp.uiuc.edu

ABSTRACT

This paper studies the problem of recognizing gender from full body images. This problem has not been addressed before, partly because of the variant nature of human bodies and clothing that can bring tough difficulties. However, gender recognition has high application potentials, e.g., security surveillance and customer statistics collection in restaurants, supermarkets, and even building entrances. In this paper, we build a system of recognizing gender from full body images, taken from frontal or back views. Our contributions are three-fold. First, to handle the variety of human body characteristics, we represent each image by a collection of patch features, which model different body parts and provide a set of clues for gender recognition. To combine the clues, we build an ensemble learning algorithm from those body parts to recognize gender from fixed view body images (frontal or back). Second, we relax the fixed view constraint and show the possibility to train a flexible classifier for mixed view images with the almost same accuracy as the fixed view case. At last, our approach is shown to be robust to small alignment errors, which is a preferred property in many applications.

Categories and Subject Descriptors

I.4.7 [Image Processing and Computer Vision]: Feature Measurement; I.5.4 [Pattern Recognition]: Applications

General Terms

Algorithms, Experimentation

Keywords

Gender Recognition, Human Body

1. INTRODUCTION

A recently emergent demand for multimedia system design is to recognize the gender of the subjects from body images



Figure 1: Examples of human body images. The top row shows female images and the bottom row shows male images.

in frontal or back views. Such a system can find applications in wide range of scenarios, e.g., security surveillance in building entrances and parking lots, or customer statistics collection in supermarkets, cafes, McDonald's, etc.

Gender recognition has been an attractive problem in image understanding. SEXNET [1] is the first attempt to recognize gender from faces. Later, this problem has attracted many other researchers [2, 3, 4, 5]. However, all these previous works are constrained to analyzing human faces. In many scenarios, gender recognition from face images is not a preferred approach, especially when it is too intrusive to take face pictures; when resolutions of the face images are too low to be tractable; and when the faces are occluded by glasses, masks and facial hair. Some researchers noticed this problem and proposed to recognize gender by human gait analysis [6]. However, capturing and describing human gait in general situations is still an open problem. With these concerns, we believe that it is crucial to develop an automatic recognition system for inferring gender from human body appearances.

Figure 1 shows a few examples of human body images. We can see that human body based gender recognition is harder than faces in several aspects. Both females and males can dress the same color or not. People of the same gender may choose clothes and hairs of different styles. The background of body images is more cluttered than the cropped face images. Due to these difficulties, the problem of gender from body has not been touched before. In the rest of this paper, we will design a reliable approach to recognize gender from

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'08, October 26–31, 2008, Vancouver, British Columbia, Canada.
Copyright 2008 ACM 978-1-60558-303-7/08/10 ...\$5.00.

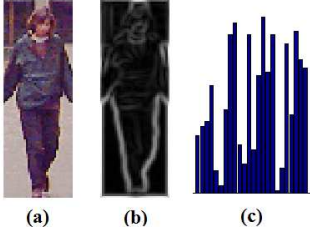


Figure 2: Representing human body by different features. (a) raw feature, (b) edge map, (c) histogram of oriented gradients.

a single body image that can overcome these difficulties.

2. APPROACHES

To recognize gender from human body, we are facing two questions. First, what characteristics are crucial for gender recognition? Second, how to make decisions based on these characteristics? To answer these questions, we will discuss several feature representation methods in section 2.1 together with two ensemble learning algorithms in section 2.2. Based on these discussions, we will present our part-based gender recognition approach in section 2.3.

2.1 Representations

Raw pixel is the most straightforward representation for images. However, it may be not a good choice for recognition task considering the variations of clothes colors. To remove the affects of visual varieties, we also consider to employ edge maps instead of raw pixels. In our approach, an edge map is obtained by Canny’s method [7], which is broadly acknowledged as a effective solution for edge detection.

Although the edge map provides rich information for shape recognition, it is not robust enough. A small change of illumination and position will result in significant differences in the edge map. To obtain a better representation, we employ the Histogram of Oriented Gradients (HOG) feature [8], which represents edges with a magnitude-weighted histogram, grouped according to edge directions. As described in [8], each body figure is divided into 3×3 blocks, and each block is represented by a HOG feature, as a vector of length 8 describing the gradient in 8 orientations. Note that HOG is originally designed for pedestrian detection [8], but we first adopt it for gender recognition due to its ability of characterizing the shape of body. Figure 2 shows examples of the three kinds of features.

2.2 Classifiers

After images are represented with different features, different classification methods can be applied. Here we introduce two powerful classification methods—Adaboost and Random Forests (RF).

Adaboost is an ensemble learning algorithm proposed by Freund and Schapire [9], which iteratively constructs optimal weak classifiers from reweighing training samples in each round. The output of Adaboost is a weighted additive model combining all these weak classifiers into a strong one. The weak classifiers can be of many forms, however, we choose the most discriminative decision stumps as in [10]. The details of training Adaboost are shown in Algorithm 1.

Algorithm 1 : Adaboost Algorithm.

Input: n_+ male images and n_- female images. The image labels y_i is 1 for male, and -1 for female, and the image feature $\mathbf{x}_i = (x_i(1), x_i(2), \dots, x_i(d))^T$, where i is the image index with $1 \leq i \leq n_+ + n_-$ and d is the feature dimension.

- 1: Initialize the weights of training samples $D_1(i) = 0.5/|n_+|$ if $y_i = 1$, or $D_1(i) = 0.5/|n_-|$, if $y_i = -1$.
- 2: **for** $t = 1, \dots, T$ **do**
- 3: Based on D_t , select the weak classifier $h_t(\mathbf{x}_i) = h_t(x_i(k_t))$, where k_t is the optimal feature position.
- 4: Compute the training error ε_t of weak classifier h_t .
- 5: Compute $\alpha_t = 0.5 \log \frac{1-\varepsilon_t}{\varepsilon_t}$.
- 6: Update $D_{t+1}(i) = \frac{D_t(i)}{Z_t} \exp(-\alpha_t y_i h_t(\mathbf{x}_i))$, where Z_t is a normalized factor.
- 7: **end for**

Output: The final classifier is $H(\mathbf{x}) = \text{sign}(\sum_{t=1}^T \alpha_t h_t(\mathbf{x}))$.

Algorithm 2 : Random Forest Algorithm.

- 1: Draw bootstrap samples from the original data.
- 2: **for each** bootstrap samples **do**
- 3: Randomly draw a subset of all the predictors.
- 4: Using the selected subset, grow an unpruned classification tree.
- 5: **end for**

Output: The trained random forest is the majority voting of the decision trees.

In [11], Breiman proposed the random forests, which train successive decision trees and combine them for prediction. Each decision tree is independently trained using a bootstrap sample of the data set. In the training process, each node is split using the best among a subset of predictors randomly chosen at that node. Random forests try to maximize the independence of features, while keeping individual classifiers as accurate as possible. Compared with many other classifiers including discriminant analysis, SVMs and neural networks, random forests perform very well or even better [11]. Random forests are also preferred since they are robust to overfitting, and are also very fast to train and test. The procedure of training a random forest is summarized in Algorithm 2.

2.3 Part-based Gender Recognition

Based on the discussion of section 2.1 and 2.2, we next introduce our Part-Based Gender Recognition (PBGR) algorithm which addresses the specialty of human body images. The intuition is summarized as follows. Although it is very hard to describe the determinant gender characteristics, it is possible to judge some parts of human body whether they are more likely to be female or male. For example, not all the females have long hair, but one person with long hair is more likely to be female than male. Similarly, a skirt is probably worn by a female, while a suit is more likely to be on a male.

To model different parts of human body, we use grid sampling to partition the image into patches. One patch corresponds to some part of human body, such as shoes and hair, which can provide useful cues for recognizing gender. To obtain a robust representation, we extract a HOG feature vector for each part. Our feature is different from those in

Algorithm 3 : Part-based Gender Recognition.

```

1: For each image  $i$ , extract part-based features  $\mathbf{x}_i$ , where
    $\mathbf{x}_i = \{x_i(p)\}$ , with index  $1 \leq p \leq P$ .
2: for  $t = 1, \dots, T$  do
3:   Select a body part  $p_t$ .
4:   Get weak classifier  $h_t(\mathbf{x}) = h_t(x(p_t))$ .
5:   Compute the voting weight  $\alpha_t$  as Adaboost.
6: end for
Output: The final classifier is  $H(I) = \text{sign}(\sum_{t=1}^T \alpha_t h_t)$ .

```

[8] since we use a finer grid to describe human body. Our grid is 6×19 in the experiment, while [8] has 2×2 or 3×3 . Moreover, our patches are overlapping with each other while the origin HOG features do not overlap.

The differences between our features and those in [8] are coming from our belief that the gender is more related to parts of the body instead of the whole. Hence we consider each part individually instead of concatenating all features into a long vector. We will then discuss how to construct a reliable classifier based on the part-based representations.

In our method, every body part provides a useful cue for the gender recognition task. To combine these cues, we adopt the ensemble learning algorithm which constructs a succession of weak classifiers. In each round, our algorithm first selects the most discriminative patch position, and then computes the optimal classifier only using the patch features of the selected position. This in fact is to build a combined classifier with the weighted voting of the recognition results based on different body parts. Although individual cues from a single body part may be doubtful, e.g., a male subject may also have long hairs, the final classifier combining multiple cues will be strong and robust. The proposed part-based gender recognition algorithm is summarized in Algorithm 3.

Our PBGR algorithm can be viewed as a variant of Adaboost algorithm. Gaining insights from random forest algorithm, the weak classifier in our algorithm is based on a subset of features (body parts), instead of the whole images. We will show in the next section that our method outperforms the traditional Adaboost and random forests.

3. EXPERIMENTAL RESULTS

3.1 Dataset

Currently, there is no existing database of human body pictures with gender labels. To train a classifier, we manually labeled publicly available MIT pedestrian database [12] into male and female classes. Some examples from the MIT pedestrian database are shown in Figure 1. All pedestrian images are 128 pixels high and 64 pixels wide. The people are centered and aligned so that their height is approximately 100 pixels and their midpoint lies in the middle column of the image. From the original database we labeled 600 images as male and 288 images as female. The other 31 images are left out because the genders of the subjects were unidentifiable even by human experts, due to heavy clothing, dense artifacts or posture. The final dataset is composed of 53% back view images and 47% frontal view images.

3.2 Comparison of Recognition Results

We compare the performance on both frontal and back view images using different features—raw feature, edge map

Table 1: Gender recognition accuracy for frontal or back view images. Each method is measured by the average accuracy and the standard deviation.

Methods	Frontal View	Back View
Ada + Raw	$71.7 \pm 3.6\%$	$61.5 \pm 4.3\%$
RF + Raw	$72.8 \pm 3.7\%$	$60.2 \pm 4.4\%$
Ada + Edge Map	$71.1 \pm 3.6\%$	$70.1 \pm 3.7\%$
RF + Edge Map	$73.1 \pm 3.7\%$	$63.9 \pm 4.4\%$
Ada + HOG	$70.9 \pm 4.4\%$	$63.0 \pm 4.1\%$
RF + HOG	$73.1 \pm 3.5\%$	$65.8 \pm 4.2\%$
Ada + Part	$71.9 \pm 3.4\%$	$71.2 \pm 3.3\%$
RF + Part	$73.2 \pm 2.4\%$	$65.4 \pm 0.0\%$
PBGR	$76.0 \pm 1.2\%$	$74.6 \pm 3.4\%$

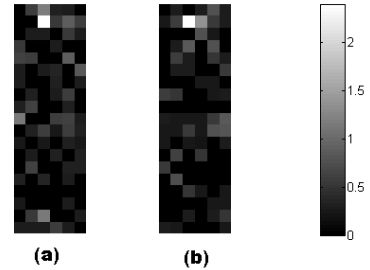


Figure 3: Voting weights for different patches. (a) weights from the frontal view classifier. (b) weights from the back view classifier.

using Canny’s method [7], HOG in [8] and our part-based representation—employed by different classifiers—Adaboost using decision stumps (Ada), Random Forest (RF) and our PBGR. The recognition accuracy is measured using a 5-fold cross validation.

Table 1 shows the recognition accuracy using these methods. We can see that HOG is a better feature than edge map, while the raw feature is the worst. Random forest performs better in recognizing gender from front view images, while Adaboost does better for back view images. However, of all these combinations, our PBGR outperforms all the other algorithms and obtain a 76.0% accuracy for frontal view images and 74.6% for back view images. This validates the effectiveness of our approach.

To understand the results of our PBGR classifier better, we go through the ensemble classifier and sum up the voting weights for weak classifiers according to different patch positions. The larger weights one patch has, the more important role it plays in gender recognition. Figure 3 shows the weights for different patches. We can see that gender recognition relies more on the patches in the positions of head and body silhouettes.

3.3 Gender Recognition from Mixed View

By observing Figure 3, we can see that the larger voting weights for the patches fall into the position of head and body, for both frontal and back view classifier. This observation suggests that the classifiers of both views share some similarity. A natural question comes: can we recognize the gender without the effort of distinguishing the frontal view and back view? If the answer is “Yes”, we will have a more flexible gender recognition classifier.

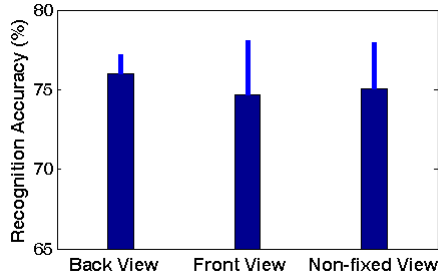


Figure 4: Comparing the recognition accuracy for the fixed view classifiers and flexible classifier. The thick bars stand for the average accuracy for each experiment, and the thin bars are the standard deviation of the 5-fold cross validation.

We train our PBGR algorithm on the non-fixed view images. In the training stage, images are taken randomly from both frontal and back views. In the testing stage, we still adopt a 5-fold cross validation, without distinguishing frontal or back views. The overall accuracy is $75.0 \pm 2.9\%$. As compared in Figure 4, non-fixed view accuracy is very similar to the fixed view accuracy. This result is amazing in some degree: we can build a more flexible classifier without degrading the accuracy. Figure 5 shows some examples of flexible gender recognition results.

3.4 Robustness Analysis

Real surveillance systems prefer a gender recognition algorithm robust to misalignment. Since histogram-based features and part-based representation will not change much with small local displacement, we expect our PBGR algorithm is robust to misalignment. To verify this, we train our model using original data while testing the model with misaligned images. Here a misaligned image is shifted horizontally with a distance so that the center of human body is no longer the center of the image. Figure 6 shows the recognition accuracy versus different misaligned distances. We can see that the accuracy remains almost unchanged when misalignment is 5% of the body width and the accuracy only drops to 69% subject to 15% misalignment. These results suggest that our approach is robust to tolerate small misalignment errors.

4. CONCLUSION

To the best of our knowledge, this paper is the first attempt to investigate gender recognition from still human



Figure 5: Examples of gender recognition from body. Red and blue rectangles indicate the subjects are recognized as female and male, respectively. Our approach fails to recognize the last example probably because of the unusual pose of the subject.

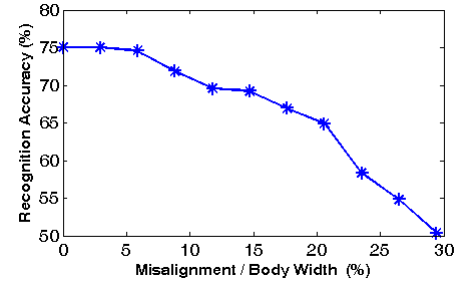


Figure 6: Accuracy of gender recognition from misaligned images.

body images. By integrating the part-based representation and ensemble learning algorithm, our PBGR algorithm can recognize the gender from either a single frontal or back view image with the accuracy of 75.0% and also behave robustly subject to small misalignment. Our future work is to extend the algorithm to side view images. In addition, we will consider the problem of gender recognition from video, which is supposed to analyze both appearances and motions.

Acknowledgement

This research was funded in part by the U.S. Government VACE program and the NSF Grant CCF 04-26627. The views and conclusions are those of the authors, not of the US Government or its Agencies.

5. REFERENCES

- [1] D. T. Lawrence B. A. Golomb and T. J. Sejnowski, "Sexnet: A neural network identifies sex from human faces," *Neural Information Processing Systems*, pp. 572–577, 1991.
- [2] R. Brunelli and T. Poggio, "Hyperbf networks for gender classification," in *DARPA Image Understanding Workshop*, 1992, pp. 311–314.
- [3] B. Moghaddam and M.-H. Yang, "Gender classification with support vector machines," in *IEEE Conf. on Automatic Face and Gesture Recognition*, 2000, pp. 306–311.
- [4] H. Wechsler S. Gutta and P. J. Phillips, "Gender and ethnic classification," in *IEEE Conf. on Automatic Face and Gesture Recognition*, 1998, pp. 194–199.
- [5] X. Xu and T. S. Huang, "Soda-boosting and its application to gender recognition," *IEEE Workshop on Analysis and Modeling of Faces and Gestures*, 2007.
- [6] X. Li, et al., "Gait components and their applications to gender recognition," *IEEE Trans. Systems, Man, and Cybernetics, Part C*, vol. 38, no. 2, 2008.
- [7] J. Canny, "A computational approach to edge detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.
- [9] Y. Freund and R. Schapire, "Experiments with a new boosting algorithm," *Int'l Conf. Machine Learning*, 1996.
- [10] P. Viola and M. Jones, "Robust real-time face detection," *Int'l Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [12] M. Oren, et al., "Pedestrian detection using wavelet templates," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 1997, pp. 193–199.