ON THE IMPORTANCE OF THE GRID SIZE FOR GENDER RECOGNITION USING FULL BODY STATIC IMAGES

Carlos Serra-Toro[1,3], V. Javier Traver[1,3], Raúl Montoliu[2,3] and José M. Sotoca[1,3]
[1]*DLSI,* [2]*DICC &* [3]*iNIT, Universitat Jaume I, 12071 Castellón, Spain*

Keywords: Gender recognition, Soft biometrics, Machine learning.

Abstract: In this paper we present an study on the importance of the grid configuration in gender recognition from whole body static images. By using a simple classifier (AdaBoost) and the well-known Histogram of Oriented Gradients features we test several grid configurations. Compared with previous approaches, which use more complicated classifiers or feature extractors, our approach outperforms them in the case of the frontal view recognition and almost equals them in the case of the mixed view (i.e. frontal and back views combined without distinction).

1 INTRODUCTION

The characterization of people according to some criterion (e.g. age, ethnicity, or gender) in digital images and videos is relevant for many applications of scientific and social interest. Some recent research has been done concerning the classification of people according to their gender. Most of them use a facial approach (Moghaddam and Yang, 2002) while others try to classify the gender of people according to their gait (Yu et al., 2009). However, there is not much work done concerning gender classification using whole body static images of standing people and the first contribution is as recent as (Cao et al., 2008).

In this paper we study the gender recognition of a single, standing person using just one whole body static image. This is a complex problem since gender recognition is a difficult task even for human beings since, although there are a number of heuristics that can partially guide the design of an algorithm, there are many exceptions that make them unreliable in general. Furthermore, detecting those conditions is also a problem in itself.

The tendency in the machine learning community seems to be to give more importance to find complicated or newer descriptors, or combination of several existing descriptors or classifiers, to achieve more accuracy. This tendency sometimes leads to overlook other simpler aspects of the problems that can impact on the global accuracy. In this paper we show that, in the case of gender recognition from whole body static images, choosing an optimal grid for classification can be as important as, and sometimes more important than, choosing a complicated feature extractor or classifier to perform the recognition. We present an study on the importance of the grid configuration for gender classification, achieving results comparable to those obtained by previous works (Section 2).

2 STATE OF THE ART

To the best of our knowledge, there are currently as few as four published papers addressing the problem of gender recognition from whole body static images. This section is a review of these works.

The first documented approach to gender recognition from static images was (Cao et al., 2008). They manually labeled (see Section 4.1) the CBCL pedestrian database (Oren et al., 1997), releasing the first publicly available dataset for the evaluation of gender classification. They created a classifier inspired by AdaBoost (Freund and Schapire, 1995), based on a part-based representation of the body, named Part-based Gender Recognition (PBGR), in which every part provided a clue of the gender of the person in the image. In each round of the algorithm, they first selected the most optimal patch of the image and then trained a learner using only the Histogram of Oriented Gradients (HOG) features (Dalal and Triggs, 2005) corresponding to that part of the image. They achieved a recognition rate of 75.0% for the mixed view, and 76.0% and 74.6% when considering just the

Serra-Toro C., Javer Traver V., Montoliu R. and M. Sotoca J..
ON THE IMPORTANCE OF THE GRID SIZE FOR GENDER RECOGNITION USING FULL BODY STATIC IMAGES.
DOI: 10.5220/0003323803340339
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2011), pages 334-339
ISBN: 978-989-8425-47-8

frontal or the back view images, respectively.

The second contribution was (Collins et al., 2009), who investigated a number of feature extractors and noticed that the best results in gender recognition were achieved using a combination of both a shape (based on an edge map) and a colour (based on hue histograms (van de Sande et al., 2008)) extractors combined using a linear kernel support vector machine (SVM) classifier. They only focused on frontal view images and randomly balanced the CBCLdataset so that there were 123 images of each gender. Also, they cropped each image so that its size was approximately the bounding box of the person represented in it. They reported an accuracy of 76.0% (the same rate as (Cao et al., 2008) for the frontal views) with a good balance between the male and the female accuracy. (Collins et al., 2009) repeated the experiment with the VIPeR dataset (Gray et al., 2007) (a larger dataset with almost 300 images for each gender) and achieved 80.6% of accuracy in frontal view images.

In a newer contribution, (Collins et al., 2010) combined the VIPeR and CBCL images to create a database of 413 images of each gender, all in frontal view. They obtained the "eigenbody" images, computed using the principal component analysis (PCA) (Jolliffe, 2005) over the raw images data and over the edge map of the images. They realized that the gender from whole body image discrimination seems to be encoded by a combination of several PCA components. They chose the top components from both the raw and edgemap data and combined them using a SVM, resulting in a recognition accuracy of 66%.

Very recently, (Guo et al., 2010) reported an accuracy of 80.6% using only the CBCL dataset with (Cao et al., 2008) manual labelling, without balancing and without cropping the images and considering both frontal and back view combined (mixed view). They represented each image with biologically-inspired features (BIF) (Serre et al., 2007) combined with several manifold learning techniques. They designed a classification framework where the type of view was considered. Their accuracy rate was 80.6%, which is, to the best of our knowledge, currently the best gender recognition rate published with the CBCL dataset.

In this paper we intend to explore the impact of the grid configuration in the gender recognition task, an aspect that has not been yet taken into consideration, by using a simple classifier.

3 APPROACH

We consider images showing a whole body picture of a single, still standing person in frontal or back view. The persons shown in all the pictures are approximately aligned and scaled so that different persons in different images all take up a similar space.

Each gray-scale image I is described using a feature vector, v_I. Each image is divided into smaller rectangles, called *cells*, whose size are defined by a $r \times c$ grid applied over the picture, with r being the number of windows across the height of the image, and c being the number of windows across its width. There is an overlap between adjacent cells of 50%, both vertically and horizontally. For each cell, a Rectangular HOG (R-HOG) (Dalal and Triggs, 2005) feature vector is obtained, so that the resulting feature vector v_I for the image I is the concatenation of all the feature vectors obtained for each of its cells: $v_I = v_{I_1} v_{I_2} \dots v_{I_{rc}}$, with v_{I_γ} being the feature vector corresponding to cell γ of the image I.

In order to classify each image as representing a male or a female we use AdaBoost (Freund and Schapire, 1995). We use decision stumps as the weak learner in the same way as (Cao et al., 2008) do, and use their same variant of AdaBoost (see Algorithm 1 of (Cao et al., 2008) for the details).

4 EXPERIMENTATION

4.1 Image Dataset

We use the CBCL pedestrian image database[1] (Oren et al., 1997), not designed initially to gender recognition, but used by other authors (Section 2) for this purpose. Images are all 64×128 pixel size, showing one pedestrian standing in frontal or back view, horizontally and vertically aligned so that their height is about 80 pixels from their shoulders to their feet.

We use (Cao et al., 2008)'s publicly available manual labelling of the CBCL database according to pedestrians' gender. This labelling consists of 600 men and 288 women. The wiews are also classified: 51% (frontal) and 49% (back) for male images, and 39% (frontal) and 61% (back) for female images.

4.2 Implementation Details

The experiments were executed using an implementation of a version of the R-HOG (Dalal and Triggs, 2005). In our implementation we consider only gray-level images and therefore the information of the

[1] http://cbcl.mit.edu/cbcl/software-datasets/PedestrianData.html

colour is lost, since we believe that the gender information is primarily codified in the shape of the figure.

Since our purpose was the study of the importance of the grid configuration, we have simplified the block schema used by (Dalal and Triggs, 2005) and we do not group adjacent cells into a block, and therefore no normalization is done between the cells within a block. Since (Dalal and Triggs, 2005) report best results with respect to pedestrian detection when block normalization is done, it is possible that gender recognition benefit too from this schema. Since our purpose was the study of the grid configuration and not the clustering between its cells, we left the study of clustering them into blocks as future work.

4.3 Experiments with the Grid Size

In order to verify our hipothesis (i.e. choosing an optimal grid is as important as using a complex algorithm with respect to gender recognition from static images) we have tested several grid configurations over the image. The idea is to find an optimal grid for gender recognition and then achieve an accuracy similar to those proposed in the literature (Section 2) using a simpler algorithm (AdaBoost, in our case).

As stated in Section 3, our grid consists of an uniformly sampled cartesian $r \times c$ grid with 50% of cell overlap. We decided to test a wide range of grid configurations, resulting in several cell sizes from considerable big (about 21×21 pixels) to quite small (about 3×3 pixels), and therefore we tested $r \in S_r$, $S_r = \{6, 9, \ldots, 42\}$ and $c \in S_c$, $S_c = \{3, 6, \ldots, 24\}$ values. Therefore, $|S_r| \times |S_c| = 13 \times 8 = 104$ possible grids were explored. In all cases we train and classify using the AdaBoost described in Section 3.

The results are averaged using a 5-fold cross validation, in the same way as the previous contributions do (Section 2). Our results are shown in Tables 1, 2 and 3 for frontal, back and mixed view, respectively. We highlight the optimal grid (the one with best accuracy) for each view. According to our experimentation, the optimal grid configurations for each view are: 21×12 for the frontal view, 36×21 for the back view and 15×15 for the mixed view.

It is interesting to notice that the optimal recognition grid is denser than the one used by (Dalal and Triggs, 2005) in their pedestrian detector since they propose a grid with cells of 6×6 pixels resulting in a grid of 210 cells for a 128×64 image, while our optimal grid results to be of 252 cells for the frontal view, 756 cells for the back view and 225 for the mixed view. We think this suggests that gender recognition depends more on certain parts of the silhouette rather than in the silhouette of the whole body, since finer grids allows the classifier to be more focused on particular aspects of the shape than grids with less divisions are able to. This finding is in agreement with those recently reported by (Collins et al., 2010).

Figures 1 and 2 summarizes the results of the Tables 1, 2 and 3, showing the mean accuracy for each combination of the value of the number of windows across the height (r) or across the width (c) of the image, for each view. As it can be seen, in general the results improve as the grid makes denser, up to a certain point at which the accuracy degrades gradually.

It is interesting to notice that the top for the frontal view is more to the left than the top for the back view. This indicates that denser grids are needed to recognize gender from back view images, probably because this view is more difficult, even for humans beings. The highest accuracy for the mixed view is more on the left but, contrary to what happens with the other views, high values for r or c result in a high variance of the results, and thus the recognition behaviour is more unstable with denser grids. This is probably the reason why the optimal recognition grid for the mixed view, 15×15, is the one with less density of cells.

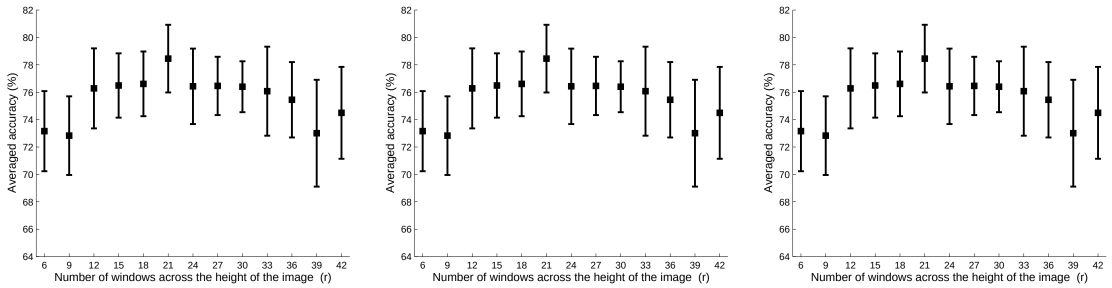
4.4 Study of the Overfitting of AdaBoost

The results shown in Tables 1, 2 and 3 are obtained using 400 iterations for AdaBoost. There is some controversy about the convenience of stopping early in AdaBoost to not overfit the data, as (Zhang and Yu, 2005) claims, or to perform a large number iterations so that the overfitting reduces, as (Mease and Wyner, 2008) experimentally shows. Results probably depend on the nature of each problem, so we have studied the evolution of the accuracy as the number of iterations of AdaBoost increases from 100 iterations to 1500, in steps of 100, for each view.

The results obtained, using the optimal grid found in Section 4.3 for each view, are shown in Figure 3. As it can be seen, the increase of the number of weak learners up to 300 and 400 in the case of frontal and mixed view, respectively, and up to 600 in the case of the back view, increases the accuracy, and then the recognition rate becomes more or less stable in the three cases. We think in our case AdaBoost is not overfitting the data because, if that were the case, then this would result in an increase of the accuracy with the increasing of the number of the iterations.

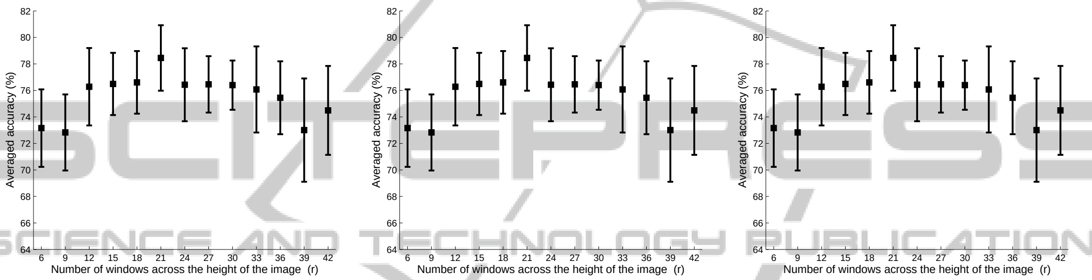
4.5 Comparison with other proposed Methods

We compare our results with those reported by the other existing published approaches (Section 2) in



(a) Frontal view (b) Back view (c) Mixed view

Figure 1: Mean accuracies (%) and standard deviations for each value of the number of windows across the height of the image (r) for each view.



(a) Frontal view (b) Back view (c) Mixed view

Figure 2: Mean accuracies (%) and standard deviations for each value of the number of windows across the width of the image (c) for each view.

Table 1: Accuracies (%) obtained using only the frontal view images using several grids of sizes $r \times c$.

		Number of windows across the width (c)							
		3	6	9	12	15	18	21	24
No. of windows across height (r)	6	68.6 ± 3.4	69.8 ± 3.0	72.9 ± 4.8	75.0 ± 4.8	75.7 ± 7.3	76.4 ± 4.2	75.2 ± 4.4	71.7 ± 3.7
	9	66.7 ± 2.4	71.9 ± 2.5	74.0 ± 3.9	74.0 ± 4.2	73.8 ± 5.5	75.0 ± 2.2	75.7 ± 2.5	71.4 ± 1.9
	12	70.5 ± 1.8	75.5 ± 3.6	80.0 ± 5.2	76.7 ± 2.5	78.8 ± 1.0	78.1 ± 3.5	75.5 ± 2.7	75.2 ± 3.4
	15	71.9 ± 5.4	75.2 ± 6.0	77.4 ± 4.5	79.0 ± 2.5	79.0 ± 2.5	76.9 ± 3.6	77.1 ± 5.2	75.2 ± 5.9
	18	72.6 ± 5.0	78.6 ± 4.9	79.0 ± 2.7	79.0 ± 3.1	77.6 ± 4.1	74.5 ± 3.9	76.2 ± 3.5	75.2 ± 4.6
	21	74.5 ± 4.6	78.1 ± 3.9	78.3 ± 3.9	**81.9 ± 5.0**	81.0 ± 5.3	80.0 ± 4.9	77.9 ± 3.4	76.0 ± 4.3
	24	71.9 ± 6.0	76.9 ± 5.0	79.3 ± 1.8	78.1 ± 3.9	79.0 ± 4.8	75.2 ± 3.3	77.9 ± 2.7	73.1 ± 4.3
	27	72.6 ± 3.0	76.7 ± 2.9	76.4 ± 3.7	77.1 ± 3.3	79.3 ± 3.7	76.9 ± 5.4	78.3 ± 1.6	74.3 ± 5.0
	30	75.2 ± 2.8	76.9 ± 5.2	76.0 ± 2.7	78.1 ± 1.8	78.3 ± 2.4	77.6 ± 5.4	76.4 ± 3.3	72.6 ± 3.6
	33	71.4 ± 6.2	75.7 ± 6.3	79.0 ± 2.6	80.5 ± 3.8	79.5 ± 1.8	74.0 ± 5.0	74.8 ± 3.4	73.6 ± 6.3
	36	72.9 ± 4.3	74.3 ± 4.5	75.7 ± 1.8	79.3 ± 3.0	79.3 ± 3.8	76.0 ± 7.2	74.5 ± 5.3	71.7 ± 6.0
	39	68.3 ± 3.4	75.5 ± 4.7	76.4 ± 1.6	77.1 ± 3.9	74.8 ± 4.7	70.0 ± 7.6	74.8 ± 5.7	67.1 ± 6.4
	42	73.1 ± 2.0	76.9 ± 3.0	77.4 ± 4.8	76.2 ± 2.7	76.2 ± 2.7	76.2 ± 3.0	72.6 ± 5.5	67.4 ± 4.3

Table 4. We left out of the comparison the work by (Collins et al., 2010) since they do not use the same image dataset as the other works (and this paper) use and therefore the results are not comparable.

As it can be seen, we achieve the highest published accuracy in the frontal view (+2.4%), nearly match the highest rate on the mixed view (−1.2%) and stay below the highest accuracy with the back view (−3.2%), always using a simpler classifier (AdaBoost, Section 3) and a reduced implementation of a simple feature extractor (R-HOG, Section 3).

5 CONCLUSIONS

We have shown that denser grids than those originally proposed for pedestrian detection by (Dalal and

Table 2: Accuracies (%) obtained using only the back view images using several grids of sizes $r \times c$.

		Number of windows across the width (c)							
		3	6	9	12	15	18	21	24
No. of windows across height (r)	6	64.5 ± 6.7	67.3 ± 2.7	67.3 ± 6.3	69.0 ± 2.5	74.2 ± 3.7	69.5 ± 2.4	70.9 ± 3.5	68.0 ± 2.9
	9	67.1 ± 3.4	69.9 ± 3.1	69.9 ± 4.7	75.4 ± 3.8	72.7 ± 3.8	70.3 ± 5.6	70.5 ± 3.1	71.2 ± 1.4
	12	73.9 ± 5.6	74.6 ± 3.2	74.2 ± 3.3	74.6 ± 1.9	77.1 ± 4.7	71.6 ± 1.5	75.6 ± 2.9	75.4 ± 3.7
	15	68.2 ± 4.4	70.7 ± 5.4	73.9 ± 3.1	73.3 ± 5.1	76.3 ± 2.4	77.6 ± 2.6	76.3 ± 4.4	76.3 ± 4.3
	18	72.2 ± 2.1	71.8 ± 2.3	72.9 ± 2.3	77.3 ± 4.2	75.7 ± 5.1	75.4 ± 2.1	80.6 ± 2.9	75.7 ± 3.4
	21	72.7 ± 2.1	73.1 ± 3.7	73.9 ± 1.1	77.6 ± 5.5	76.7 ± 6.9	75.2 ± 3.6	79.9 ± 1.7	78.4 ± 2.2
	24	66.2 ± 2.8	73.3 ± 2.5	76.1 ± 3.7	78.8 ± 2.9	76.3 ± 3.5	77.6 ± 6.3	78.4 ± 2.7	76.5 ± 3.9
	27	70.7 ± 4.4	73.1 ± 1.9	78.4 ± 2.2	79.3 ± 2.4	77.1 ± 2.1	78.4 ± 2.4	78.2 ± 3.6	73.9 ± 4.0
	30	70.7 ± 3.4	75.6 ± 4.4	75.9 ± 0.9	79.5 ± 3.2	74.6 ± 3.7	79.7 ± 3.4	79.9 ± 2.5	76.3 ± 2.8
	33	72.0 ± 2.2	74.3 ± 3.9	74.4 ± 2.0	76.9 ± 4.9	78.2 ± 3.2	79.3 ± 3.3	78.8 ± 1.6	72.5 ± 6.9
	36	72.4 ± 5.4	72.2 ± 2.3	73.9 ± 2.5	77.8 ± 2.2	75.4 ± 5.7	76.1 ± 2.9	$\mathbf{80.8 \pm 2.3}$	71.4 ± 8.8
	39	68.2 ± 3.5	69.0 ± 4.8	74.6 ± 4.4	76.3 ± 3.4	78.6 ± 2.0	75.6 ± 4.4	71.4 ± 6.1	69.4 ± 1.4
	42	68.2 ± 3.5	72.4 ± 7.0	73.7 ± 3.1	77.6 ± 3.5	75.4 ± 3.8	73.1 ± 4.9	72.8 ± 3.7	63.7 ± 7.1

Table 3: Accuracies (%) obtained using both the frontal and back view images (i.e. considering both views without distinction between them) using several grids of sizes $r \times c$.

		Number of windows across the width (c)							
		3	6	9	12	15	18	21	24
No. of windows across height (r)	6	63.1 ± 2.9	68.0 ± 1.5	70.7 ± 0.5	67.8 ± 2.9	71.5 ± 4.7	69.9 ± 1.8	69.9 ± 3.6	70.5 ± 2.7
	9	68.9 ± 4.0	73.3 ± 2.2	70.4 ± 2.9	72.4 ± 2.8	72.4 ± 1.1	72.6 ± 4.9	73.1 ± 3.4	73.9 ± 1.6
	12	69.6 ± 2.3	73.1 ± 3.1	75.0 ± 2.2	73.0 ± 1.6	76.1 ± 5.3	74.8 ± 2.7	75.9 ± 4.3	71.6 ± 3.3
	15	69.6 ± 1.8	73.7 ± 2.0	75.0 ± 2.5	76.2 ± 2.3	$\mathbf{79.4 \pm 1.4}$	73.5 ± 2.4	74.2 ± 2.2	78.3 ± 1.6
	18	71.2 ± 3.8	74.3 ± 3.2	74.7 ± 1.1	77.0 ± 4.0	75.9 ± 1.1	76.8 ± 2.8	77.8 ± 2.2	74.2 ± 1.6
	21	73.5 ± 3.0	76.9 ± 2.6	77.1 ± 1.9	75.9 ± 3.5	76.6 ± 2.3	77.0 ± 2.1	77.9 ± 3.1	76.1 ± 4.1
	24	70.1 ± 4.0	72.2 ± 1.2	75.1 ± 2.1	75.8 ± 2.1	75.6 ± 2.7	76.9 ± 4.0	77.8 ± 4.0	69.6 ± 3.9
	27	70.4 ± 2.3	73.4 ± 3.7	74.2 ± 3.5	78.0 ± 1.7	77.0 ± 4.4	75.1 ± 4.4	75.1 ± 2.5	70.9 ± 3.2
	30	73.3 ± 2.4	74.5 ± 2.5	74.9 ± 2.8	75.8 ± 2.7	76.0 ± 2.5	72.4 ± 2.3	72.9 ± 3.2	72.0 ± 3.9
	33	71.8 ± 2.0	73.8 ± 4.3	77.1 ± 2.4	77.7 ± 2.9	76.6 ± 4.3	74.4 ± 2.4	73.9 ± 5.5	62.9 ± 8.9
	36	69.5 ± 3.3	74.3 ± 4.0	77.5 ± 3.1	74.3 ± 2.9	77.0 ± 3.6	70.3 ± 4.2	67.9 ± 6.0	66.2 ± 0.8
	39	72.4 ± 0.6	72.2 ± 2.5	75.0 ± 3.5	78.2 ± 3.7	75.6 ± 2.2	62.9 ± 6.8	64.0 ± 1.8	62.5 ± 4.9
	42	70.7 ± 2.2	73.0 ± 2.3	76.2 ± 3.2	77.4 ± 2.8	72.3 ± 3.1	66.7 ± 5.1	63.6 ± 2.5	62.8 ± 2.3

Table 4: Comparison between our approach and the previous published works addressing the gender recognition from whole body static images (Section 2) reporting results using the same dataset as ours (CBCL pedestrian database (Oren et al., 1997)).

	Balanced dataset?	Uses (Cao et al., 2008) manual labelling?	Frontal view accuracy	Back view accuracy	Mixed view accuracy
(Cao et al., 2008)	No	Yes	76.0 ± 1.2	74.6 ± 3.4	75.0 ± 2.9
(Collins et al., 2009)	Yes	No	76.0 ± 8.1	Not reported	Not reported
(Guo et al., 2010)	No	Yes	79.5 ± 2.6	$\mathbf{84.0 \pm 3.9}$	$\mathbf{80.6 \pm 1.2}$
Ours	No	Yes	$\mathbf{81.9 \pm 5.0}$	80.8 ± 2.3	79.4 ± 1.4

Triggs, 2005) are needed for gender recognition. The optimal grid varies with the point of view of the figure resulting in a 360% times denser grid in the case of the back view and 120% times denser in the case of the frontal view, and comparable density while recognizing in the mixed view. This variation leads us to guess that (Guo et al., 2010) approach (i.e. first detecting the view and then recognizing the gender using an optimal classifier for that view) is possibly a directive to follow and requires further study.

The importance of the grid is evidenced by the state of the art results, outperformed in the case of the frontal view and nearly equalled in the case of the mixed view, using classifiers simpler than those proposed in the literature (Section 2) and a simple feature extractor (R-HOG, Section 3).

We think there is a need for a dataset specifically created to test gender recognition algorithms, large

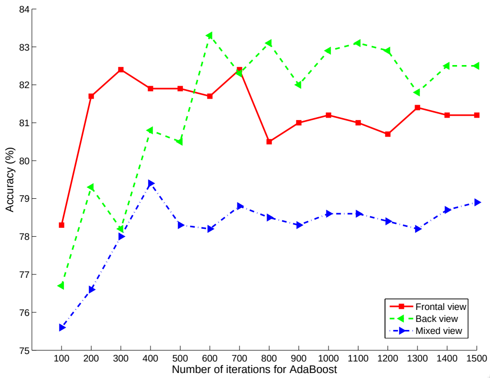


Figure 3: Evolution of the accuracy (%) as the number of iterations for AdaBoost increases, obtained for the optimal grid found in Section 4.3 for each view: 21×12 for the frontal, 36×21 for the back, and 15×15 for the mixed.

enough to allow the use of a separate test set different from that used in the validation scheme in order to obtain a more realistic accuracy rate (Alpaydın, 2010).

Another aspect to be considered in the future is the unbalanced distribution of the classes. One way of managing the unbalanced nature of a dataset is the method proposed by (Kang and Cho, 2006), used for example in a recent work dealing with gender recognition through gait (Martín-Félez et al., 2010).

ACKNOWLEDGEMENTS

The authors acknowledge the Spanish research programme Consolider Ingenio-2010 CSD2007-00018, and Fundació Caixa-Castelló Bancaixa under project P1·1A2010-11. Carlos Serra-Toro is funded by Generalitat Valenciana under the "VALi+d program for research personnel in training" with grant code ACIF/2010/135.

We thank Liangliang Cao (Cao et al., 2008) for his help in understanding and reproducing their algorithm PBGR (Section 2). This research uses the CBCL pedestrian database (Oren et al., 1997) collected by the Center for Biological & Computational Learning (CBCL) at MIT.

REFERENCES

Alpaydın, E. (2010). *Introduction to Machine Learning*. The MIT Press, Cambridge, Massachusetts.

Cao, L., Dikmen, M., Fu, Y., and Huang, T. S. (2008). Gender recognition from body. In *MM '08: Proceeding of the 16th ACM international conference on Multimedia*, pages 725–728, New York, NY, USA.

Collins, M., Zhang, J., Miller, P., and Wang, H. (2009). Full body image feature representations for gender profiling. In *IEEE ICCV Workshops*, pages 1235–1242.

Collins, M., Zhang, J., Miller, P., Wang, H., and Zhou, H. (2010). Eigenbody: Analysis of body shape for gender from noisy images. In *International Machine Vision and Image Processing Conference*.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE CVPR*, volume 1, pages 886–893.

Freund, Y. and Schapire, R. (1995). A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational Learning Theory*, volume 904 of *LNCS*, pages 23–37. Springer Berlin / Heidelberg.

Gray, D., Brennan, S., and Tao, H. (2007). Evaluating appearance models for recognition, reacquisition, and tracking. *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*.

Guo, G., Mu, G., and Fu, Y. (2010). Gender from body: A biologically-inspired approach with manifold learning. In *Computer Vision ACCV 2009*, volume 5996 of *LNCS*, pages 236–245. Springer Berlin / Heidelberg.

Jolliffe, I. (2005). *Principal Component Analysis. Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons, Ltd.

Kang, P. and Cho, S. (2006). EUS SVMs: Ensemble of under-sampled SVMs for data imbalance problems. In *Neural Information Processing*, volume 4232 of *LNCS*, pages 837–846. Springer Berlin / Heidelberg.

Martín-Félez, R., Mollineda, R. A., and Sánchez, J. S. (2010). A gender recognition experiment on the CASIA gait database dealing with its imbalanced nature. In *VISAPP*, volume 2, pages 439–444, Angers (France).

Mease, D. and Wyner, A. (2008). Evidence contrary to the statistical view of boosting. *Journal of Machine Learning Research*, 9:131–156.

Moghaddam, B. and Yang, M.-H. (2002). Learning gender with support faces. *IEEE PAMI*, 24(5):707–711.

Oren, M., Papageorgiou, C., Sinha, P., Osuna, E., and Poggio, T. (1997). Pedestrian detection using wavelet templates. In *IEEE CVPR*, pages 193–99.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE PAMI*, 29(3):411–426.

van de Sande, K. E. A., Gevers, T., and Snoek, C. G. M. (2008). Evaluation of color descriptors for object and scene recognition. In *IEEE CVPR*, Anchorage, Alaska, USA.

Yu, S., Tan, T., Huang, K., Jia, K., and Wu, X. (2009). A study on gait-based gender classification. *IEEE TIP*, 18(8):1905–1910.

Zhang, T. and Yu, B. (2005). Boosting with early stopping: Convergence and consistency. *The Annals of Statistics*, 33(4):1538–1579.