

Gait recognition from corrupted silhouettes: a robust statistical approach*

Javier Ortells¹, Ramón A. Mollineda¹, Boris Mederos², Raúl Martín-Félez¹

¹Institute of New Imaging Technologies, Universitat Jaume I, Castellón de la Plana (Spain)

²Departamento de Física y Matemática, Universidad Autónoma de Ciudad Juárez, Ciudad Juárez (México)

¹{jortells,mollined,martinr}@uji.es, ²boris.mederos@uacj.mx

Abstract

This paper introduces a method based on *robust statistics* to build reliable gait signatures from averaging silhouette descriptions, mainly when gait sequences are affected by severe and persistent defects. The term *robust* refers to the ability of reducing the impact of silhouette defects (outliers) on the average gait pattern, while taking advantage of clean silhouette regions. An extensive experimental framework was defined based on injecting three types of realistic defects (salt and pepper noise, static occlusion, dynamic occlusion) to clean gait sequences, both separately in an easy setting and jointly in a hard setting. The robust approach was compared against two other operation modes: i) simple mean (weak baseline), and ii) defect exclusion (strong benchmark). Three gait representation methods based on silhouette averaging were used: Gait Energy Image (GEI), Gradient Histogram Energy Image (GHEI), and the joint use of GEI and HOG descriptors. Quality of gait signatures was assessed by their discriminant power in a large number of gait recognition tasks. Non-parametric statistical tests were applied on recognition results, searching for significant differences between operation modes.

*This work has been supported by the grants P1-1B2012-22 and PREDOC/2012/05 from Universitat Jaume I, PROMETEOII/2014/062 from Generalitat Valenciana, and TIN2013-46522-P from Spanish Ministry of Economy and Competitiveness.

Keywords— Gait recognition, model-free, noisy silhouettes, occluded silhouettes, robust statistics

1 Introduction

Biometric systems have become increasingly widespread as the need for higher security levels has grown. They address person identification or verification by analyzing physiological or behavioral traits. The use of physical portions of the body, such as face, fingerprint and hand geometry, has been more popular due to the traditional cooperation of the subject and the inherent static nature of these biometric sources. Conversely, the use of behavioral traits is intended to encode the singular way in which a human performs a common action (walking, signing, typing, etc.). Their main advantage over physiological traits is the potential of capturing people doing in real life, although the need of managing (possibly unconstrained) dynamic information could lead to inaccurate biometric patterns.

This paper focuses on recognizing human gait, a behavioral biometric source that has received much attention in the last two decades [11, 13]. In addition to not requiring subject's cooperation, human gait is a universal action that can be captured at a distance by simple sensors, even in adverse conditions. These strengths have made this biometric a highly valuable information source in video-based security and surveillance systems.

From a biomechanical point of view, each person

is supposed to have a unique walking pattern because this action is supported by a particular musculoskeletal structure [25]. However, this uniqueness is usually very hard to elicit due to a large number of factors that can affect either the gait dynamic or the gait perception. Dynamic may be altered by surface, footwear, age, body weight, mood, physical injuries, and neurological disorders. Similarly, gait perception depends on subject appearance and video quality. Appearance can be affected by changes in clothing, load carrying, and camera viewpoint, while quality can degrade by the presence of noise or occlusions. Gait description under severe quality problems defines the scope of this work.

Representation methods are expected to be able to extract discriminant information, allowing classifiers to recognize or verify a person’s identity by their gait. They have been broadly separated into two major families: model-based and model-free approaches. Model-based methods build a predefined model of the human walking by continuously measuring dynamic attributes such as joint angles and body part locations, what makes these approaches robust to changes in viewpoints and scale. However, they tend to be error-prone and time-consuming due to the need of estimating parameters. On the contrary, model-free techniques do not use a explicit body model. They usually capture subject’s dynamic and appearance directly from binary silhouettes, being thus robust to changes in color, texture, and lighting conditions, but also sensitive to viewpoints. Besides, unlike model-based approaches, they are simpler (a model is not required), less costly, and able to better encode body shape (appearance), which carries considerable biometric information [6, 13]. These advantages have made model-free methods an appealing choice for gait analysis in uncontrolled scenarios.

Both the model-based and model-free approaches have been applied to the analysis of low-quality gait sequences. Those methods that rely on a model [12, 15, 23, 36] have showed an outstanding ability to reconstruct incomplete body parts (due to occlusions or segmentation errors), although the resulting gait representations usually suffer from standardization without much of their individual information. Poor gait recognition results reported in some of these works

validate the detrimental effect of model fitting. On the other hand, a number of general-purpose model-free methods [4, 6, 7] have been proposed and successfully assessed on imperfect silhouettes. However, most of the gait samples used in these researches are just slightly affected by a few minor defects scattered along the sequences.

With the aim of having heavily contaminated gait sequences, just a few works [4, 23, 35, 36] have injected artificial major defects to human silhouette images to simulate the impact of environmental factors in the segmentation process. Common practices have been salt and pepper noise, that can potentially arise due to a continuously changing background (e.g. water in motion, tree leaves blowing in the wind) or sensor malfunction, and partial occlusions caused by static objects (e.g. vertical and horizontal bars) or by superposition of silhouettes of the person of interest and other objects in motion (e.g. people, cars). All these works injected only one type of defect in each sequence, while the impact of the size of the contaminated part on the recognition accuracy was not assessed.

This paper introduces a method for a more reliable computation of a gait representation that results from averaging silhouette descriptions. Typical examples would be the Gait Energy Image (GEI) [6] and GEI-based methods such as the Gradient Histogram Energy Image (GHEI) [7]. The proposed approach is inspired by a statistical framework called *robust statistics* (RS) [10, 17] which, unlike the classic conjecture of data normality, assumes an approximately normal distribution where most data fit a normal shape, but there are heavy tails of atypical observations (outliers). The term *robust* refers to mitigation of the impact of outliers in parameter estimation. Within the context of gait representation, this proposal is intended to neutralize the influence of silhouette defects (outliers) on the average gait pattern under construction, while it takes advantage of clean silhouette regions.

Experiments have been organized in two levels of defect injection, as regards the portion of gait sequence that was affected. In an easy setting, only one fifth of each sequence was contaminated, while in a hard one, three fifths were corrupted. Three types

of defect (salt and pepper, static occlusion, dynamic occlusion) were added to gait sequences separately in the easy setting and jointly (in a random manner) in the hard setting. Gait patterns were obtained from combining three gait representation methods (GEI, GHEI, GEI+HOG) and three operation modes (simple mean, defect exclusion, robust mean). To assess the quality of gait patterns, a number of recognition tasks based on two classifiers (1-NN, RankSVM [19]) were designed: 108 and 36 in the easy and the hard settings, respectively. Besides, a neutral setting was defined as an auxiliary collection of 24 tasks built from the original clean sequences. Non-parametric statistical tests were applied on recognition results, searching for significant differences between operation modes and representation methods.

Summarizing, the main contributions of this paper are: i) a robust statistical approach to obtain reliable gait patterns from averaging silhouette descriptions, and ii) a thorough experimental study of gait recognition on a wide range of defective sequences.

The rest of the paper is structured as follows. Section 2 examines the literature that concerns quality of gait representations. Section 3 provides a theoretical basis for the method proposed, which is introduced in Sect. 4. Experimental methodology is outlined in Sect. 5. Section 6 presents statistical results and performance curves. Finally, Sect. 7 discusses conclusions and promising directions for future research.

2 Related work

Some works in the literature have dealt with low-quality gait sequences. They could be roughly divided into two groups. On the one hand, some authors have focused on devising methods to grade the level of complexity or degradation of gait samples. Within this context, some strategies have been proposed to improve or reconstruct the quality of degraded silhouettes. On the other hand, a number of gait representations have been designed to work directly with defective gait samples. Hereinafter, most relevant works from both groups are reviewed.

A method for measuring the quality of a range of silhouettes from their 1-D foreground-sum signal

was proposed in [14]. This metric, named Silhouette Quality Quantification (SQQ), was exploited to weight gait patterns by their quality in order to improve the recognition rate. Experiments on raw sequences from videos recorded in a challenging environment (complex background, variations in illumination) yielded recognition results higher than those of a baseline. In a related work [20], a GEI complexity index is computed based on a probabilistic model, to quantify how far the sample under analysis is from a normality model. Experiments showed a high correlation between the complexity index and the recognition error. Recently, a novel occlusion model was presented in [22], to statistically describe the level of occlusion in videos based on three parameters: the initial phases of motion of both the target and the occluder subjects, and the duration of the occlusion. It was employed to synthesize dynamic and static occlusions in clean videos, as well as to characterize real occlusions in defective videos from challenging databases. Experiments showed the precision of the occlusion model, in addition to its usefulness in designing realistic gait recognition tasks.

Within this scope, some few works have focused on strategies to improve the quality of degraded gait silhouettes. In [23], gait cycles were modeled as chains of estimated key poses. This scheme allows to detect partially occluded and missing silhouettes in the frames of a gait sequence, which are then reconstructed using a Balanced Gaussian Process Dynamical Model. This solution was tested on sequences occluded by real static and dynamic objects, and on frames degraded by normal distributions. A simpler approach based on the amount of foreground pixels was proposed in [9]. It is aimed to detect gait sub-sequences where silhouettes appear partially or totally occluded. Affected silhouettes are then replaced by similar-pose clean silhouettes retrieved from non-affected cycles. Although this method effectively handles occlusions, it is at the expense of replicating information. In [4], the problem of silhouette incompleteness was also addressed. The method consists in classifying the raw silhouettes into clusters, and computing a GEI from each cluster. Then, each GEI is denoised resulting in a Dominant Energy Image (DEI). Finally, each original silhouette is substituted

by a new image (FDEI) that arises from the summation of its cluster’s DEI and the positive portion of its difference with the preceding silhouette. A different strategy to deal with problems in silhouettes is to exclude affected regions as in [30], where covariate factors are removed from gait representations.

Other works rely on prior models to clean noisy silhouettes or to reconstruct missing parts of them. In [12], a global pedestrian population model and subject-dependent HMM-based models were created to refine and fill in missing parts of silhouettes, possibly caused by a faulty segmentation. A similar procedure was followed in [15], where an eigen-stance gait model was created from a series of manually selected silhouettes. This model, along with an HMM-driven strategy, was applied to match silhouettes to stances so that noise could be detected and removed. Despite of the improvement in representing appearance, these approaches led to low recognition rates due to the loss of individual clues.

Some model-based and model-free approaches have been proposed to obtain gait patterns directly from low-quality gait sequences. In [36], authors built a simplified articulated model which accurately fits the silhouettes, even in presence of noise or occlusions. The model is defined by a series of static and dynamic parameters used to characterize gait poses. The method was tested in recognition tasks, outperforming a baseline in outdoor scenarios but not indoors. As regards the model-free family, methods range from simple gait characterizations as in [33], where a contour-based approach is designed to mitigate little defects on silhouettes, to intricate methodologies as in [35], where a fractal-based gait description is introduced. However, the most popular model-free methods are those based on the GEI, which can be considered a de-facto standard. GEI is computed as the average image of a series of normalized binary silhouettes previously extracted from a gait video. A similar strategy is followed in [7], where histograms of oriented gradients (HOG) are first computed from individual silhouettes. Then, histograms are averaged to produce the Gradient Histogram Energy Image (GHEI). Other approaches [27, 29] also make use of HOG to describe a gait sequence, but HOG descriptors are extracted directly from the GEI. All

these proposals share a key operational issue: they compress silhouette information by averaging, which effectively reduces the negative effects of scattered defects. However, they fail in case of major and persistent over time defects.

3 Statistical framework

3.1 Robustness

This section introduces basic concepts of *robust statistics* that appear in the book [17]. Readers are encouraged to consult it for a more in-depth understanding.

Given a set of observed values $\{x_1, x_2, \dots, x_n\}$, the maximum likelihood estimate (MLE) of the mean μ can be expressed by the following optimization problem:

$$\hat{\mu} = \operatorname{argmin}_{\mu} \sum_{i=1}^n \rho(x_i - \mu) \quad (1)$$

where $\rho = -\log f$, with f being the underlying probability density function of the error $e_i = x_i - \mu$. If ρ is differentiable, then differentiating (1) with respect to μ and equating to zero lead to:

$$\sum_{i=1}^n \psi(x_i - \hat{\mu}) = 0, \text{ with } \psi = \rho' \quad (2)$$

Assuming that $e_i \sim N(0, \sigma^2)$, the solution of Eq. (2) is:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

This way of estimating the mean equally weights all observed values, which opens the door to the negative impact of outliers. It suggests a straightforward strategy to deal with outliers: to detect them [2] and to leave them out. However, as discussed in [17], the decision of removing outliers is inherently subjective because the outlyingness of observations should be measured and thresholded. In addition, it encloses the risk of discarding “genuine” observations which could induce a bias in the mean estimation. Another alternative is to use the sample median, as it has proven to be less sensitive to outliers. Nevertheless,

the statistical performance provided by the median is generally poorer than that of the mean when data contain no outliers. Thus, a good solution should behave like the mean when no outliers affect the data, while it should ignore outliers otherwise. This is just the ultimate goal of robust estimation.

In most cases of interest, $\psi(0) = 0$ and $\psi'(0)$ exists. Let $W(x)$ be a function defined from $\psi(x)$ as follows:

$$W(x) = \begin{cases} \psi(x)/x & : x \neq 0 \\ \psi'(0) & : x = 0 \end{cases} \quad (4)$$

Then, using the $W(x)$ function, Eq. (2) can be reformulated as:

$$\sum_{i=1}^n W(x_i - \hat{\mu})(x_i - \hat{\mu}) = 0 \quad (5)$$

From Eq. (5), the sample mean can be expressed in terms of a weighted mean:

$$\hat{\mu} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \text{ where } w_i = W(x_i - \hat{\mu}) \quad (6)$$

The Eq. (6) establishes a weighted computation of the sample mean $\hat{\mu}$, where the term $w_i = W(x_i - \hat{\mu})$ weights the observation x_i . Note that $\hat{\mu}$ appears in both the left- and the right-hand sides of the Eq. (6), thus it can be rewritten as a recurrence and solved by a numerical iterative method, typically a fixed point algorithm.

Within this robust framework, ρ is chosen in order to ensure that $W(x)$ is a symmetric, non-increasing function of $|x|$, and that $W(x) \rightarrow 0$ when $x \rightarrow +\infty$. Thus, the farther away is an observation x_i from the sample mean $\hat{\mu}$, the smaller will be the associated weight $W(x_i - \hat{\mu})$. Accordingly, outliers should receive small weights, reducing their impact in the mean estimation.

There exist several examples of robust functions [3]. A popular choice for ρ and ψ is the bisquare family, from which the following weight function can be deduced:

$$W_b(x|t) = \begin{cases} \left[1 - \left(\frac{x}{t}\right)^2\right]^2 & : |x| \leq t \\ 0 & : |x| > t \end{cases} \quad (7)$$

Note that $W_b(x|t)$ has non-trivial zeros at $x = t$ and $x = -t$, beyond which the function vanishes. That is, any x_i located at a distance of $\hat{\mu}$ greater than or equal to t , will have no impact on the mean estimation. Figure 1(a) shows a plot of $W_b(x|t)$ for $t = 3$, which could correspond to $\pm 3\sigma$ when $e_i \sim N(0, 1)$.

3.2 Quasi-robustness

Reducing the influence of outliers in the mean estimation is usually performed by weight functions $W(x)$ that decrease very quickly and, unlike the bisquare-based function, approach zero at infinity. Despite weights keep rigorously greater than zero, they become very small even for moderately deviated samples, thus involving a considerable risk of losing genuine information.

A more reasonable approach could be the one known as *quasi-robust*, which is described “as being much more robust than ordinary solutions without being strictly robust” [21]. This principle can be shaped in terms of a (quasi-robust) weight function that guarantees at least a minimum weight $\epsilon > 0$ to any uncommon value (possibly an outlier), disregarding how far from the mean it is located. Formally, a quasi-robust weight function $w(x)$ should satisfy the following properties:

1. $w(x)$ is symmetric
2. $w(x)$ is a non-increasing function of $|x|$
3. $\exists \epsilon > 0$, such that $w(x) \geq \epsilon$

For instance, the bisquare-based function $W_b(x|t)$ can be generalized into a new quasi-robust realization $w_b(x|t, \epsilon)$ so that $w_b(\cdot) \rightarrow \epsilon$ when $|x| \rightarrow t$:

$$w_b(x|t, \epsilon) = \begin{cases} \left[1 - \left(\frac{x}{t}\right)^2(1 - \sqrt{\epsilon})\right]^2 & : |x| \leq t \\ \epsilon & : |x| > t \end{cases} \quad (8)$$

Figure 1(b) illustrates $w_b(x|t = 3, \epsilon = 0.05)$.

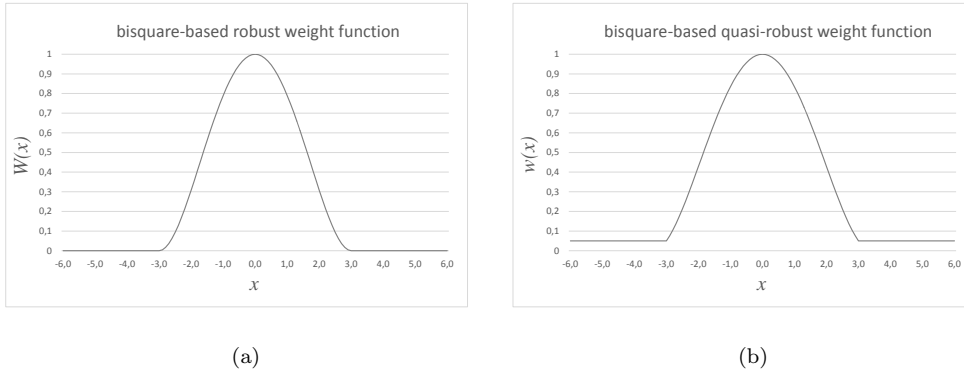


Figure 1: Bisquare-based weight functions.

4 Robust gait representation

4.1 A general recurrence

This section introduces a method inspired by robust statistics to build more reliable gait patterns.

Let $S = \{s_1, s_2, \dots, s_n\}$ be a sequence of n binary silhouette images gathered from a gait video, and let $X = \{x_1, x_2, \dots, x_n\}$ be a related set of silhouette descriptions, such that x_i is a vector of d numerical features that describes s_i . For instance, x_i could be a vector of pixel values or HOG descriptors. A gait pattern $g \in \mathbb{R}^d$ can be obtained by averaging all x_i :

$$g = \frac{1}{n} \sum_{i=1}^n x_i \quad (9)$$

Some particular g -like representations are GEI, GHEI, and the joint use of GEI+HOG.

Let us assume now that S can be divided into m disjoint subsequences S^j , $1 \leq j \leq m$, such that $S = S^1 \cup S^2 \cup \dots \cup S^m$, with each S^j corresponding to the j -th gait cycle (one stride or two steps)¹. Thus, it can be assumed that all subsequences S^j are chronologically ordered. This structure induces an equivalent partition on the set X , leading to a related collection of silhouette description subsets $\{X^1, X^2, \dots, X^m\}$. Let g_j be a cycle-based gait representation obtained by averaging all silhouette de-

¹This scheme admits a series of half cycles.

scriptions in X^j . An alternative (and in general approximate) way of computing g is:

$$g \approx \frac{1}{m} \sum_{j=1}^m g_j \quad (10)$$

Since all g_j result from averaging and given the cyclical nature of gait, each of the d features over the set $\{g_1, g_2, \dots, g_m\}$ is expected to approach a normal distribution. Then, it makes sense to measure the deviation of a particular feature value with respect to that feature's mean, $|g_j(k) - g(k)|$, where k , $1 \leq k \leq d$, denotes the k -th feature of g_j . As each g_j summarizes silhouette information within the j -th cycle, a high deviation value $|g_j(k) - g(k)|$, assuming a reliable $g(k)$, could be a symptom of poor quality data in regions that contribute to the feature $g_j(k)$. It is important to note that any perceived anomaly in g_j necessarily comes from a persistent cause along that cycle j , such as partial occlusions, serious segmentation errors, etc.

This paper proposes a *quasi-robust* formulation of Eq. (10) based on an incremental method introduced in [18], which computes cumulative gait representations from cycles in the order they occur. Given a gait sequence that consists of m cycles, the Eq. (10) can be rewritten as the following recurrence:

$$g \approx g_{1:m} = (1 - \alpha_m) g_{1:(m-1)} + \alpha_m g_m \quad (11)$$

with $\alpha_j = 1/j$, $g_{1:j}$ denoting a gait representation

that aggregates the first j cycles, and $g_{1:1} = g_1$ being the seed value. Equation (11) equally weight (by $1/m$) all features in all gait patterns, no matter how corrupted they can be. The solution of this recurrence leads to an incremental computation of the gait pattern, from the first cycle (seed) to the final one. All cycle-based patterns g_j are ultimately weighted by $1/m$, hence the order of cycles does not affect the resulting $g_{1:m}$.

Let us assume that the seed value $g_{1:1} = g_1$ is a clean cycle-based gait pattern built from high-quality silhouettes comprised in the first cycle, and let α_{jk} be a weight function defined as follows:

$$\alpha_{jk} = \frac{\omega_{jk}}{\sum_{i=1}^j \omega_{ik}},$$

with $\omega_{ik} = \begin{cases} w_k(g_i(k) - g_{1:i-1}(k)) & : i > 1 \\ 1 & : i = 1 \end{cases}$ (12)

where $w_k(\cdot)$ is a quasi-robust weight function as characterized in Sect. 3.2, and $g_{1:i}(k)$ is the accumulated value of the feature k along the first i cycles. Then, a feature-dependent generalization of Eq. (11) can be:

$$G(k) = g_{1:m}(k) = (1 - \alpha_{mk}) g_{1:m-1}(k) + \alpha_{mk} g_m(k) \quad (13)$$

Equation (13) can be seen as an approximate and incremental way of computing Eq. (6). Since the contribution of each $g_j(k)$ is affected by a history-dependent factor $w_k(\cdot)$, embedded in α_{jk} , Eq. (13) depends on the order of cycles. In order to make Eq. (13) easier to understand, an analysis of $w_k(\cdot)$ is provided:

- As suggested above, $w_k(\cdot)$ fulfills the three properties stated in Sect. 3.2.
- There are d feature-dependent weight functions $w_k(\cdot)$, one for each feature k .
- The function $w_k(\cdot)$ is intended to model a threshold over the distribution of the deviation (as a random variable) of $g_i(k)$ from $g_{1:i-1}(k)$, $\forall i > 1$, to separate genuine deviations from those considered as irregular (large deviations produced by outliers).

- In the hypothetical case of zero deviations of all features across all cycles, i.e. $g_i(k) - g_{1:i-1}(k) = 0 \forall i, k$, by properties 1 and 2 (Sect. 3.2), $\omega_{ik} = \max w_k(\cdot)$ and Eq. (13) turns into the simple mean. The function $w_k(\cdot)$ should be chosen such that small deviations receive $\omega_{ik} \approx \max w_k(\cdot)$.
- In case of large deviations (greater than the threshold encoded in $w_k(\cdot)$), by property 3 (Sect. 3.2), $\omega_{ik} \approx \epsilon$. That is, the related $g_i(k)$ (possibly an outlier) will contribute by a minor weight.
- An area for a further generalization is to encapsulate the computation of a deviation measure of a new observation $g_i(k)$ from its expected value $g_{1:i-1}(k)$ within a function $v(g_i(k), g_{1:i-1}(k))$, and to create a function composition $w_k(v(\cdot))$ as proposed next:

$$\alpha_{jk} = \frac{\omega_{jk}}{\sum_{i=1}^j \omega_{ik}},$$

with $\omega_{ik} = \begin{cases} w_k(v(g_i(k), g_{1:i-1}(k))) & : i > 1 \\ 1 & : i = 1 \end{cases}$ (14)

Until now, the implicit form of $v(\cdot)$ has been the absolute value of the difference $|g_i(k) - g_{1:i-1}(k)|$. However, as will be seen later, $w_k(v(\cdot))$ allows for a greater flexibility when using different gait representation methods.

The quasi-robust approach formulated by the Eq. (13) is expected to approach the simple mean when there are no outliers, while it should underweight largely deviated samples (possibly outliers) otherwise.

4.1.1 On the applicability of the Eq. (13)

The recurrence defined in Eq. (13) can be considered a simple and natural way to address an inherently cyclical process (gait). It supports strategies for controlled injection of defects into cycles as they occur, so as to track the quality of the gait pattern under construction as a function of time/cycle. This serial

scheme could also provide reliable and early identification hypotheses, by only considering a few initial cycles, that can be of great value in real-time systems. In addition, Eq. (13) entails a much more simple solution than numerical computation required by Eq. (6), avoiding potential convergence problems inherent to the latter.

A typical approach to robust estimation consists in the iterative optimization of an initial estimate. This first state should be built from well-behaved data, so that it can be gradually refined by samples that differ slightly or moderately from it (samples that differ substantially are underrated). As Eq. (13) can only be initialized with the first cycle, it is required to be of an acceptable quality. This issue turns into critical since there are usually a few cycles (iterations) to improve the estimate. On the contrary, in case of a noisy first cycle, and thus a poor-quality initial estimate (outlier), a few iterations would be insufficient to push the estimate to a satisfactory state.

Unlike Eq. (13), conventional approaches that fit Eq. (6) operate on all samples at once. This allows for more standard initialization choices such as the sample median [17], which has proven to be resistant to outliers. This operational context could inspire more general solutions to the problem of computing robust gait patterns, provided that the full gait sequence is given as input. In this regard, well-behaved initial estimates could be obtained from any clean cycle, not necessarily the first one, or from a synthetic cycle built by linking complementary clean silhouettes picked from different partially-affected cycles. Previous goals could be even further relaxed if they are reduced to half a cycle.

However, dealing with (6) would require new solution methods (e.g. a numerical analysis algorithms) that involve a number of challenges such as convergence to local minima, low convergence speed, poor choices of initialization, etc. Besides, such a free setting would have made it impractical to thoroughly study the behavior of the robust approach.

4.2 A logistic-based weight function

In this work, a weight function based on the logistic curve is chosen to implement Eq. (13). This function,

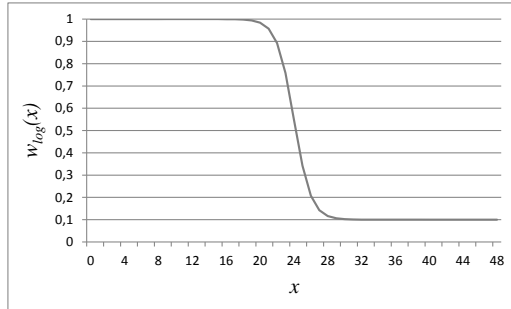


Figure 2: Logistic-based weight function, with $\epsilon = 0.1$, $s = 1$, and $t = 24$.

denoted by $w_{log}(\cdot)$, is formulated as follows:

$$w_{log}(x|t, s, \epsilon) = 1 - \frac{1 - \epsilon}{1 + e^{-s(x-t)}} \quad (15)$$

where $\epsilon = \min w_{log}(\cdot)$ (the minimum possible weight), s is the curve steepness, x is a measure of the deviation of an observation from its expected value (in the sense of $v(\cdot)$), and t is a threshold to discriminate between acceptable deviations and ill-suited ones.

Figure 2 shows a $w_{log}(\cdot)$ example, which looks similar to a step function with a sharp fall at $x = t$. This function assigns the maximum weight to those observations which are close enough to their expected values (their deviations are lower than t), while it gives a small weight otherwise. With respect to a true step function, $w_{log}(\cdot)$ is a more adaptable function that allows for more diverse shapes and smoother transitions. As can be easily proven, $w_{log}(\cdot)$ is a particular case of the class of quasi-robust functions $w(\cdot)$ defined in Sect. 3.2 and characterized in Sect. 4.1.

From an implementation point of view, two issues of the function composition $w_{log}(v(\cdot))$ should be closely examined. A first point is related to model complexity. Equation (13) involves d feature-dependent weight functions $w_k(\cdot)$, $1 \leq k \leq d$, each of them requiring parameter fitting. However, the dimensionality of most gait representation models is high, usually of thousands of features, even for low

resolution images. For the sake of simplicity, this work uses a common weight function $w(\cdot)$ to all features, which is adjusted from the distribution of deviations of all feature values together. A second key issue concerns the definition of the function $v(\cdot)$, which provides a measure of the deviation of a feature value from its expectation. Because features are closely connected with image pixels, to have a higher tolerance to segmentation and alignment inaccuracies the computation of $v(\cdot)$ should involve some contextual information of the feature under analysis. As this relationship usually depends on the representation method, different $v(\cdot)$ functions have been proposed (see Sect. 5.3.1).

5 Experimental methodology

Figure 3 depicts a methodology overview, where five areas can be identified: *Data usage*, *Silhouette defect injection*, *Gait representation*, *Parameter estimation and learning*, and *Classification and performance evaluation*. Next subsections describe these stages.

5.1 Data usage

5.1.1 Data partitioning

Given a database of gait sequences, the methodology begins by roughly equally distributing subjects (along with all their sequences) into the *Training* and the *Test* subsets. On the one hand, training data are used for two purposes: 1) to learn some transferable knowledge required by a ranking-based method used in the classification stage (RankSVM), and 2) to estimate parameters of the weight function $w_{log}(x)$. Both tasks are detailed in Sect. 5.4. On the other hand, following a classical supervised approach, test sequences are equally divided by chance into the *Gallery* and *Probe* subsets, in such a way each subject has samples in both of them.

As depicted in the Fig. 3, clean *gallery* sequences are represented by full gait patterns g to be used as reference data within a template matching strategy. This decision-making process is tested on in-between gait patterns $g_{1:i}$ built from clean and corrupted *probe*

sequences, following three operation modes: simple mean, defect exclusion, and robust mean (see Sect. 5.3.2). More details of this process are given in Sect. 5.5.2.

5.1.2 Dataset analysis

The experimental methodology proposed in this work can benefit from certain database characteristics:

1. An adequate number of subjects is recommended to have enough data for the two-level partition: first into training and test subjects, and then the test data into gallery and probe sequences. An acceptable amount could be no less than 50-60 people.
2. At least two different gait sequences *per* person under neutral appearance are required, in order to ensure at least one sample of every test subject in each of the Gallery and the Probe subsets.
3. All sequences must comprise at least four gait cycles to permit an incremental data processing as proposed here, including up to three types of defect injected in intermediate cycles.

After a thorough search of publicly available databases, two well-known collections were chosen: the OU-ISIR Treadmill Dataset B [16], and the USF Human ID Gait Database [24]. Both are large sets of sequences from people recorded several times, which broadly meet the requirements. The former is composed of indoor recordings of 68 subjects from their side view with variations on clothing up to 32 combinations. To fulfill the requirement of appearance neutrality, only two types of sequences close to a neutral appearance were used in this work. One shows subjects in regular pants and full shirt whereas the other type, in regular pants and parka. The second dataset consists of videos of 122 subjects recorded outdoors under combinations of up to five covariate conditions: 1) surface (concrete or grass), 2) view angle (left or right), 3) footwear (two types of shoes), 4) carrying condition (with or without a briefcase), and 5) time of recording (May or November). Again, sequences of the two combinations that best represent a neutral

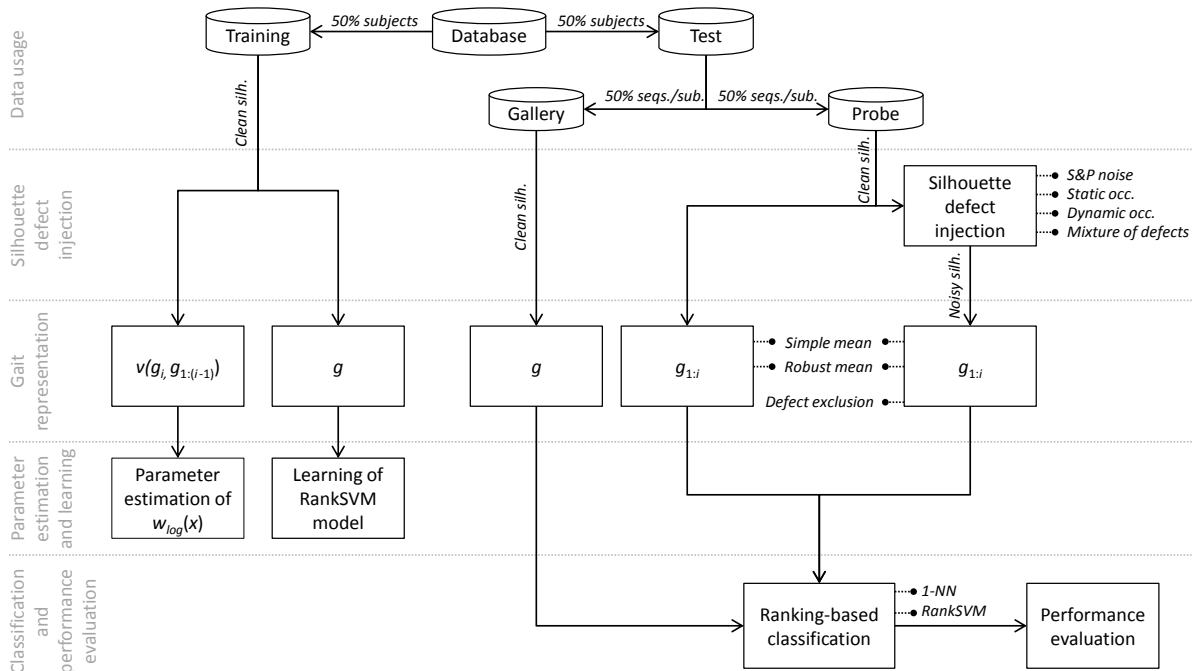


Figure 3: Overall methodology graph.

appearance were chosen for each subject. They agree in the values of four covariates (*concrete*, *shoe type A*, *no briefcase* and *May*), and just differ in the view angle. Since the goal is to recognize gait under low-quality samples, three types of defect were simulated and injected into gait sequences. Next subsection explains how this process was carried out.

It is also worthy to mention the TUM-IITKGP gait dataset [8], since it includes sequences with real static and dynamic occlusions that severely affect the gait perception. Nevertheless, some crucial dataset properties have dissuaded us from using it in this work. First, this database consists of recordings of binarized frames from 35 individuals, an amount that can be considered lacking as regards the previously stated criteria. Second, defects seem to be specific to sequences of a same individual, thus making an exogenous positive contribution to the related biometric signature. Finally, some sequences do not include even a first complete clean cycle, which is an assumption of the Eq. (13) in Sect. 4.1. Other widely used

gait databases like CMU MoBo [5] and CASIA [34] were also discarded, because they did not satisfy some of the given conditions.

5.2 Silhouette defect injection

As none of the databases analyzed comprises heavily contaminated gait sequences, three types of defects have been artificially injected in silhouettes of OUISIR and USF, simulating different contextual factors that affect the quality of segmentation. These types of defects are:

Salt & Pepper noise: When introducing S&P noise on an image, a percentage α of the pixels are randomly turned into black or white. Here, a high-level noise of $\alpha = 75\%$ has been applied on all silhouettes of a cycle, simulating a defective segmentation resulting from a scenario with a highly variable background. This could be due to a changing light intensity (e.g. sunlight) potentially caused by physical events such as

reflection from surfaces in motion like water, or moving objects (or their shadows) as foliage blowing in the wind. Figure 4(b) shows how the noise affects the silhouettes and blurs the resulting GEI, as compared to the clean cycle in 4(a).

Static occlusion: It represents a background object that is located in a plane nearer to the camera than that of the subject of interest. In experiments, a weed silhouette has been sequentially added to human silhouettes representing a stationary element of the scene. The use of this kind of stuff was motivated by the fact that it affects mostly the lower part of the body, which is expected to contain highly relevant gait information. Figure 4(c) shows how weed is introduced along the frames of a cycle, and the damage produced in the lower part of the GEI.

Dynamic occlusion: It represents a foreground object which follows a trajectory that crosses that of the subject of interest. In a similar way to static occlusions, a car silhouette is added to human silhouettes simulating an object in motion in the scene. At this point, it is worth remarking the difference between both types of occlusion: static objects belong to the background, causing missing body parts in human silhouettes, while dynamic objects are segmented as foreground, distorting the shape of human silhouettes or generating several blobs. Figure 4(d) illustrates the superposition of silhouettes, spotting considerably the resulting GEI.

As can be appreciated in Fig. 4, all silhouettes (disregarding their nature) appear centered in boxes of the same size, which represent the region of interest. They could be the ideal output of a segmentation process based on a smart tracker able to reliably estimate the silhouette’s location (e.g. its centroid) on each video frame. This process could rely on probabilistic models of gait pose transitions over time [23], or on detecting and tracking isolated body parts (e.g. the head) [26, 32]. Once the silhouette’s centroid is located, a fixed-size window can be used to bound the silhouette. People detection and tracking are active

research areas, but they are out of the scope of this work. Thus, it is assumed that there exists such a smart tracker feeding the gait representation stage. In this work, the size of the boxes has been set to 64×44 px.

5.3 Gait representation

5.3.1 Methods based on averaging silhouette data

As suggested in Sect. 4.2, the computation of a measure of the deviation of some feature value from its expectation by the $v(\cdot)$ function should consider some contextual information, which may depend on the gait representation method. A brief description of each method, along with the proposed $v(\cdot)$, is given next:

- *Gait Energy Image* (GEI) [6]. It is a widely-known model-free method for gait representation. It computes an average image (GEI) from a set of normalized binary silhouettes, which reflects the shape and dynamic of the body parts. Pixels of the resulting GEI are used as features. Given a pixel (feature) $g_i(k)$, let $r_{i,h}(k)$ be a h^2 -dimensional vector composed of the pixels that belong to the $h \times h$ region in the GEI centered at the pixel $g_i(k)$. Let $r_{1:i-1,h}(k)$ be the corresponding vector gathered from the pattern $g_{1:i-1}(k)$. Then $v(g_i(k), g_{1:i-1}(k))$ is defined as the normalized Euclidean distance between $r_{i,h}(k)$ and $r_{1:i-1,h}(k)$. The normalization by the number of pixels allows a direct comparison between (full) internal regions and (smaller) border regions. In the experiments, a 5 by 5 neighborhood was used.
- *Gradient Histogram Energy Image* (GHEI) [7]. It is a model-free method that computes normalized histograms of oriented gradients (HOGs) on binary silhouettes, which are then averaged to obtain a vector of mean HOG descriptors (GHEI). Since each descriptor condenses contextual information by itself, the $v(\cdot)$ function keeps its implicit form, i.e. $v(g_i(k), g_{1:i-1}(k)) = |g_i(k) - g_{1:i-1}(k)|$. GHEI can also be obtained

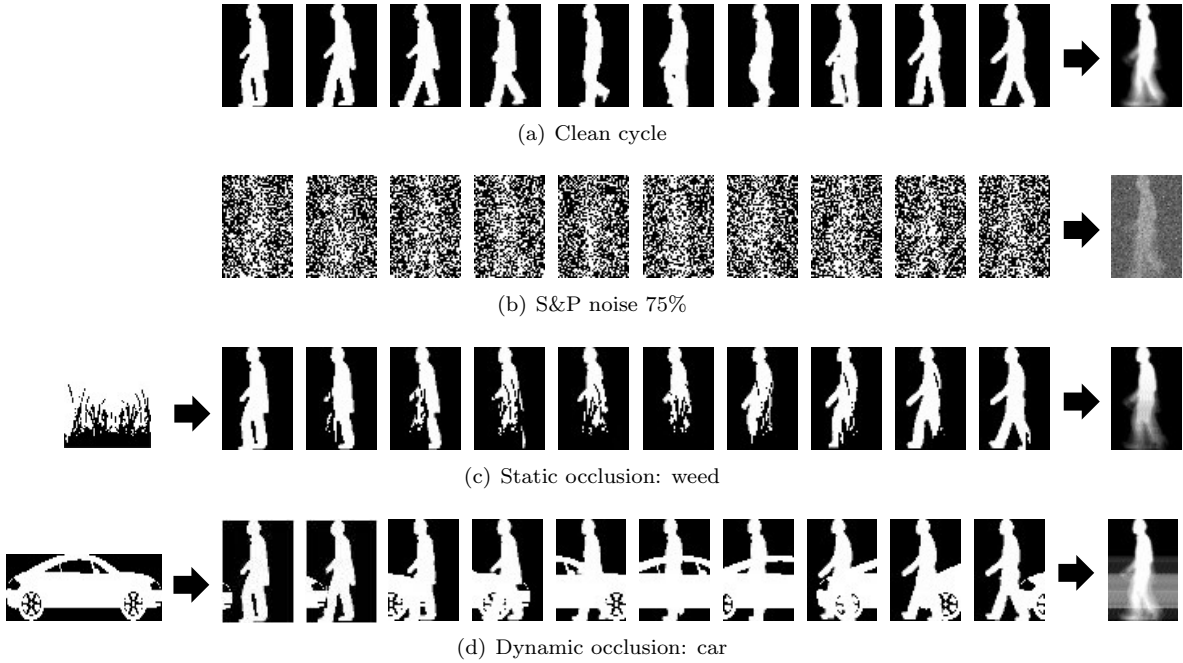


Figure 4: Gait cycle (through a number of key frames) along with its GEI including: a) no defects; b) 75% of S&P noise; c) a static occlusion (weed); and d) a dynamic occlusion (car).

from color images but, for the purpose of comparison with GEI, in this work only binary silhouette images have been used.

- *Gradient histograms from a GEI (GEI+HOG)* [27, 29]. It follows an opposite strategy to that of GHEI: binary silhouettes are first averaged to obtain a GEI; then, HOG descriptors are computed on that GEI. Since the gait pattern is encoded as in GHEI, the $v(\cdot)$ function keeps the same implicit definition.

5.3.2 Operation modes

Given a gait representation method based on averaging silhouette descriptions, three operation modes have been defined: *simple mean*, *defect exclusion*, and *robust mean*. The first refers to the standard operation of the method by the Eq. (10). It is intended to play the role of a somewhat weak baseline, because it handles defective silhouette descriptions without

any filtering technique. The second mode, denoted as *defect exclusion*, consists in solving Eq. (10), but excluding those cycles g_i that are known to be contaminated. This process simulates the existence of an ideal filtering technique, like an oracle, which is able to detect corrupted cycles so that they can be avoided. Since this mode makes the most of the a priori information about what cycles are defective, it is deemed as a very demanding benchmark. Finally, *robust mean* concerns the method introduced in Eq. (13). This mode directly acts on all cycles, without using a priori information on their quality.

5.4 Parameter estimation and learning

Next, the two learning tasks embedded in the methodology are addressed. They learn from the Training subset, which comprises only clean sequences of subjects different to Test people.

5.4.1 Logistic-based weight function

The robust characterization of gait sequences introduced in Sect. 4.1 requires the adjustment of the parameters of the weight function. Concerning the logistic-based weight function $w_{log}(x)$ proposed in Eq. (15), parameters s , t , and ϵ need to be appointed for each combination of gait representation method g and database. The parameter s , that defines the steepness of the curve, has been manually set to ensure a behavior similar to the step function. As this is ultimately determined by the distribution of x , with $x = v(\cdot)$ depending on the representation method, s was set to 1 for GEI and to 1000 for GHEI and GEI+HoG representations².

Since the slope of $w_{log}(x)$ has been adjusted to perform close to a step function, the parameter t can be understood as a threshold at which the logistic curve drops, beyond which x values are considered outliers (see Fig. 2). That is, for almost all $x < t$, $w_{log}(x)$ reaches its maximum value 1, whereas for almost all $x > t$, $w_{log}(x)$ outputs its minimum ϵ . Thus, t should separate genuine deviations from the irregular ones. The parameter t is proposed to be tuned from some distribution of $v(\cdot)$ computed over an independent set of clean gait sequences. This process is detailed below:

1. Let Y be an independent set of *clean* gait sequences with at least four cycles each one. In this work, the set Y is represented by the Training subset.
2. Given a particular gait representation, let $D_i = \{v(g_i(k), g_{1:i-1}(k))\}$, $i > 1$, be the set of all deviations from the i^{th} cycle of all clean sequences $y \in Y$, including all features k , $1 \leq k \leq d$. Thus, D_i is expected to contain only genuine deviations.
3. Let H_i be a histogram that condenses D_i . Figure 5(a) shows H_4 built from using GEI on a Training subset of the USF database. As can be

²In the case of GEI, $v(\cdot)$ computes normalized Euclidean distances between vectors of gray scale pixel values, thus $v(\cdot) \in [0, 255]$, while in GHEI and GEI+HoG, $v(\cdot)$ measures absolute differences between features normalized in $[0, 1]$.

seen in 5(b), H_4 roughly approach an exponential distribution.

4. Formally, an exponential distribution is defined by the following probability density function (PDF):

$$f(x|\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (16)$$

where the maximum likelihood estimate of λ is $\hat{\lambda} = 1/\bar{x}$, with \bar{x} being the sample mean.

5. Given an exponential PDF $f(x|\lambda)$, the Tukey criteria [28] determines a limit l beyond which data can be interpreted as outliers. This criteria establishes the following formula to compute l :

$$l = Q3 + 1.5|Q3 - Q1| = \frac{\ln(4)}{\lambda} + 1.5 \frac{\ln(3)}{\lambda} \quad (17)$$

with $Q1$ and $Q3$ being the first and the third quartiles, respectively. The amount of data higher than l (anomalies) is expected to account for 4.81%.

6. Assuming D_i follows an exponential distribution, the Tukey criteria is used to estimate t , i.e. $t = l$.
7. The parameter t was estimated from D_4 , to better exploit sequences with at least four cycles.

With regard to GEI, $D_i = \{v(g_i(k), g_{1:i-1}(k))\}$ consisted of only those $v(\cdot)$ deviations in which either $g_i(k)$ or $g_{1:i-1}(k)$, or both, was a foreground pixel. This is expected to lead to a more reliable estimation of λ , because it prevents a bias to zero deviations caused by background pixels. Conversely, this strategy cannot be applied to GHEI and GEI+HOG since their features are not directly associated to pixel locations, and all $v(\cdot)$ were taken into account to build D_i .

Finally, the parameter ϵ was set to 0.1, a minimum weight value that can be considered as reasonable.

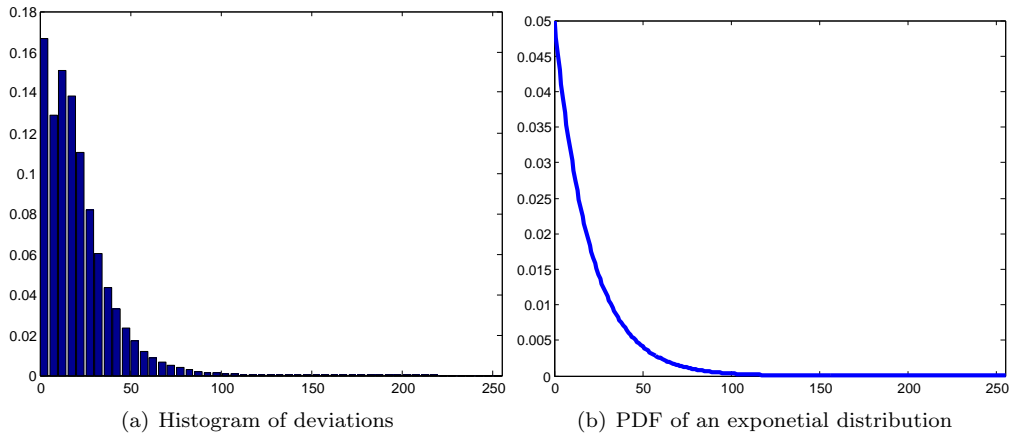


Figure 5: a) Histogram H_4 of deviations based on GEI computed on a training subset of USF sequences, and b) probability density function of an exponential distribution with λ estimated from H_4 data.

5.5 Classification and performance evaluation

5.5.1 Ranking-based classification

Ranking-based classification in Fig. 3 refers to the use of a scoring function to perform template matching between a probe sample and all available gallery samples.

The Ranking Support Vector Machine (RankSVM) [19] was chosen because its ability to suitably manage changing conditions between training and test data. In this work, training is based on clean gait sequences, while test involves defective sequences simulating uncontrolled scenarios. RankSVM learns from the Training subset how to rate gait features within a scoring (dissimilarity) function, so as to reward features that are invariant under intra-class changes. For comparison purposes, the traditional 1-Nearest Neighbor classifier (1-NN) has also been considered. Since 1NN does not require modeling, the Training subset was not used.

5.5.2 Cumulative performance curves

Let us denote as *study* an experimental design that combines a gait database, a data partition that fits the *Training+(Gallery+Probe)* scheme, a represen-

tation method g , a strategy to inject defects, an operation mode, and a classifier. Given a particular study, a recognition task can be defined for each cycle i , $1 \leq i \leq m$, with m being the number of cycles of the shortest probe sequence. At cycle i , the classifier uses the gallery sequences represented by g , to score and rank probe samples characterized by $g_{1:i}$. Then, a series of m recognition results (i, acc_i) is obtained and represented as a *cumulative performance curve* (CPC), where acc_i denotes the classification accuracy over $g_{1:i}$.

A CPC allows for continuous monitoring of method performance along the cycles. That is, it is possible to know the impact of adding both defective and clean cycles at some cycle i , when building the pattern $g_{1:i}$. Thus, CPCs can be considered as a suitable tool to benchmark the robust approach proposed.

5.5.3 Statistical analysis of results

This section introduces three experimental settings as regards the amount of defect injected, provided that all probe sequences consist of at least five cycles³:

Easy setting Only the cycle in the middle (the third cycle) is affected by a particular type of

³This is a particular case of having at least four cycles.

defect (S&P noise, static occlusion, dynamic occlusion). Thus, one fifth of the sequence is contaminated.

Hard setting The three cycles in the middle (the second, third and fourth) are affected by the three types of defects (one each) in a random way. In this case, three fifths of the sequence are contaminated.

Neutral setting No cycle is affected. Thus, recognition tasks perform on original clean sequences.

The easy setting comprises 108 studies (each one depicted by a CPC) that result from combining the two databases (OU-ISIR, USF), the two classifiers (1-NN, RankSVM), the three gait representations (GEI, GHEI, GEI+HOG), the three operation modes (simple mean, defect exclusion, robust mean), and the three defect scenarios (S&P noise, static occlusion, dynamic occlusion). Unlike the easy setting, the hard setting considers only one defect scenario (mixtures of defects), thus entailing 36 studies. Finally, the neutral setting defines 24 conventional studies from clean sequences, which arise from the combination of the two databases, two classifiers, three gait representations, and *two* operation modes (defect exclusion makes no sense).

The CPC of each study involving defective sequences was sampled twice, the first one right after summing the corrupted gait cycle(s) (early sampling), and the second one at the end of the sequence analysis (final sampling). The former occurs in the third and the fourth cycles in the easy and the hard settings, respectively. It is aimed to assess the immediate impact of noisy cycles, while final sampling allows for assessing the final pattern.

In order to ease the comparison of such a large number of results, they were conveniently grouped considering three criteria defined in Sect. 6.1.1, 6.1.2, and 6.1.3, respectively. Each of them builds equal-sized series of results of similar methods, which are pairwise compared using the Wilcoxon’s Signed Rank test [31]. For each pair, Wilcoxon’s null hypothesis assumes that both methods perform equally. Then, evidence is searched for in the data to reject the null

hypothesis, thus establishing the superiority of one method over the other.

6 Experiments

This section has been structured into two major areas of analysis. The former involves a statistical study based on the Wilcoxon’s Signed Rank test, which was conducted using the KEEL software [1]. The second area focuses on performance analysis based on CPCs.

The statistical analysis follows three perspectives. First, operation modes are compared under each type of defect, considering all results from combining the two databases, the two classifiers, and the three gait representations (Sect. 6.1.1). The second perspective entails again a comparison between operation modes, but on each database separately (Sect. 6.1.2). It is intended to be a more general view, because each series comprises results from both clean and defective scenarios. Finally, gait representations are compared on each database, taking into account results from the joint use of the two classifiers and the three operation modes on both clean and defective sequences. (Sect. 6.1.3).

The performance analysis focuses on GEI results, since it is probably the most popular gait representation method. A total of 56 studies or CPCs are examined, 36 of which belong to the easy setting, 12 to the hard setting, and 8 to the neutral setting.

Each result involved in the analyses is an average computed over five repetitions of the related experiment with different random data partitions according to Sect. 5.1.1. Finally, as a reminder, the structural parameters introduced in Sect. 5 are given below:

- level of the Salt & Pepper noise, $\alpha = 75\%$ (Sect. 5.2)
- size of silhouette boxes, 64×44 px. (Sect. 5.2)
- size of the neighborhood used to compute $v(\cdot)$ in GEI, 5×5 px. (Sect. 5.3.1)
- the weight function $w_{log}(\cdot)$ (Sect. 4.2)
- steepness of the weight function $w_{log}(\cdot)$, $s = 1$ for GEI, $s = 1000$ for GHEI and GEI+HoG (Sect. 5.4.1)

- domain value at which $w_{log}(\cdot)$ drops, t is estimated from the Training subset (Sect. 5.4.1)
- minimum of $w_{log}(\cdot)$, $\epsilon = 0.1$ (Sect. 5.4.1)

6.1 Statistical analysis

6.1.1 Defect-conditional analysis of operation modes

Given a *defect scenario* and a sampling event, this analysis consists in pairwise statistical comparisons between the three operation modes (simple mean, robust mean, and defect exclusion). Recognition results were grouped into three series, each one corresponding to an operation mode. Overall, eight groups were built:

- *Easy setting*: Six (3 defective scenarios \times 2 sampling events) three-series groups, where each series comprises 12 recognition results (2 databases \times 2 classifiers \times 3 gait representations).
- *Hard setting*: Two (1 defective scenario \times 2 sampling events) three-series groups, where each series comprises 12 recognition results (2 databases \times 2 classifiers \times 3 gait representations).

Figure 6 shows the results of the Wilcoxon’s test applied to the eight groups. A first relevant finding is that robust mean performed better than or equal to simple mean in all cases, being statistically better in five out of the eight groups with a confidence of 95%. This proves that robust mean is able to mitigate the negative impact of faulty regions, while takes advantage of clean parts. When focusing on each setting, defect exclusion outperformed both types of means in the easy setting (1/5 sequence corrupted), while robust mean was the best mode in the hard setting (3/5 sequence corrupted). The latter observation demonstrates again that robust mean leverages profitable pieces of information from faulty cycles to construct better gait patterns, whereas defect exclusion simply discards these cycles.

These results suggest that robust mean is the best choice when the contaminated portion of a gait sequence is high, while defect exclusion (ideal bench-

mark) leads to the best performance under low defect rates.

6.1.2 Database-conditional analysis of operation modes

Given a *database* and a sampling event, this analysis consists in pairwise statistical comparisons between the three operation modes. Results were grouped into three series, each one corresponding to an operation mode. Overall, eight *new* groups of results were built⁴:

- *Easy setting*: Four (2 databases \times 2 sampling events) three-series groups, where each series comprises 24 recognition results [2 classifiers \times 3 gait representations \times (3 defect + 1 clean scenarios)].
- *Hard setting*: Four (2 databases \times 2 sampling events) three-series groups, where each series comprises 12 recognition results [2 classifiers \times 3 gait representations \times (1 mixture of defects + 1 clean scenarios)].

Figure 7 summarizes the outcomes of the Wilcoxon’s test on the eight groups. As in the first analysis, robust mean always performs at least equal to simple mean, statistically overcoming it in OU-ISIR at early sampling. Thus, the robust method allows for eliciting faster and no less reliable identification hypotheses, as compared to the simple mean.

By examining each setting, defect exclusion proved to be the best mode in the easy setting, while the robust approach outperformed defect exclusion in the hard setting in both databases with a confidence of 95%. This effect is clear in the USF database, which contains low-quality silhouettes captured outdoors. As in Sect. 6.1.1, robust gait patterns significantly benefited from high-quality pieces of information retained from defective cycles, mainly in the more realistic scenario proposed in USF. However, it is imperceptible at the final sampling in OU-ISIR, composed of high-quality indoor sequences. It means that using only two clean cycles (the first and the fifth) from

⁴Note that each series includes results from both defective and clean scenarios together.

		Easy setting						Hard setting								
		S&P			Static			Dynamic			Mixture of defects					
early sampling	SM	-	o	o	SM	-	DE	RM	SM	-	o	o	SM	-	o	o
	DE	•	-	•	DE	-	-	-	DE	•	-	•	DE	•	-	o
	RM	•	o	-	RM	-	-	-	RM	•	o	-	RM	•	•	-
final sampling	SM	-	o	o	SM	-	DE	RM	SM	-	o	o	SM	-	o	o
	DE	•	-	•	DE	•	-	•	DE	•	-	•	DE	•	-	o
	RM	•	o	-	RM	-	o	-	RM	-	o	-	RM	•	-	-

Figure 6: Summary of the Wilcoxon’s test for pairwise comparisons of operation modes (SM: simple mean, DE: defect exclusion, RM: robust mean), under each combination of *defect* and sampling event. The symbol “•” (“o”) indicates that the model in the row (column) significantly outperforms that in the column (row). Results below the main diagonal are supported by a level of confidence of $\alpha = 0.95$, while results above that diagonal, by a level of confidence of $\alpha = 0.90$.

the OU-ISIR sequences, is statistically as reliable as building robust patterns.

In brief, results on the more challenging tasks prove again that it is better to robustly consider all cycles instead of blindly adding them (simple mean) or simply discarding the affected ones (defect exclusion).

6.1.3 Database-conditional analysis of gait representation methods

Given a database and a sampling event, the third analysis compares gait representation methods (GEI, GHEI, and GEI+HOG) by pairs. Results were grouped into three series, each one corresponding to a representation method. Overall, eight *new* groups were built:

- *Easy setting*: Four (2 databases \times 2 sampling events) three-series groups, where each series comprises 24 recognition results [2 classifiers \times 3 operation modes \times (3 defect + 1 clean scenarios)].
- *Hard setting*: Four (2 databases \times 2 sampling events) three-series groups, where each series

comprises 12 recognition results [2 classifiers \times 3 operation modes \times (1 mixture of defects + 1 clean scenarios)].

Figure 8 summarizes the results of the Wilcoxon’s test on the eight groups. A first observation focuses on the fact that no differences exist between the early and the final samplings. Thus, the analysis between gait representations does not seem to depend on the CPC point where the recognition accuracy is measured.

With regard to the effectiveness of each representation, GEI outperforms GHEI and GEI+HOG with a level of significance up to 95% in OU-ISIR. However, exactly the opposite occurs in USF, where the HOG-based methods beat GEI, evidencing the dependence of method behaviour on the quality of gait samples. Disregarding the type of defect, the well-defined silhouettes of OU-ISIR allowed GEI to construct more reliable gait patterns from pixel values, than those based on HOG descriptors built by GHEI and GEI+HOG. On the contrary, when managing the low-quality silhouettes of USF, GEI led to poorer gait representations as compared to the HOG-based

		Easy setting				Hard setting				
		OU-ISIR		USF		OU-ISIR		USF		
early sampling	SM	-	○	○	SM	-	○	SM	-	○
	DE	●	-	●	DE	●	-	DE	-	○
	RM	●	○	-	RM	-	-	RM	●	-
final sampling	SM	-	○	○	SM	-	○	SM	-	○
	DE	●	-	●	DE	●	-	DE	-	○
	RM	-	○	-	RM	-	○	RM	●	-

Figure 7: Summary of the Wilcoxon’s test for pairwise comparisons of operation modes (SM: simple mean, DE: defect exclusion, RM: robust mean), under each combination of *database* and sampling event. The symbol “●” (“○”) indicates that the model in the row (column) significantly outperforms that in the column (row). Results below the main diagonal are supported by a level of confidence of $\alpha = 0.95$, while results above that diagonal, by a level of confidence of $\alpha = 0.90$.

patterns. Within this context, GEI+HOG significantly outperformed GHEI when dealing with USF in the easy setting, although these differences vanished in the hard setting.

Considering all of the above, GEI representation seems to be more appropriate when high-quality silhouettes are available, due to its higher precision at encoding spatial information. Otherwise, HOG-based methods can be a better alternative under blurry or less sharpened silhouettes.

6.2 Performance analysis

This section conducts a complementary analysis on gait recognition results which, due to the total amount of experiments, involves only those studies (or CPCs) that are based on GEI⁵. The analysis consists of three parts, each one corresponding to a particular setting: neutral, easy, and hard. Within each setting, CPCs were grouped into diagrams, where each diagram comprises all results that arise from a

⁵Performance curves from the other two methods are pretty similar to those from GEI showed in this section.

particular combination of a database and a clean or a defect scenario.

Two preliminary remarks, common to the three settings, are worth making at this point: 1) each RankSVM accuracy is higher than the comparable result of 1-NN due to the learning power of the former; 2) each OU-ISIR accuracy is higher than the comparable result from USF because of the lower quality of USF silhouettes.

Figure 9 accommodates two diagrams, one for each database, with classification results from the neutral setting (clean sequences). Each diagram includes four CPCs resulting from combining the two classifiers (1NN, RankSVM) and two operation modes (simple mean, robust mean). Note that defect exclusion makes no sense within a neutral setting. The most interesting point is that the simple and the robust means performed roughly equal for each pair of classifier and database. This proves that robust statistics can be effective even on clean sequences, generating gait signatures as reliable as those built by simply averaging.

Results from the easy setting are shown in Fig. 10,

		Easy setting				Hard setting				
		OU-ISIR		USF		OU-ISIR		USF		
	early sampling	G	GH	G+H	G	GH	G+H	G	GH	G+H
final sampling	G	-	•	•	-	◦	◦	-	◦	◦
	GH	◦	-	◦	•	-	◦	•	-	◦
early sampling	G	-	•	•	-	◦	◦	-	◦	◦
	GH	◦	-	◦	•	-	◦	•	-	◦
final sampling	G	-	•	•	-	◦	◦	-	◦	◦
	GH	◦	-	◦	•	-	◦	•	-	◦
early sampling	G	-	•	•	-	◦	◦	-	◦	◦
	GH	◦	-	◦	•	-	◦	•	-	◦
final sampling	G	-	•	•	-	◦	◦	-	◦	◦
	GH	◦	-	◦	•	-	◦	•	-	◦

Figure 8: Summary of the Wilcoxon’s test for pairwise comparisons of representation methods (G: GEI; GH: GHEI; G+H: GEI+HOG), under each combination of database and sampling event. The symbol “•” (“◦”) indicates that the model in the row (column) significantly outperforms that in the column (row). Results below the main diagonal are supported by a level of confidence of $\alpha = 0.95$, while results above that diagonal, by a level of confidence of $\alpha = 0.90$.

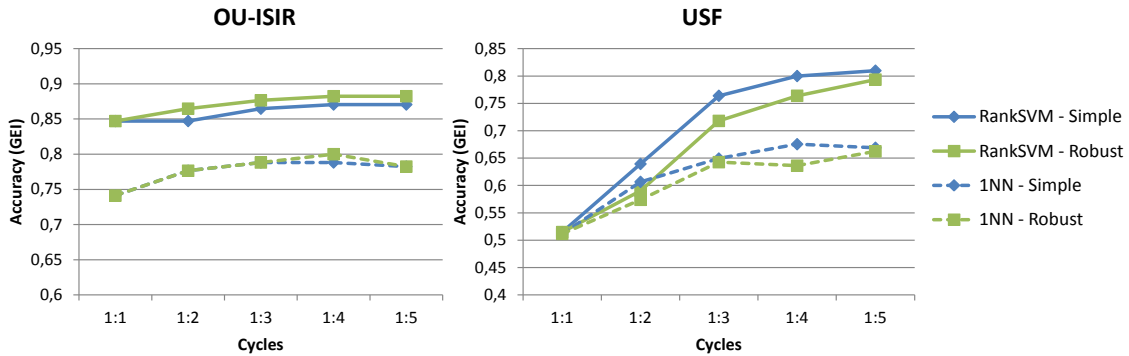


Figure 9: Cumulative performance curves based on GEI patterns built from the OU-ISIR (left) and USF (right) databases within a neutral setting, *i.e.*, without artificial defect injection.

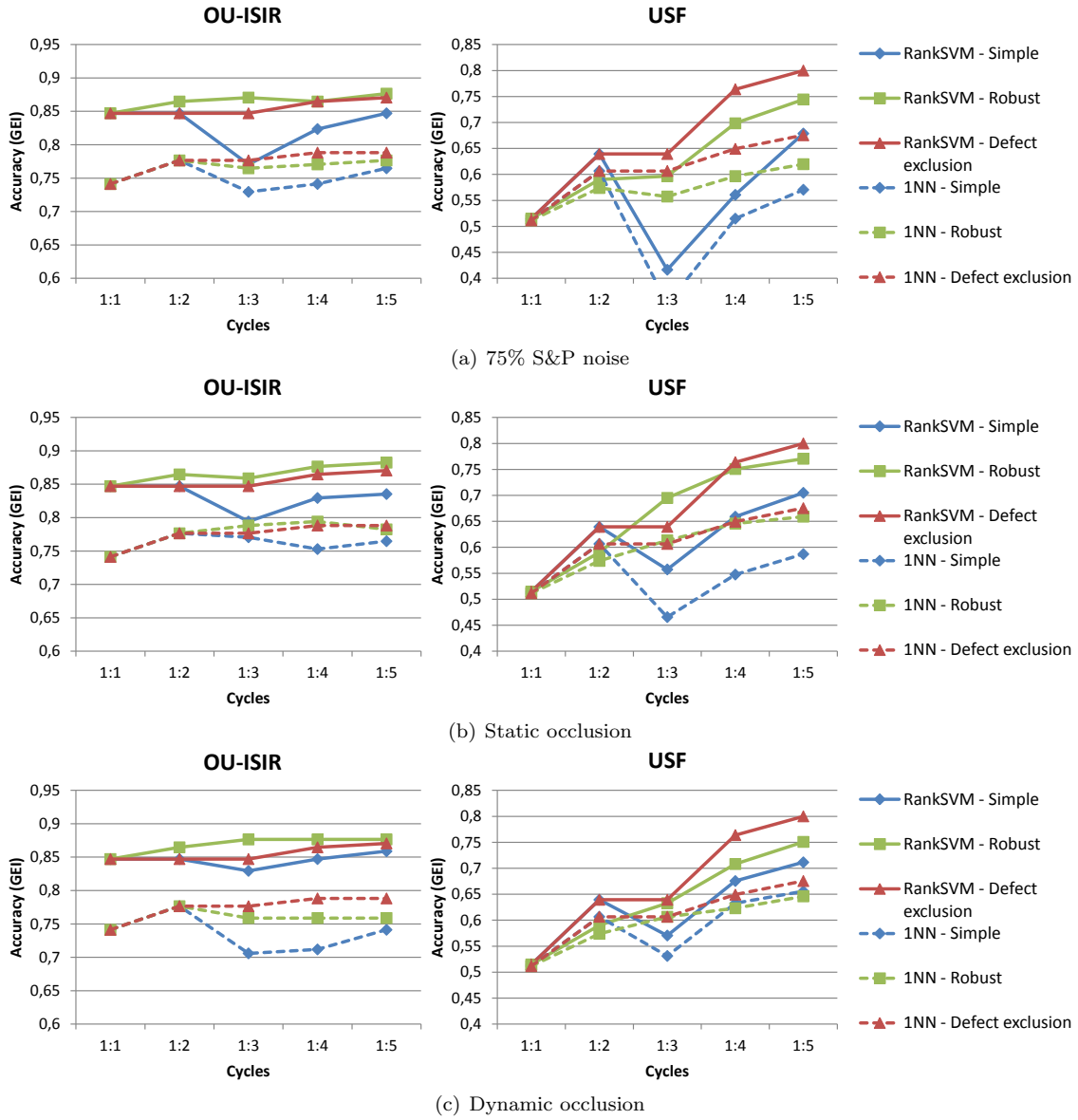


Figure 10: Cumulative performance curves based on GEI patterns built from the OU-ISIR (left) and USF (right) sequences affected by the three defect scenarios within the easy setting: a) 75% of salt & pepper noise; b) static occlusion (weed); and c) dynamic occlusion (car).

where six diagrams are arranged (3 defect scenarios \times 2 databases), with each diagram comprising six CPCs (3 operation modes \times 2 classifiers). As can be observed, a notable drop in CPCs based on simple mean took place when defects (no matter which) were injected in the third cycle, although these curves recovered slowly when the last two clean cycles (the fourth and the fifth) were added. This fall was specially patent in the USF database, since gait representations are based on low-quality USF silhouettes, more sensitive to defect insertion than those from OU-ISIR. Conversely, CPCs that came from the robust characterization showed a high stability or an upward trajectory over all cycles in all diagrams, occasionally surpassing curves based on defect exclusion (ideal benchmark).

Finally, Fig. 11 depicts the CPCs of the hard setting, where the three types of defects are injected at random into the second, third and fourth cycles, respectively. Only two diagrams were needed, one for each database. As expected, gait sequences affected by such a mixture of defects led to more pronounced falls in the accuracy of methods based on simple mean as compared to the easy setting, with these accuracies being notably lower than their comparable results from the robust mean. Again, it was particularly noticeable in USF, where accuracies decreased below 0.4 in the first affected cycle, making it impossible for them to reach again their original performances. Meanwhile, curves from the robust method kept growing along the three consecutive affected cycles, being even generally better than those derived from defect exclusion. This strengthens conclusions drawn in the statistical analysis: the robust approach is able to provide faster and more reliable gait representations than conventional averaging methods, specially when gait sequences are severely corrupted.

7 Conclusions and future work

This work introduces a weighted averaging method to build reliable gait patterns from silhouette descriptions heavily affected by major defects. It is based on a statistical approach called *robust statistics*, which assumes data follows an approximate normal distri-

bution with heavy tails of atypical samples (outliers). The proposed robust method is able to behave nearly as the simple mean when there are no outliers, while it underweights largely deviated samples (possibly outliers) otherwise.

The robust method was compared to two other modes of operating on silhouette descriptions, simple mean and defect exclusion, as regards their discriminant capabilities under a large number of biometric identification studies based on clean and defective gait sequences. Each study was designed from combining a gait representation method (GEI, GHEI, GEI+HOG), defects used to corrupt gait sequences (salt and pepper noise, static occlusions, dynamic occlusions), a strategy to inject defects (single, mixtures), an operation mode on silhouette descriptions, a gait database (USF, OU-ISIR), and a classifier (RankSVM, 1-NN).

Result assessment was carried out from two perspectives. First, the Wilcoxon’s Signed Rank test was used for a qualitative pairwise comparison of operation modes and gait representation methods. The robust approach proved to be generally more reliable than the simple mean, as well as the best choice in the more defective scenarios. That is, when the contaminated portion of a gait sequence is high, it is better to robustly consider all cycles instead of averaging them indiscriminately (weak baseline) or discarding the affected cycles (strong benchmark). Second, a complementary analysis focused on gait recognition results based on GEI supported the previous conclusion: the robust method is able to provide faster and more reliable gait representations than conventional averaging methods, specially when gait sequences are severely contaminated.

Next, some promising directions for future research are suggested. First, to keep the model complexity to a minimum, a single weight function was used in the robust method, whose parameters were adjusted from the distribution of all feature deviations together. However, it is easy to see that features (e.g. GEI pixels) do not share a common distribution pattern, nor feature deviations. That is, universal parameters inferred from a general distribution could not be able to model optimal feature-dependent criteria to separate genuine from large feature deviations (from

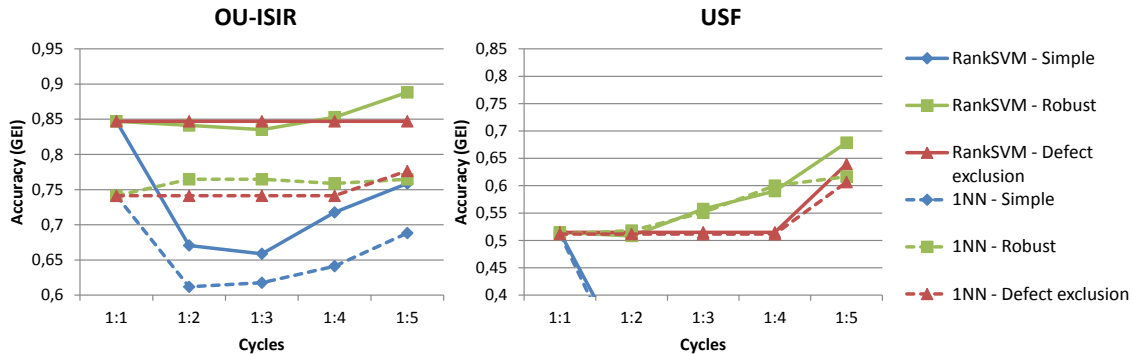


Figure 11: Cumulative performance curves based on GEI patterns built from the OU-ISIR (left) and USF (right) sequences affected by random mixtures of defects within the hard setting.

their feature means). Instead, it can be expected that feature-dependent weight functions will more accurately fit the deviation distributions by features, allowing a more reliable outlier detection. Second, due to the spatio-temporal nature of a gait sequence, the robust method has been implemented as a recurrence over cycles. This formulation assumes a clean first gait cycle, so that it can lead to a well-behaved first estimate. However, more general robust approaches could be devised, which can rely on any gait cycle that fits a normality model, or on a synthetic cycle built by concatenating clean silhouettes chosen from different partially-affected cycles. Third, some structural parameters have been manually set to widely accepted values. Thus, it would be interesting to design new experiments to explore optimality in parameter values, as well as their interactions. Finally, in order to facilitate the interpretation of results, defect injection was restricted to cycles. However, more realistic scenarios can be generated if defects were freely added to any subsequence, disregarding cycle limits.

References

- [1] Alcalá-Fdez, J., Fernandez, A., Luengo, J., Derac, J., Garca, S., Snchez, L., Herrera, F.: Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *J MULT-VALUED LOG S* **17**, 255–287 (2011)
- [2] Barnett, V., Lewis, T.: *Outliers in statistical data*, vol. 3. Wiley New York (1994)
- [3] Black, M., Rangarajan, A.: On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *INT J COMPUT VISION* **19**(1), 57–91 (1996)
- [4] Chen, C., Liang, J., Zhao, H., Hu, H., Tian, J.: Frame difference energy image for gait recognition with incomplete silhouettes. *PATTERN RECOGN LETT* **30**(11), 977–984 (2009)
- [5] Gross, R., Shi, J.: The CMU Motion of Body (MoBo) database. Tech. Rep. CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (2001)
- [6] Han, J., Bhanu, B.: Individual recognition using Gait Energy Image. *IEEE T PATTERN ANAL* **28**(2), 316–322 (2006)
- [7] Hofmann, M., Rigoll, G.: Exploiting gradient histograms for gait-based person identification. In: *IEEE IMAGE PROC*, pp. 4171–4175 (2013)
- [8] Hofmann, M., Sural, S., Rigoll, G.: Gait recognition in presence of occlusion: A new dataset and baseline algorithms. In: *WSCG*, pp. 99–104 (2011)

- [9] Hofmann, M., Wolf, D., Rigoll, G.: Identification and reconstruction of complete gait cycles for person identification in crowded scenes. In: VISAPP, pp. 594–597 (2011)
- [10] Huber, P.: Robust statistics. Wiley New York (1981)
- [11] Jin, W., She, M., Nahavandi, S., Kouzani, A.: A review of vision-based gait recognition methods for human identification. In: DICTA, pp. 320–327 (2010)
- [12] Lee, L., Dalley, G., Tieu, K.: Learning pedestrian models for silhouette refinement. In: IEEE I CONF COMP VIS, vol. 1, pp. 663–670 (2003)
- [13] Liang, W., Tieniu, T., Huazhong, N., Weiming, H.: Silhouette analysis-based gait recognition for human identification. IEEE T PATTERN ANAL **25**(12), 1505–1518 (2003)
- [14] Liu, J., Zheng, N., Xiong, L.: Silhouette quality quantification for gait sequence analysis and recognition. SIGNAL PROCESS **89**(7), 1417–1427 (2009)
- [15] Liu, Z., Sarkar, S.: Effect of silhouette quality on hard problems in gait recognition. IEEE T SYST MAN CY B **35**(2), 170–183 (2005)
- [16] Makiyama, Y., Mannami, H., Tsuji, A., Hosain, M., Sugiura, K., Mori, A., Yagi, Y.: The OU-ISIR Gait Database comprising the treadmill dataset. IPSJ Transactions on Computer Vision and Applications **4**, 53–62 (2012)
- [17] Maronna, R.A., Martin, D.R., Yohai, V.J.: Robust Statistics: Theory and Methods. John Wiley and Sons (2006)
- [18] Martín-Félez, R., Ortells, J., Mollineda, R.A.: Exploring the effects of video length on gait recognition. In: INT C PATT RECOG, pp. 3411–3414 (2012)
- [19] Martín-Félez, R., Xiang, T.: Gait recognition by ranking. In: ECCV, LECT NOTES COMPUT SC, vol. 7572, pp. 328–341 (2012)
- [20] Ortells, J., Martín-Félez, R., Mollineda, R.A.: A complexity measure of gait perception. In: Pattern Recognition and Image Analysis, LECT NOTES COMPUT SC, vol. 7887, pp. 492–499 (2013)
- [21] Rey, W.: Introduction to Robust and Quasi-Robust Statistical Methods. Universitext. Springer Berlin Heidelberg (1983)
- [22] Roy, A., Chattopadhyay, P., Sural, S., Mukherjee, J., Rigoll, G.: Modelling, synthesis and characterization of occlusion in videos. IET COMPUT VIS **9**(6), 821–830 (2015)
- [23] Roy, A., Sural, S., Mukherjee, J., Rigoll, G.: Occlusion detection and gait silhouette reconstruction from degraded scenes. SIGNAL IMAGE VIDEO P **5**(4), 415–430 (2011)
- [24] Sarkar, S., Phillips, P., Liu, Z., Vega, I., Grother, P., Bowyer, K.: The HumanId Gait Challenge problem: Data sets, performance and analysis. IEEE T PATTERN ANAL **27**(2), 162–177 (2005)
- [25] Schafer, R.C.: Clinical biomechanics: musculoskeletal actions and reactions. Williams & Wilkins (1987)
- [26] Shu, G., Dehghan, A., Oreifej, O., Hand, E., Shah, M.: Part-based multiple-person tracking with partial occlusion handling. In: PROC CVPR IEEE, pp. 1815–1821 (2012)
- [27] Sivapalan, S., Chen, D., Denman, S., Sridharan, S., Fookes, C.: Histogram of weighted local directions for gait recognition. In: Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Conference on, pp. 125–130 (2013)
- [28] Tukey, J.W.: Exploratory Data Analysis. Addison-Wesley (1977)
- [29] Whytock, T., Belyaev, A., Robertson, N.: Improving robustness and precision in GEI + HOG action recognition. In: Advances in Visual Computing, LECT NOTES COMPUT SC, vol. 8033, pp. 119–128 (2013)

- [30] Whytock, T., Belyaev, A., Robertson, N.: On covariate factor detection and removal for robust gait recognition. *MACH VISION APPL* **26**(5), 661–674 (2015)
- [31] Wilcoxon, F.: Individual comparisons by ranking methods. *BIOMETRICS BULL* pp. 80–83 (1945)
- [32] Wu, B., Nevatia, R.: Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. *INT J COMPUT VISION* **75**(2), 247–266 (2007)
- [33] Yu, S., Tan, D., Huang, K., Tan, T.: Reducing the effect of noise on human contour in gait recognition. In: *Advances in Biometrics, LECT NOTES COMPUT SC*, vol. 4642, pp. 338–346 (2007)
- [34] Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: *INT C PATT RECOG*, pp. 441–444 (2006)
- [35] Zhao, G., Cui, L., Li, H.: Gait recognition using fractal scale. *PATTERN ANAL APPL* **10**(3), 235–246 (2007)
- [36] Zhou, Z., Prugel-Bennett, A., Damper, R.: A bayesian framework for extracting human gait using strong prior knowledge. *IEEE T PATTERN ANAL* **28**(11), 1738–1752 (2006)