Towards a More Realistic Appearance-Based Gait Representation for Gender Recognition

Raúl Martín-Félez, Ramón A. Mollineda and J. Salvador Sánchez
Institute of New Imaging Technologies, DLSI, Universitat Jaume I, Castellón (Spain)
{martinr, mollined, sanchez}@uji.es

Abstract

A realistic appearance-based representation of side-view gait sequences is here introduced. It is based on a prior method where a set of appearance-based features of a gait sample is used for gender recognition. These features are computed from parameter values of ellipses that fit body parts enclosed by regions previously defined while ignoring well-known facts of the human body structure. This work presents an improved regionalization method supported by some adaptive heuristic rules to better adjust regions to body parts. As a result, more realistic ellipses and a more meaningful feature space are obtained. Gender recognition experiments conducted on the CASIA Gait Database show better classification results when using the new features.

1. Introduction

Humans are very accurate at recognizing gender from a face, a voice or the manner in which an individual walks (gait). Nevertheless, unlike voice or a face, gait can be perceived at a distance. Furthermore, gait has additional advantages with regard to other biometric features: it is non-contact, non-invasive and, in general, does not require subjects’ willingness.

These issues have stirred up the interest of the computer vision community for conceiving gait-based gender recognition systems [4, 7, 5, 3]. A number of applications may benefit from the development of such systems: demographic analysis of a population, advanced interaction of robots, biometric systems that use gender recognition to reduce the search space to half, etc.

Most of gait research has been addressed to biometric identification, which consists of predicting the identity of a person according to her/his way of walk. However, few recent works have used gait analysis for other classification tasks, such as gender recognition or age estimation. Regardless of the purposes, two different approaches to describe gait can be considered. Some works extract dynamic features from subject’s movements [2, 6], while others take static attributes related with the appearance of the subject [4, 3, 7], what implicitly might contain motion information.

An early appearance-based method to recognize gender from gait [4] divides silhouettes bounding boxes into seven regions following fixed rules, disregarding well-known facts of the human body structure and also, to some extent, the variability of human body proportions. The silhouette parts enclosed by these regions were fitted by ellipses, whose parameters were concatenated and averaged, across all frames of a gait sequence, to built a feature vector that represents a gait sample.

This paper proposes a method to divide a body silhouette into more natural body parts, which can be fitted by more realistic ellipses. Unlike previous work, adaptive rules based on simple image analysis techniques are devised to better adjust regions to body parts. The new regionalization method is able to better deal with deficient segmentation of silhouettes. Both the original method [4] and the new proposal are implemented and used to characterize gait samples from the CASIA Gait Database [1] for gender recognition experiments. As a result, the classifier learning from the new features achieves higher classification accuracy computed on the basis of a more balanced class recognition rates. This is particularly important because of the imbalance between the distributions of samples per class in the CASIA dataset, which has 31 women and 93 men.

2. On the extraction of body parts

The approach here proposed for human gait characterization is inspired in the method introduced by Lee and Grimson [4] (LGM), where the silhouette appearance of a side-view human walking sequence was described by static features that were further used for person identification and gender recognition. Given a gait
video sample, the original LGM can be defined in terms of the following steps (see Fig. 1):

**Foreground segmentation.** For each frame, the foreground (silhouette) is segmented from the background. A new binary frame with the silhouette highlighted is obtained.

**Silhouette extraction.** A reduced image defined by the bounding box that encloses the silhouette is extracted to normalize its size and location.

**Regionalization.** The bounding box is divided into seven regions obtained from the silhouette centroid and from fixed percentages of the box sizes.

**Ellipse fitting.** Seven ellipses, one per region, fit those silhouette pixels enclosed by the regions.

**Feature extraction.** Four features per ellipse are extracted: the x and y-coordinates of the centroid, the orientation of the major axis \((\alpha)\) and the aspect ratio \((axis_1[axis_2])\). An extra global feature, which consists of the quotient of the y-coordinate of the silhouette centroid to the silhouette height, is also considered.

**Gait representation.** To represent the gait video sample (changes in silhouette poses across the frames), the mean and the standard deviation of the four parameters of each ellipse are computed across all the frames of the sequence. The eight resulting features of each of the seven ellipses are concatenated, along with the mean of the extra global feature, to built a 57-dimensional vector.

This paper proposes an improvement on the regionalization stage, according to new rules that provide a more realistic division of the human body.

### 2.1 Regionalization stage in LGM

LGM begins by dividing the body into two horizontal parts, one above and one below the centroid of the silhouette. In turn, the part located above de centroid is blindly subdivided into two halves: the upper half is associated to the head, while the lower one is related to the torso. Thus, frequently the head region also includes the neck and the shoulders. In a subsequent horizontal division, the part located below the centroid is then separated into two new halves, the upper one for thighs and the lower one for calves. Then, the x-coordinate of the centroid defines a vertical division of the torso, thigh and calf regions into their front and rear parts. The two regions located at the torso level get torso segments mix up with arms, what produces unrealistic body parts. See Fig. 1 for details of this regionalization.

### 2.2 A more realistic regionalization stage

The proposed regionalization approach aims at getting a better and more comprehensive fit of the human body. For such a purpose, unlike the seven regions defined by the original LGM, the new proposal divides the body silhouette into eight more realistic regions: head, torso, two arms, two thighs and two calves/feet. A comparative overview of both approaches is illustrated in Figure 1. From a procedural perspective, this approach performs three main improvements with respect to the early LGM:

**Adaptive head location.** The neck is automatically detected to define the head region. To this end, the vertical projection of the silhouette is computed and, considering a medium-size windows on the projection, the uppermost minimum is chosen as the neck position. It has proved to be robust to noise and holes, as can be seen in Figure 2a).

**Independent regions for torso and arms.** Unlike LGM, the torso is fully segmented, and the two arms are independently located (if they are visible). To detect the torso, the convex hull of the top half of the rectangular region delimited by the y-coordinates of the neck and the centroid, is computed. This structure provides its minimum and maximum x-coordinates that, along with the previous y-coordinates, define the torso region. These x-coordinates also determine, to both sides of the torso, the regions that may probably enclose the arms (see Fig. 2b). These arm regions might provide important information about the movement extent of a gait sample.

**Adaptive leg location.** The segmentation of thighs and calves/feet has also been modified by an adaptive heuristic rule (see Figure 2c). As in LGM, the rectangular area located below the silhouette centroid was divided into four parts for thighs and calves/feet, keeping a common height. However, they are not divided vertically from the silhouette centroid. For each horizontal half, thighs or calves, if only one blob (foreground connected component) is detected, the vertical division is set by the blob centroid. Otherwise, if the region has two blobs, for example, when the two legs are not connected, the vertical division is set by the mid-point between both blobs. In addition, small blobs containing hands or noises are ignored, contributing to a better ellipse fitting.

Once the new eight regions are defined, the rest of the process follows as in LGM: each region is fitted by
an ellipse, four parameters are extracted from each ellipse, their mean and standard deviation are computed across all frames generating eight features per region, which are finally putting together, along with the extra global quotient, to build a 65-dimensional vector (instead of the 57 attributes of LGM).

3. Experiments

The aim of the experiments is to compare the discriminant power, on a gender recognition task, of the two feature sets generated by the regionalization approaches described in Section 2.

Experiments involve the CASIA Gait Database [1] due to its availability and moderate size. It contains gait video samples of 124 people, 31 women and 93 men. From each person, six side-view gait sequences of segmented silhouettes are available, resulting in a collection of 744 sequences distributed in 186 and 558 samples from women and men, respectively. Each sequence was represented by both a 57-dimensional vector and a 65-dimensional vector as regards the approach followed, as explained in Section 2.

Some works [7, 3, 4] have used this database for gender recognition tasks. However, their experiments have been formulated on the basis of small subsets with equal number of subjects per class, disregarding the imbalance nature of the original data. In addition, the classification performance was measured in terms of the global accuracy, ignoring individual class error rates and therefore, possible biased behaviours of the classifiers. Empirical evidence shows that accuracy can be strongly biased with respect to class imbalance.

In this work, apart from accuracy, other measures more appropriate for imbalanced problems are also considered. In particular, the individual class recognition rates, the True Positive rate (TPr) and the True Negative rate (TNr), are used alone, and they are also combined to compute their Geometric mean, $G_{mean} = \sqrt{TPr \times TNr}$, and the Area Under the ROC Curve for a single classification result, $AUC = (TPr + TNr)/2$. These last two metrics are unbiased as regards the distribution of samples between classes.

The performance measures were estimated by repeating 10 times a 10-fold cross-validation scheme with a Support Vector Machine (SVM) with a linear kernel. Training and test partitions keep the a priori class probabilities.

3.1 Experimental results

Table 1 shows the mean values of the 10 repetitions of the 10-fold cross validation scheme and the 95% confidence intervals from a Student’s t-test, which was used to find out whether two series of classification results were statistically significantly different.

When focusing on classification accuracy, the new proposal significantly outperforms the LGM. This higher accuracy is supported by a considerable increase of the recognition rate in the women class (TPr), from...
Table 1. Experimental results.

<table>
<thead>
<tr>
<th></th>
<th>LGM</th>
<th>New proposal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>93.1% ± 0.73%</td>
<td>94.7% ± 0.19%</td>
</tr>
<tr>
<td>TPr</td>
<td>78.4% ± 1.92%</td>
<td>85.1% ± 0.33%</td>
</tr>
<tr>
<td>TNr</td>
<td>98% ± 0.50%</td>
<td>97.9% ± 0.21%</td>
</tr>
<tr>
<td>Gmean</td>
<td>87.6% ± 1.17%</td>
<td>91.3% ± 0.21%</td>
</tr>
<tr>
<td>AUC</td>
<td>88.2% ± 1.07%</td>
<td>91.5% ± 0.2%</td>
</tr>
</tbody>
</table>

78.4% to 85.1%, and by an imperceptible degradation in the classification rate of the men class (TNr), from 98% to 97.9%. These results prove that the new feature set was able to correct part of the biased behaviour of the SVM classifier using the LGM attributes. In other words, the new feature space reduced the harmful influence of the imbalanced nature of CASIA database on the classifier behaviour. This fact is confirmed by the Gmean and AUC metrics, which give unbiased joint views of the two class rates. Gmean and AUC results obtained from the use of the new feature set were better than those achieved by SVM with the LGM attributes.

A straightforward comparison between these classification results and those reported on CASIA database by other previous related works is not appropriate, because all of them are defined regarding reduced database subsets, different feature sets, classification models and error evaluation schemes. Nevertheless, to allow for a broader analysis, the main gender recognition results are presented in terms of accuracy, which has been the only measure priorly used. In [3], the use of the LGM led to an accuracy of 85%, that is much smaller than the one here obtained. A better result of 95.97% was achieved in [7], but the database was artificially balanced by randomly picking up 31 men. However, as demonstrated above, accuracy is not a reliable measure for imbalanced scenarios.

4. Conclusion and future work

This paper introduces a realistic appearance-based representation of gait sequences for automatic gender recognition. It is based on a prior method proposed by Lee and Grimson [4] (LGM), where the silhouette appearance of a gait sequence was regionalized and then, the resulting parts of the silhouette were fitted by a collection of ellipses. This work presents an improved regionalization method supported by some adaptive heuristic rules to better adjust regions to body parts. As a result, more realistic ellipses and a more meaningful feature space are obtained.

Both approaches have here been compared as regards the discriminant power of the generated feature sets, in a gender recognition task on the CASIA Gait Database [1]. It contains gait samples from 124 subjects unequally distributed into 31 women and 93 men. Apart from the plain accuracy, the classification performance has also been measured by two metrics appropriate for imbalanced databases: the geometric mean of class recognition rates, and the Area Under the ROC Curve. For the three metrics used, the classification results obtained from the new features have been better than those achieved from the LGM-based ones. When focusing on the two unbiased metrics, differences have been statistically significant. These results can be explained by scrutinizing the recognition rates for each individual class: a considerable increase in the minority class (women), and a small degradation in the majority class (men).

To the best of our knowledge, the papers that previously have worked on this database within the gender recognition problem have used reduced balanced subsets. Unlike them, this contribution manages all the samples and analyzes the results from an imbalance viewpoint. Therefore, this paper might be particularly useful for benchmark purposes.

ACKNOWLEDGEMENTS

Partially supported by projects CSD2007-00018 and CICYT TIN2009-14205-C04-04 from the Spanish Ministry of Innovation and Science, P1-1B2009-04 from Fundació Bancaixa and PREDOC/2008/04 grant from Universitat Jaume I.

References


