CrossMark

ORIGINAL PAPER

On covariate factor detection and removal for robust gait recognition

Tenika Whytock[1] · Alexander Belyaev[1] · Neil M. Robertson[1]

Abstract We propose a novel bolt-on module capable of boosting the robustness of various single compact 2D gait representations. Gait recognition is negatively influenced by covariate factors including clothing and time which alter the natural gait appearance and motion. Contrary to traditional gait recognition, our bolt-on module remedies this by a dedicated covariate factor detection and removal procedure which we quantitatively and qualitatively evaluate. The fundamental concept of the bolt-on module is founded on exploiting the pixel-wise composition of covariate factors. Results demonstrate how our bolt-on module is a powerful component leading to significant improvements across gait representations and datasets yielding state-of-the-art results.

Keywords Gait recognition · Covariate factor detection · Covariate factor removal

1 Introduction

This paper focusses on the gait recognition problem and specifically boosting the robustness of popular single compact 2D gait representations. To achieve this, contrary to the traditional gait recognition flow, a dedicated covariate factor detection and removal procedure is performed post

✉ Tenika Whytock
tpw3@hw.ac.uk

Alexander Belyaev
a.belyaev@hw.ac.uk

Neil M. Robertson
n.m.robertson@hw.ac.uk

[1] Institute of Sensors, Signals and Systems, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, Scotland, UK

representing gait; the flow therefore becomes gait representation, bolt-on module (covariate factor detection and removal), feature vector extraction, dimensionality reduction and classification. Our contribution is threefold.

- We present our novel bolt-on module which combines extensive covariate factor detection and aggressive removal, and easily fits into existing gait recognition schemes post gait representation construction.
- We define a "typical" gait representation which exposes the underlying pixel-wise composition of covariate factors; this contributes a 15.9 % average state-of-the-art increase across validation dataset covariate factor sequences.
- In particular, we demonstrate that our bolt-on module can generalise over a diverse set single compact 2D gait representations, varying in feature content and natural robustness, to yield an average performance increase of 15.1 %.

The initial concept of our bolt-on module [32] was designed to boost Gait Energy Image [14] robustness and continued to advance state-of-the-art results when validated [31] on a more complex dataset. With these encouraging results, we felt our bolt-on module could serve a greater purpose of dedicated covariate factor detection and removal to enhance the performance of analogous single compact 2D gait representations. This paper therefore describes the quantitative and qualitative evaluation generalising our bolt-on module for deployment on a diverse set of single compact 2D gait representations varying in feature content and natural robustness.

1.1 Motivation

Consider a subject carrying a rucksack (covariate factor). It is natural for the rucksack to be deemed static with respect to the human body, however this is a simplifying assumption. Nat-

ural gait motion cases the rucksack to subsequently undergo motion; therefore pixel-wise confusion occurs between covariate factor motion and natural gait motion.

Currently few gait recognition approaches perform dedicated covariate factor detection and removal; the primary focus instead considers gait representations and features which yield a degree of covariate factor mitigation. Therefore, we are motivated by exploring how covariate factors and their motion affect pixel intensity values. This is achieved by developing an understanding of how covariate factors cause gait appearance and motion alterations. Pixel-wise, covariate factors cause (a) pixel addition, e.g. bags and coats add bulk uniformly or in specific locations about the body, (b) pixel occlusion, e.g. bags occlude trailing arm motion and (c) pixel shifting, e.g. leaning due to compensation for a shifted centre of gravity when carrying a bag.

1.2 Gait recognition

Gait recognition identifies a walking person; walking is standard where alternative actions [34] are less common. The fundamental walking pattern is similar across healthy subjects where a unique gait is achieved through subtle magnitude and timing variations attributed to walking manner and posture [26]; this is established during early medical [27] and psychophysics [9] research. Applications are real world and manifold, where a considerable effort is dedicated to surveillance tasks including crime detection; regardless, gait recognition is a popular computer vision research topic.

Gait recognition is closely related to classical human motion analysis studies, however modern gait recognition surveys reflect current debates [6]. This paper is dedicated to computer vision gait recognition, however alternative means include wearable or floor pressure sensors [13] and acoustic gait recognition [16]. Be aware that gait is capable not only of determining identity, but also age [24] and gender [21], however, these tasks are outside the scope of this paper.

Gait is classed as a behavioural biometric and is desirable over physiological biometrics, e.g. fingerprint, due to capture (1) at distance, (2) low resolution, (3) without consent, (4) without cooperation and (5) unobtrusively; further still, gait is difficult to fake, disguise and forget. Therefore it is imperative for gait recognition to deploy a means to represent gait which is (1) discriminative during high subject numbers (hundred minimum), (2) robust to real-world covariate factors, e.g. clothing, carrying a bag and complex couples thereof and (3) efficient with respect to memory, computational and processing costs.

1.3 Related work

Since early computer vision attempts [8,28], gait recognition has developed significantly and datasets have matured

equally to validate their robustness; this results in numerous implementation debates. We discuss three major debates, (1) model-based versus model-free approaches, (2) the quantity of images utilised to represent gait and (3) existing research deploying covariate factor detection and/or removal techniques.

1.3.1 Model-based versus model-free approaches

Model-based approaches, such as [18,36], utilise human body structure to track or model body segments, e.g. head, legs, arms, extracted via anthropometric data [11,12]. Models yield static (e.g. body segment and stride lengths) and dynamic (e.g. joint angle trajectories) features. Model-free approaches, such as [14,15], conversely disregard human body structure and instead target the appearance and motion of gait. Representations are typically founded on silhouettes given colour and texture are disregarded thus avoiding bias to appearance given its inconsistency over time; popular derivatives include contours, optical flow and skeletons.

Discussion. While model-based approaches are popular due to their view and scale invariant properties, their reliance on anthropometrics encourages sensitivity to body pose and image quality. Conversely, model-free approaches exhibit favourable qualities outweighing model-based approaches, namely (1) insensitivity to image quality and noise, (2) lower computational complexity and (3) handling low resolution images typical to surveillance applications.

1.3.2 Quantity of images utilised to represent gait

A complete gait sequence contains an entire gait cycle, e.g. left-to-left heel striking the ground. Model-based approaches tend to utilise the entire gait sequence which is uneconomical with respect to memory and computational cost. Key frames [7] are selected for their saliency at fixed points during the gait sequence, however this rejects a quantity of information. Single compact 2D gait representations are popular with model-free approaches and are constructed by space- and time-normalising a gait sequence; this enables natural robustness to noise and short term occlusion whilst constructing an economical representation with respect to memory and computational cost. Note that our novel bolt-on module is designed for single compact 2D representations.

Manifold single compact 2D gait representations exist promoting static features (torso), dynamic features (limb motion) [14,17,19,33,35] or a combination thereof [1,2, 33,38]; those containing only dynamic features tend to be naturally robust given their saliency [25] over time compared to only static features. Regardless of representations, misclassification occurs from neglecting the following: (a) covariate factor pixel-wise confusion with natural gait motion and (b)

the degree, severity and uniqueness in which covariate factors affect gait appearance and motion.

1.3.3 Covariate factor detection

Gait recognition strives to mitigate the effects of covariate factors. The majority of approaches achieve covariate factor mitigation as a by-product of robust gait representations or features. However, there are relatively few approaches which explicitly attempt to detect and/or remove the influence of covariate factors; a select few are described. Li et al. [19] specifically attempt to mitigate covariate factors relating to carrying a bag and clothing. Anthropometrics are employed to segment the body into six large regions. By monitoring the pixel distribution within each region, those outside normal ranges are deemed to be influenced by a covariate factor and are therefore disregarded from subsequent processing. This approach neglects the fact that covariate factors may only partially affect a region, thus removing potentially salient regions unhampered by covariate factors. Bashir et al. [1] propose a feature selection mask. Given the saliency of lower limb motion, the body is segmented into an upper and lower region and pixel thresholds applied to reject the typical appearance of covariate factors. This approach neglects one of the fundamental limitations of gait recognition, namely pixel-wise confusion between covariate factor motion and natural gait motion. Das Choudhury and Tjahjadi [10] utilise Fourier descriptors based on the contour of the silhouette. Specifically, the presence of a briefcase covariate factor is confirmed should the number of contour points increase above a set threshold. This approach neglects the fact that covariate factors can exist within the confines of the silhouette contour, thus causing covariate factors to go undetected.

Discussion. From the aforementioned examples, it is clear that our bolt-on module requires a flexible covariate factor detection approach to (1) target covariate factors and thus avoid removing potentially salient body areas unaffected by covariate factors, (2) minimise the pixel-wise confusion between covariate factor motion and natural gait motion and (3) ensure that covariate factors which exist within the human figure silhouette are not overlooked.
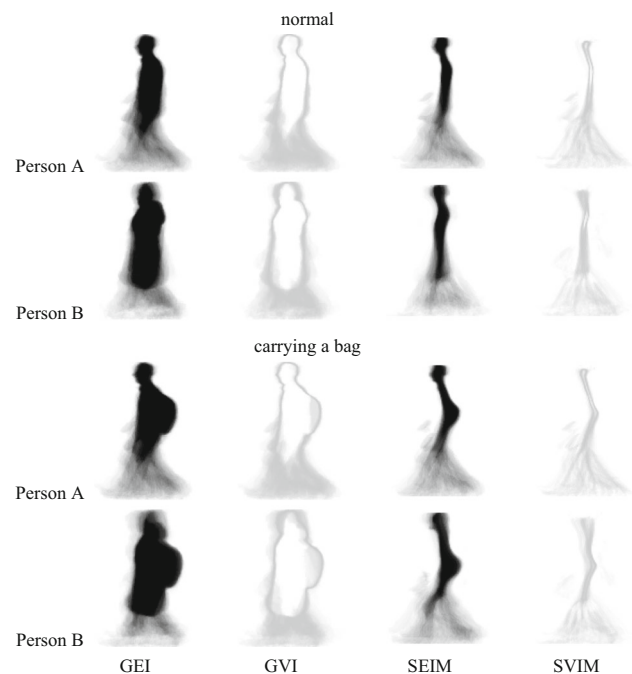
1.4 Validation

Our bolt-on module is applied to the Gait Energy Image [14], Gait Variance Image [33], Skeleton Energy Image [33] and Skeleton Variance Image [33] which vary in feature content and natural robustness, demonstrated in Fig. 1. Validation is performed on two of the largest and covariate factor rich standardised publicly available datasets: CASIA B [37,39] and TUM GAID [15,16]. Overall, our bolt-on module enables representations to achieve superior robustness thanks to ded-

icated covariate factor detection and removal procedures; these ensure a significantly greater quantity of covariate factor related pixel intensity values are targeted (this is vital for the trade-off between incorporating the natural inter-class and intra-class variance versus minimising the pixel-wise confusion between covariate factor and natural gait motion).

2 Gait representations

Our bolt-on module is designed for single compact 2D gait representations which are popular for model-free gait recognition. Single compact 2D gait representations boast (1) a compact nature due to condensing a gait sequence into a single 2D image, (2) natural robustness to noise and short term occlusion due to the normalisation procedures applied and (3) low computational and memory costs. Despite the compactness, these representations remain discriminative.

We select a mixture of traditional and recent gait representations which naturally vary in feature content and natural robustness, namely the Gait Energy Image (GEI), Gait Variance Image (GVI), Skeleton Energy Image (SEIM) and Skeleton Variance Image (SVIM) demonstrated in Fig. 1; baseline performances on which to improve are presented in Table 1 for validation on the CASIA B and TUM GAID datasets. This combination permits a rounded and enhanced robustness evaluation of our novel bolt-on module.

Fig. 1 Notice that despite the varying feature content within the GEI, GVI, SEIM and SVIM validation representations, the unique nature of gait is clearly evident. More importantly, notice the various ways in which the carried bag covariate factor manifests itself within the representations

Table 1 Covariate factor removal results for the CASIA B dataset (normal (nm), carrying a bag (bg), clothing (cl)) and TUM GAID dataset (normal (N), carrying a bag (B), shoes (S), time and normal (TN), time and carrying a bag (TB), time and shoes (TS)) across all gait representations (GR); optimal "typical" GR leniency and covariate factor threshold parameters are utilised for their highest average performance across all covariate factor types in a dataset. With exception of the SEIM validated on the CASIA B dataset, notice how the bolt on module enhances the robustness of all GRs across each dataset

GR	CASIA B (%): remove		nm	bg	cl	Average
GEI	Baseline		**100.0**	53.2	22.2	58.5
	Bolt-on module	CV only	99.2	75.4	64.1	79.6
		CV outwards	99.2	76.6	65.7	80.5
		CV rows completely	98.4	**77.4**	**93.1**	**89.7**
GVI	Baseline		**97.2**	77.8	50.4	75.1
	Bolt-on module	CV only	**97.2**	78.6	50.8	75.5
		CV outwards	**97.2**	78.2	50.8	75.4
		CV rows completely	95.6	**85.9**	**71.4**	**84.3**
SEIM	Baseline		98.0	93.1	69.8	87.0
	Bolt-on module	CV only	98.0	93.1	69.8	87.0
		CV outwards	98.0	93.1	69.8	87.0
		CV rows completely	98.0	93.1	69.8	87.0
SVIM	Baseline		**98.4**	92.7	71.8	87.6
	Bolt-on module	CV only	98.0	96.4	72.6	89.0
		CV outwards	98.0	**96.8**	73.0	**89.2**
		CV rows completely	97.2	94.8	**73.8**	88.6

GR	TUM GAID (%): remove		N	B	S	TN	TB	TS	Average
GEI	Baseline		**99.7**	19.0	**96.5**	34.4	0.0	43.8	67.4
	Bolt-on module	CV only	98.1	53.9	88.1	**43.8**	**28.1**	37.5	75.9
		CV outwards	99.0	42.3	92.3	40.6	15.6	43.8	73.7
		CV rows completely	98.7	**58.1**	87.4	37.5	21.9	**46.9**	**77.1**
GVI	Baseline		**99.0**	47.7	94.5	62.5	15.6	**62.5**	77.3
	Bolt-on module	CV only	98.1	64.2	94.2	**65.6**	28.1	**62.5**	82.4
		CV outwards	98.4	**68.1**	95.8	59.4	25.0	50.0	83.4
		CV rows completely	98.7	**68.1**	93.9	62.5	**34.4**	59.4	**83.6**
SEIM	Baseline		98.7	18.4	**96.1**	31.3	0.0	31.3	66.4
	Bolt-on module	CV only	98.7	45.2	91.6	**37.5**	**25.0**	34.4	74.2
		CV outwards	**99.0**	**47.1**	92.9	**37.5**	12.5	25.0	**74.6**
		CV rows completely	98.7	45.5	92.3	**37.5**	**25.0**	**37.5**	**74.6**
SVIM	Baseline		**98.4**	64.2	**91.6**	65.6	31.3	50.0	81.4
	Bolt-on module	CV only	98.1	73.2	89.4	65.6	40.6	**56.3**	83.6
		CV outwards	**98.4**	70.3	**91.6**	71.9	34.4	53.1	83.6
		CV rows completely	**98.4**	**74.8**	89.7	**68.8**	**43.8**	43.8	**84.3**

Bold values indicate the highest performance for each sequence and gait representation

2.1 Gait Energy Image

The Gait Energy Image (GEI) [14] exhibits static (torso appearance) and dynamic (limb motion) features differentiated by high and low pixel intensity values respectively. Construction first requires the silhouette sequence to be space-normalised to ensure silhouettes are (1) of constant size and (2) horizontally aligned with respect to a reference point, e.g. head or centre mass. Next, time-normalisation, via the pixel-wise mean, condenses the silhouette sequence into a single compact 2D gait representation; this aspect is advantageous for noise and short-term occlusion robustness. The key limitation relates to the pixel-wise confusion between covariate factor and natural gait motion which is an unfortunate consequence of (1) the construction process and (2) incorporating a mixture of static and dynamic features.

2.2 Gait Variance Image

The Gait Variance Image (GVI) [33] extracts only dynamic features due to their saliency [25]. The construction process mimics the GEI where time-normalisation employs the pixel-wise variance in place of the pixel-wise mean; this process exploits dynamic features whilst suppressing static features. The inherent risk with the GVI is enhancing noise (instead of suppressing noise due to utilising the pixel-wise mean in the case of the GEI) due to computing the pixel-wise variance. Regardless, the saliency of dynamic features outweighs this limitation to yield often double the performance achieved by the GEI, demonstrated in Table 1.

2.3 Skeleton Variance Image

The Skeleton Variance Image (SVIM) [33] utilises skeletons derived from silhouettes and extracts dynamic features utilising similar space- and time-normalisation procedures as the GVI. In brief, computing the gradient of a smoothed distance function based on the Screened Poisson equation yields a skeleton robust to silhouette quality (see [33] for further details). The SVIM is successful due to the (1) saliency of dynamic features and (2) skeletons naturally emphasizing gait motion whilst suppressing covariate factor motion.

2.4 Skeleton Energy Image

The Skeleton Energy Image (SEIM) [33] is analogous to the GEI where skeletons (same procedures as the SVIM) replace silhouettes. The advantages of utilising skeletons over silhouettes remain; however the inclusion of static features due to computing the pixel-wise mean can be a double edged sword performance-wise depending on silhouette quality, i.e. missing head or limbs due to silhouette extraction techniques.

3 Covariate factor detection

The GEI, GVI, SEIM and SVIM are founded on silhouettes which disregard colour and texture thus avoiding bias to subject appearance, e.g. clothing, which is inconsistent over time; this is fundamental during the complex and coupled time-based covariate factors presented by the TUM GAID dataset. Covariate factor detection is not a trivial task; detection in an RGB image is easier by human eye compared to silhouettes. Considering the silhouettes in Fig. 2, it is relatively straightforward to identify the bags should they disrupt the known outline of the human body, however the extent to which they encroach the silhouette is unknown. Therefore when utilising silhouettes, it is impossible to identify and remove every covariate factor related pixel.

Fig. 2 Detecting the extent of covariate factors is easier by eye in RGB images. Silhouette image covariate factor detection relies on disruptions of the known human body outline

GEI   GVI

SEIM   SVIM
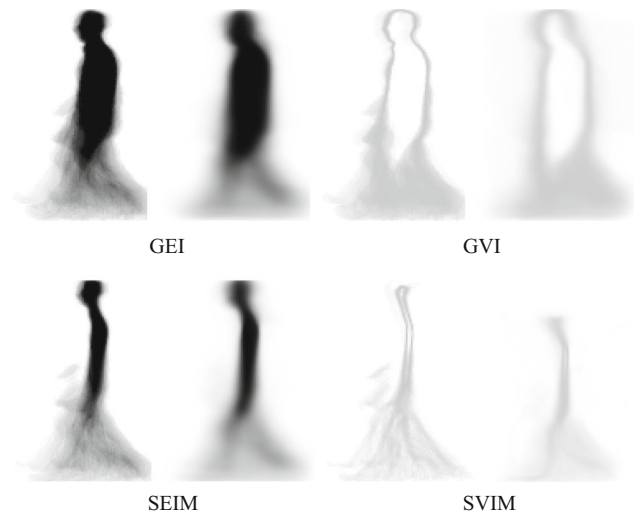
Fig. 3 Test GR (*left in pair*) versus "typical" GR (*right of pair*); notice the smoothing especially around limb areas

Considering Fig. 1, it is easy for humans to roughly detect the carried bag covariate factor in these gait representations (GR i.e. the GEI, GVI, SEIM and SVIM) if a "typical" covariate factor free (normal) GR is known. Therefore, it is imperative for computer vision to mimic this ability in order to achieve a similar skill set. Since gait recognition datasets provide specific training and test image sequences, it is possible to construct a covariate factor free "typical" GR by computing the pixel-wise mean of all training GRs; the averaging applied causes "typical" GR smoothing demonstrated in Fig. 3. The "typical" GR enables the bolt-on module to understand how the body is posed and distributed with respect to pixel intensity values; this is crucial for maximising covariate factor pixel detection and minimising pixel-wise confusion between covariate factor and natural gait motion.

Gait recognition is concerned with the walk action which follows the same fundamental pattern of movement; subtle differences relating to magnitude and timing yield inter-class and intra-class variance demonstrated in Fig. 1. It is imperative to incorporate this subtle inter-class and intra-class variance within the "typical" GR; this is achieved by incorporating the pixel-wise standard deviation of all training GRs

666
                                                                    T. Whytock et al.

to simulate a degree of pixel-wise leniency. Should the pixel intensity values in the GRs display a Gaussian distribution, the three-sigma rule (stating that nearly all values lie within three standard deviations of the mean) could be applied, however this is not true due to the unequal quantity of static and dynamic features within the GRs (especially when covariate factors are present).

Regardless, we define four levels of leniency for each "typical" GR

$$tGR_0 = \overline{GR}, \quad tGR_1 = \overline{GR} \pm \sigma,$$
$$tGR_2 = \overline{GR} \pm 2\sigma, \quad tGR_3 = \overline{GR} \pm 3\sigma \quad (1)$$

where $tGR_{0,1,2,3}$ are "typical" GRs, and $\overline{GR}$ and σ are the pixel-wise mean and pixel-wise standard deviation of all training GRs respectively. tGR_0 is simply the "typical" GR sans leniency, while $tGR_{1,2,3}$ incorporate $\pm 1, 2, 3\sigma$ to simulate successively greater leniency within the "typical" GR; notice $\pm$ which is essential given pixel intensity values may lie either side of the mean due to the uniqueness of gait. These leniency definitions are vital to determine their contribution to satisfy the trade-off between incorporating inter-class and intra-class variance versus minimising the pixel-wise confusion between covariate factor motion and natural gait motion. Further still, this trade-off is likely to vary across the validation GRs given the features utilised and natural robustness therein.

Therefore areas of the body affected by covariate factors are detected by comparing a "typical" GR against a test GR

$$CV = |GR - tGR_{0,1,2,3}| \quad (2)$$

where CV are detected covariate factor areas, and GR and $tGR_{0,1,2,3}$ are test and "typical" GRs respectively. For illustrative purposes, consider the GEI in Fig. 4 to help visualise the detected covariate factor areas with respect to leniency included in the "typical" GEI during the CASIA B dataset. Starting with the a normal (covariate factor free) test GEI, it is clear to see the advantage of utilising a large σ (tGR_3) to incorporate the greatest possible amount of inter-class and intra-class variance. However, now consider the carrying a bag test GEI. A significant amount of covariate factor areas are detected when σ is small (tGR_0) due to (1) the bag and (2) leaning due to a shifted centre of gravity; i.e. there is considerable pixel-wise confusion between covariate factor motion and natural gait motion. When σ is larger (tGR_3), obviously fewer covariate factor areas are detected given the subsequent decrease in pixel-wise confusion between covariate factor and natural gait motion. Remember while covariate factors are static with respect to the body, covariate factors undergo motion due to the nature of human gait. This observation is mimicked during the clothing test GEI. Given covariate

test GEI   Detected covariate factor areas CV

using

$tGR_0 \qquad tGR_1 \qquad tGR_2 \qquad tGR_3$
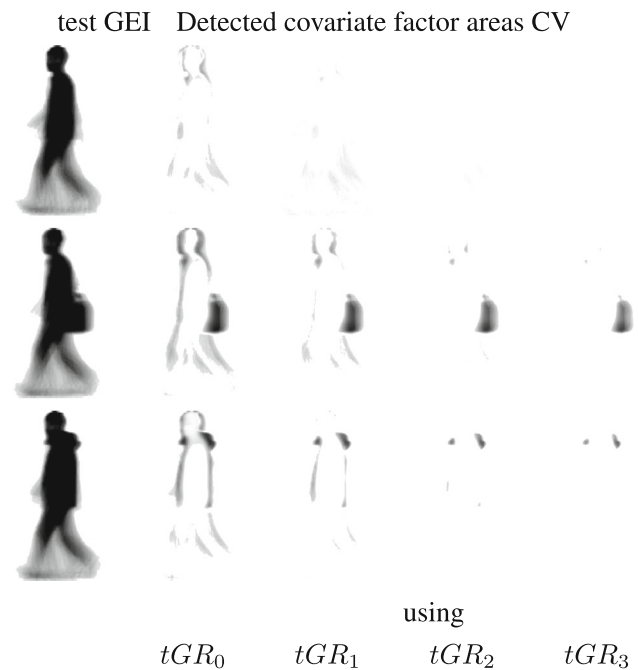
Fig. 4 Detected covariate factor areas in the CASIA B dataset with respect to leniency included within the "typical" GEI $tGR_{0,1,2,3}$ for normal (*top*), carrying a bag (*middle*) and clothing (*bottom*) covariate factors

factor performance is paramount for robust gait recognition, Fig. 4 clearly demonstrates that $> \pm 3\sigma$ is not beneficial and therefore represents the boundary of leniency values considered.

4 Covariate factor removal

While our covariate factor detection stage aims to locate the greatest extent of covariate factor areas possible, the employed removal technique is the final opportunity to contribute an effective pixel-wise distinction between covariate factors and natural gait motion. Covariate factor areas must be removed as their covariate factor free equivalent is unknown. Removal therefore requires a three stage process: (1) apply a threshold to the detected covariate factor areas to satisfy the trade-off between incorporating inter-class and intra-class variance versus minimising the pixel-wise confusion between covariate factor and natural gait motion, (2) remove these areas and finally (3) reclaim any salient limb-based dynamic features if removed by preceding stages.

Stage 1: covariate factor threshold

All test GRs are normalised with respect to pixel intensity values due to the GR construction process. Therefore, detected covariate factor areas vary with respect to pixel intensity value depending on the covariate factor present; normalising these values is ill advised as they are indicative of person iden-

tity. A threshold is therefore employed to satisfy the trade-off for incorporating inter-class and intra-class variance versus minimising the pixel-wise confusion between covariate factor and natural gait motion. This is especially beneficial in the case of Fig. 4 where a low leniency "typical" GEI is combined with a covariate factor free test GEI, this leads to incorrectly detected covariate factor areas due to the lack of inter-class and intra-class variance incorporated; a threshold can alleviate this issue which also affects the GVI, SEIM and SVIM. The application of a threshold therefore ultimately decides the detected covariate factor areas for removal. A broad range of values are chosen {$T_h = 0.1$ to 1 in steps of 0.1} to analyse their contribution to effective and complete covariate factor removal; these values are chosen as test GRs contain normalised pixel intensity values, i.e. lying between 0 and 1. Note that the head area is often deemed a covariate factor area in Fig. 4 due to either natural body rotations or leaning due to compensation for a shifted centre of gravity when e.g. carrying a bag.

Stage 2: covariate factor removal techniques

The primary objective is to ensure maximum covariate factor removal. Three covariate factor removal techniques (pseudocode provided in Algorithm 1) are evaluated which vary in aggression and demonstrated in Fig. 5 for the GEI in the CASIA B dataset.

Result: Remove covariate factors only
for *every pixel value* **do**
    if *pixel value > threshold* **then**
        set pixel value to zero;
    end
end

Result: Remove covariate factors outwards from the centreline
        of the body
for *every row* **do**
    calculate midpoint of the GR
    for *RHS (centreline $\longrightarrow$ RHS)* **do**
        for *every pixel value* **do**
            if *pixel value > threshold* **then**
                set all pixel values in the row to zero;
            end
        end
    end
    repeat process for LHS
end

Result: Remove covariate factor rows completely
for *every row* **do**
    for *every pixel value* **do**
        if *pixel value > threshold* **then**
            set all pixel values in the row to zero;
        end
    end
end
Algorithm 1: Pseudocode for each covariate factor removal technique

---
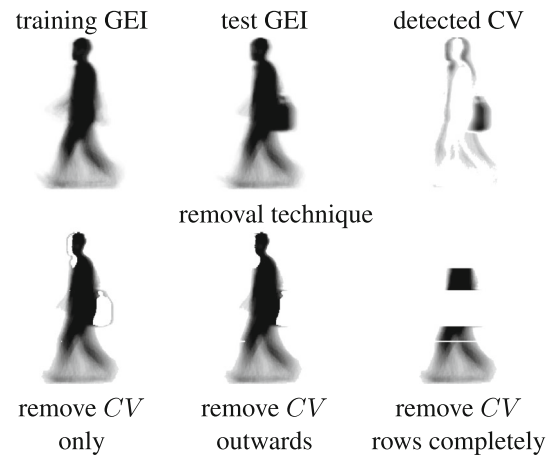
Fig. 5 Detected covariate factor areas CV are removed using three removal techniques: remove CV only, remove CV outwards from the centreline of the body and remove CV rows completely; notice the increasing removal aggression and the similarity with remaining pixel intensity values when compared to a training GEI

Removing covariate factors only. This is the least aggressive removal technique and is classed as high risk if the detection stage cannot effectively differentiate pixel-wise between covariate factor motion and natural gait motion. The effect of this limitation is clearly demonstrated in Fig. 5 where an outline (consisting of covariate factor motion related pixel intensity values) surrounds the removed pixels. As a result, poorer performance is naturally expected.

Removing covariate factors outwards from the centreline of the body. Figure 5 demonstrates a proactive technique resolving the limitations of removing covariate factor areas only. Any residual effects are targeted by removing all covariate factor areas outwards from the centreline of the body; therefore increased robustness is naturally anticipated.

Removing covariate factor rows completely. Presented in our preliminary studies [31,32], this removal technique is aggressive and we therefore expect high performance. Figure 5 demonstrates complete rows are removed only where covariate factor areas are detected. This works on the premise that covariate factors can be found within, or at the boundary of, the figure thus causing increased pixel-wise confusion between covariate factor motion and natural gait motion. While this technique may cause a significant quantity of the GR to be removed and risk jeopardising person identity, this limitation is specifically addressed during stage 3.

Stage 3: reclaiming salient leg features

Limb motion is salient [25] and removal from GRs is ill advised; baseline performance in Table 1 demonstrates GRs containing dynamic features only tend to outperform those containing both static and dynamic features i.e. the GVI is

superior to the GEI. Given covariate factor detection depends on "typical" GR leniency and tunable threshold value, a significant quantity of the test GR in question can be removed during Stage 2. Therefore, any salient leg motion (dynamic features) is reclaimed by initiating a bottom-up search to identify the highest GR image row containing a low pixel intensity value; all rows up to this point are retained completely despite their initial removal by Stages 1 and 2 (this process is similar in nature to [1]).

5 Experimental procedure

5.1 Dataset

The CASIA B and TUM GAID datasets are employed for validation. The CASIA B dataset is nearing a decade old however deployment for validation remains popular to this day. In an indoor environment, 124 subjects are captured under three covariate factors: (1) normal i.e. covariate factor free, (2) carrying a bag which varies across the dataset, e.g. handbags, rucksacks and (3) clothing in the form of bulky outdoor jackets varying in length and shape. A clear divide separates training and test data. Training data utilises four normal image sequences per person; test data utilises two image sequences per covariate factor per person.

The TUM GAID dataset is also based indoors however its recency means fewer gait recognition validation results exist. Containing 305 and 155 training and test subjects respectively, this dataset also boasts complex coupled covariate factors which are highly attractive traits. Six covariate factors are presented, three single: (1) carrying a bag (rucksack) consistent across the dataset and (2) shoes, i.e. wearing clean room shoe covers; an additional three complex coupled covariate factors are utilised where image sequences are captured three months later and therefore include clothing as an additional and hidden covariate factor due to a weather season change: (3) time and normal, (4) time and bag and (5) time and shoes. Image sequence division for test and training data mirrors the CASIA B dataset.

Both datasets provide silhouettes which permits research to focus on the gait recognition problem as opposed to silhouette extraction techniques. Overall, the CASIA B dataset contains poorer quality silhouettes compared to the TUM GAID dataset. The TUM GAID dataset utilises the Microsoft Kinect to extract depth information enabling relatively clean and intact silhouettes. The CASIA B dataset utilises background subtraction [37] which extracts imperfect silhouettes containing extraneous noise with missing heads or limbs. This difference in silhouette quality is advantageous as silhouette quality robustness is an aim for our bolt-on module.

5.2 Validation representations

Commonly employed, the GEI, GVI, SEIM and SVIM are space-normalised utilising the head as a reference point for horizontal alignment; note that space-normalisation is performed prior to skeletonisation for the SEIM and SVIM. The SEIM and SVIM are constructed utilising optimal parameters set out by [33]; smoothing parameters respectively are (1) CASIA B dataset: $t = 30$ and $t = 70$ and (2) TUM GAID dataset: $t = 90$ and $t = 5$.

5.3 Dimensionality reduction and classification

Standard to single compact 2D gait representations the GEI, GVI, SEIM, SVIM represent gait (standard dataset image sizes CASIA B: 240×240, TUM GAID: 128×178) and describe gait when reshaped to a 1D vector (CASIA B: $57600D$, TUM GAID: $22784D$). This yields undesirably high dimensionality feature vectors which are alleviated by typical gait recognition procedures [14]. Principle Component Analysis and Linear Discriminant Analysis [23] are combined to satisfy the best data representation with respect to covariance and class separability respectively (CASIA B: $123D$, TUM GAID: $154D$ account for approximately 97 % variance). Nearest Neighbour classification is performed alongside the Euclidean and Cosine distance metrics which are standards set by the CASIA B and TUM GAID datasets respectively. This dimensionality reduction and classification combination is very effective and typical [14] for single compact 2D gait representations due to the small number of training sequences available in gait recognition datasets.

5.4 Recognition procedure

Each test GR is considered in turn and detected covariate factor areas are removed from both test and training GRs to ensure dimensionality reduction and classification are performed on areas deemed covariate factor free. This is essential for Nearest Neighbour classification where it is unfair to compare test GRs with covariate factor areas removed against complete "typical" GRs; this promotes increased distances during Nearest Neighbour classification thus triggering unnecessary misclassification.

For each covariate factor type, Nearest Neighbour classification assigns a predicted subject ID to each test GR given the dataset specified distance metric. A confusion matrix is therefore constructed based on the predicted and known subject IDs. Therefore, the performance of our bolt-on module (percentages seen in Table 1) is calculated by averaging the diagonals of the confusion matrix and dividing by the subject number in the dataset.

6 Results

Four key factors are considered when analysing the results from each GR, namely performance effects due to (1) covariate factors, (2) "typical" GR leniency, (3) covariate factor threshold and (4) covariate factor removal techniques.

Figure 6 provides detailed analysis of "typical" GR leniency and covariate factor threshold for each GR and dataset during covariate factor free, covariate factor and the average of all sequences. For simplicity, covariate factor removal technique performance is presented in Table 1 for optimal "typical" GR leniency and covariate factor threshold parameters yielding the highest average performance across all sequences. Across datasets, this table demonstrates how our bolt-on module increases GR performance by 15.1 %, and increases covariate factor sequence performance by 15.9 %.

Note that we are primarily interested in the average performance across covariate factors in each dataset as this demonstrates the ability for our bolt-on module to generalise over covariate factors; this is standard for gait recognition. The performance achieved during individual covariate factors is important to determine weaknesses in our bolt-on module. All tables and figures highlight the baseline performance enabling easy comparison post bolt-on module application.

6.1 Covariate factor effect on performance

Consider Table 1 which demonstrates performance across each covariate factor (utilising optimum "typical" GR leniency and covariate factor threshold). Covariate factor free test GRs (CASIA B: nm; TUM GAID: N, S) are visually similar to training GRs which explains their high performance and therefore serve as proof of concept. While the TUM GAID dataset provides shoe sequences, their appearance is visually similar to training GRs thus yielding high performance; note that the clean room shoe covers are primarily aimed towards being an acoustic gait recognition covariate factor. Alternative shoe types such as flip flops [5] and heels will cause greater misclassification due to greater alterations in gait appearance and motion.

Covariate factor GRs (CASIA B: bg, cl; TUM GAID: B, TN, TB, TS) yield significant gait appearance and motion alterations thus causing detrimental performance drops. One major difference between the CASIA B and TUM GAID dataset is the bag carried; the TUM GAID dataset utilises a consistent rucksack while the CASIA B dataset employs a range of bags, e.g. rucksacks, handbags; on top of image size differences, this variation likely accounts for the performance differences encountered between these sequences. The time-based covariate factor sequences utilised by the TUM GAID dataset show particularly poor performance due to the cou-

pled covariate factors in play i.e. time and clothing on top of the named covariate factor. Notice in Table 1 how the performance of such sequences are near half that achieved by single covariate factor sequences. The aforementioned patterns occur in the GEI, GVI, SEIM and SVIM.
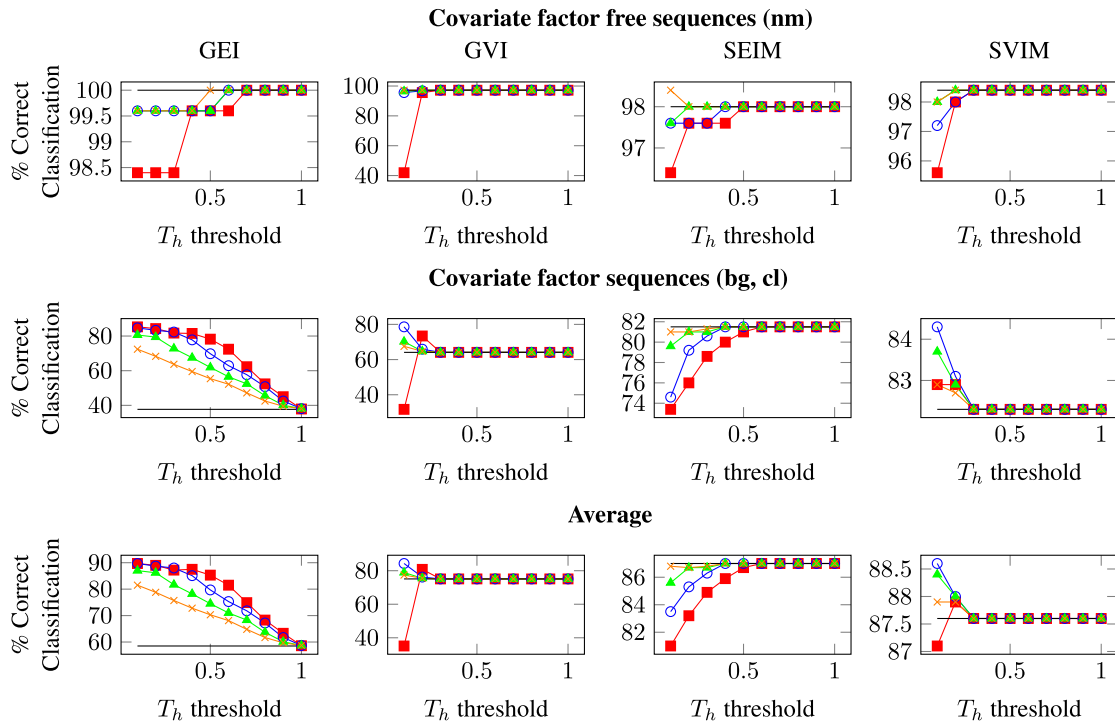
6.2 "Typical" GR leniency

Consider Fig. 6 which demonstrates "typical" GR leniency across covariate factors for the "remove covariate factor rows completely" covariate factor removal technique (all removal techniques demonstrate similar performance patterns). Including leniency in the "typical" GR is aimed towards incorporating the abundant inter-class and intra-class variance encountered in human gait; however inclusion is a double edged sword due to potential pixel-wise confusion between covariate factor motion and natural gait motion.

Consider the GEI which is the least naturally robust GR in this paper. Covariate factor free GEIs (CASIA B: nm, TUM GAID: N, S) benefit from higher "typical" GR leniency (tGR_3) as no covariate factors are present and therefore significant inter-class and intra-class variance can be incorporated to boost performance. Conversely covariate factor GEIs (CASIA B: bg, cl, TUM GAID: B, TN, TB, TS) require negligible "typical" GR leniency (tGR_0) as greater inter-class and intra-class variance promotes unnecessary pixel-wise confusion between covariate factor and natural gait motion. To achieve the greatest performance across covariate factors, the GEI prefers tGR_0 regardless of dataset.

The GVI, SEIM and SVIM demonstrate a degree of natural robustness compared to the GEI which is demonstrated in Table 1 baseline results; robustness is achieved by (1) the GVI extracting dynamic features only and (2) the SEIM and SVIM placing emphasis on natural gait motion as opposed to covariate factor motion. Covariate factor free GVI, SEIM and SVIM (CASIA B: nm, TUM GAID: N, S) similarly benefit from higher "typical" GR leniency (tGR_3). However their natural robustness permits covariate factor GRs (CASIA B: bg, cl, TUM GAID: B, TN, TB, TS) higher "typical" GR leniency (tGR_1); this enables a greater quantity of inter-class and intra-class variance to be incorporated in the GRs whilst simultaneously minimising the pixel-wise confusion between covariate factor motion and natural gait motion. Notice that the SEIM validated on the CASIA B dataset does not achieve any performance increases which is attributed to SEIM skeleton sensitivity to poorer quality image sequences causing often significant differences between training and test SEIMs; the SVIM receives only a minor performance increase during the CASIA B dataset. To achieve the greatest performance across covariate factors the GVI, SEIM and SVIM prefer tGR_1 regardless of dataset.

GRs validated on the CASIA B dataset
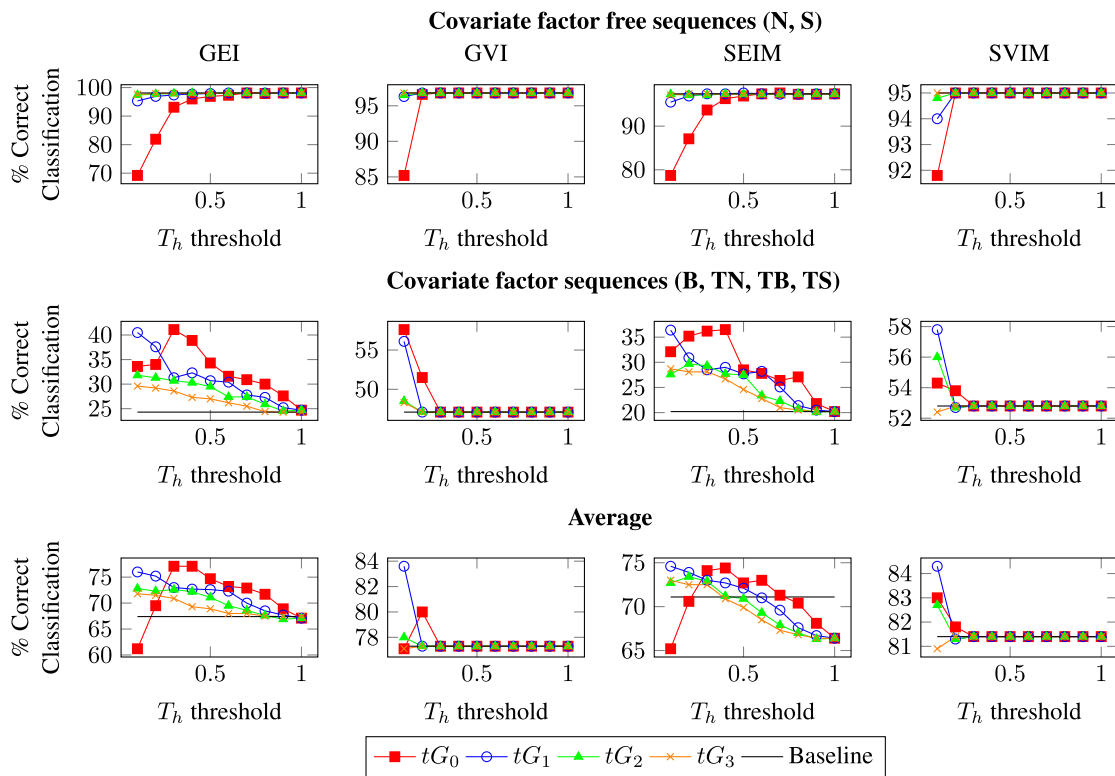
Fig. 6 Covariate factor threshold results for each GR validated on the CASIA B and TUM GAID datasets based on covariate factor free sequences (CASIA B: normal (nm), TUM GAID: normal (N), shoes (S)), covariate factor sequences (CASIA B: carrying a bag (bg), clothing (cl), TUM GAID: bag (B), time and normal (TN), time and bag (TB), time and shoes (TS)) and averaged across all covariate factor types. Results with respect to leniency $tGR_{0,1,2,3}$, covariate factor threshold T_h and "remove covariate factor rows completely" covariate factor removal technique (all covariate factor removal techniques exhibit similar patterns). Notice how the bolt-on module considerably increases robustness during covariate factor sequences

6.3 Covariate factor threshold

Consider Fig. 6 which demonstrates the covariate factor threshold behaviour during covariate factor free (CASIA B: nm, TUM GAID: N, S), covariate factor (CASIA B: bg, cl, TUM GAID: B, TN, TB, TS) and the average of all sequences. The covariate factor threshold is designed to further satisfy the trade-off between incorporating inter-class and intra-class variance versus minimising the pixel-wise confusion between covariate factor and natural gait motion. Similar to "typical" GR leniency, the optimal covariate factor threshold value varies between covariate factor and covariate factor free GRs. A high threshold value is ideal for covariate factor free GRs to maximise the incorporation of natural inter-class and intra-class variance (given no covariate factors are present to induce pixel-wise confusion). However, it is paramount to utilise a low threshold value for covariate factor sequences to satisfy the aforementioned trade-off and thus yield robust gait recognition. Notice how Fig. 6 clearly demonstrates significant robustness improvements for each GR during covariate factor sequences (the SEIM receives fewer performance increases during the CASIA B dataset due to the lack of similarity between training and test SEIMs and skeleton sensitivity to noise); note that the covariate factor free sequences match the baseline results.

When averaging across all presented covariate factor types, the CASIA B and TUM GAID datasets unanimously prefer $T_h = 0.1$ for the GVI, SEIM and SVIM despite their different image sizes (standard dataset image sizes CASIA B: 240×240, TUM GAID: 128×178) and across presented covariate factors. However the GEI varies between $T_h = 0.1$ and $T_h = 0.3$ for the CASIA B and TUM GAID datasets respectively; this difference is attributed to the inclusion of static and dynamic features and the inherent pixel-wise confusion between covariate factor and natural gait motion. Note that these threshold values are for gait recognition and require further investigation for analogous applications.

6.4 Covariate factor removal technique

This is the final opportunity to remove covariate factors which may have previously evaded detection. Table 1 presents the best combinations of "typical" GR leniency and threshold for the CASIA B and TUM GAID datasets achieving the highest average performance across covariate factors (Fig. 5 helps visualise each removal technique).

6.4.1 Removing covariate factors only

The least aggressive and most risky technique is low and middle ranking dataset dependent. Despite the best efforts of covariate factor detection, pixel-wise confusion between covariate factor and natural gait motion occurs frequently causing covariate factor motion related pixel intensity values to enclose removed covariate factor areas; this is demonstrated in Fig. 5.

6.4.2 Removing covariate factors outwards from the centreline of the body

While this approach visually resolves the limitations of removing covariate factors only, increased performance is not exhibited. Removing covariate factors only and removing covariate factors outwards from the centreline of the body neglect an important point. Covariate factors can lie within, and at the boundary of, the human figure which increases the complexity of differentiating between covariate factor and natural gait motion. As such, the ranking is also middle and low ranking dataset dependent.

6.4.3 Removing covariate factor rows completely

The most aggressive and originally implemented removal technique remains superior across datasets. Compared to aforementioned removal techniques which gingerly remove covariate factor areas, this technique is aggressive to the point of appearing risky; remember stage 3 of covariate factor removal ensures salient limb dynamic features are retained if removed during preceding stages. This technique could be considered sensitive with respect to natural inter-class and intra-class variance, however we demonstrate it is best to err on the side of caution to ensure minimal pixel-wise confusion between covariate factor and natural gait motion.

7 Comparison to state of the art

The best performing GR bolt-on module parameters (1) "typical" GR leniency, (2) covariate factor threshold and (3) removal technique, for the CASIA B and TUM GAID datasets are posed against state-of-the-art results in Table 2. We compare against individual covariate factors, and more importantly those achieving the highest average performance to demonstrate the ability to generalise over covariate factors.

We therefore successfully demonstrate our bolt-on module enhances GR robustness with new state-of-the-art results

CASIA B: carrying a bag (bg) + 4.0 %, clothing (cl) + 11.0 %
TUM GAID: bag (B) + 16.5 %, time + normal (TN) + 4.9 %, time + bag (TB) + 39.9 %, time + shoes (TS) + 18.8 %, weighted average + 3.6 %

where on average, our bolt-on module provides a 15.9 % increase to covariate factor sequences.

quality must permit similarities between training and test GRs.

8 Conclusion and future work

We have clearly demonstrated the benefits of our bolt-on module which incorporates covariate factor detection and removal within standard gait recognition procedures; this significantly enhances the robustness of single compact 2D gait representations by an average of 15.1 % thanks primarily to the tunable parameters. Extensive covariate factor detection and aggressive removal combine to contribute significant improvements by ensuring a favourable trade-off between incorporating the natural inter-class and intra-class variance versus minimising the confusion between covariate factor and natural gait motion. This yields multiple new state-of-the-art results across validation gait representations and datasets. Our bolt-on module is capable of generalising over covariate factors at the boundary of, and hidden within, the human figure where such sequences receive a 15.9 % increase.

Three future directions exist for our bolt-on module and gait recognition in general: (1) there exists no single optimal threshold or leniency parameter, which effectively combats every covariate factor, and this is natural as each affects gait appearance and motion uniquely; one possible solution exists whereby covariate factors are initially detected and the preferred threshold and leniency parameter subsequently applied—note that this process is best achieved utilising RGB images as too many visual cues are lost when utilising silhouettes, (2) greater research should focus on the complex and coupled time-based covariate factors given Table 2 highlights the interesting and open problem of exceeding 50 % in performance, and (3) analysis of the bolt-on module for single compact 2D representations during analogous research topics, e.g. action recognition.

References

1. Bashir, K., Xiang, T., Gong, S.: Feature selection for gait recognition without subject cooperation. In: Proceedings of the British Machine Vision Conference (2008)

2. Bashir, K., Xiang, T., Gong, S.: Gait recognition using gait entropy image. In: International Conference on Crime Detection and Prevention, pp. 1–6 (2009)

3. Bashir, K., Xiang, T., Gong, S.: Gait representation using flow fields. In: Proceedings of the British Machine Vision Conference (2009)

4. Bashir, K., Xiang, T., Gong, S.: Gait recognition without subject cooperation. Pattern Recognit. Lett. **31**(13), 2052–2060 (2010)

5. Bouchrika, I., Nixon, M.: Exploratory factor analysis of gait recognition. In: IEEE International Conference on Automatic Face and Gesture Recognition (2008)

6. Chai, Y., Ren, J., Han, W., Li, H.: Human gait recognition: approaches, datasets and challenges. In: International Conference on Imaging for Crime Detection and Prevention, pp. 1–6 (2011)

7. Collins, R., Gross, R., Shi, J.: Silhouette-based human identification from body shape and gait. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (2002)

8. Cunado, D., Nixon, M., Carter, J.: Using gait as a biometric, via phase-weighted magnitude spectra. In: Proceedings of the International Conference on Audio- and Video-Based Biometric Person Authentication, pp. 95–102. Springer, Berlin (1997)

9. Cutting, J., Kozlowski, L.: Recognising friends by their walk: gait perception without familiarity cues. Bull. Psychon. Soc. **9**(5), 353–356 (1977)

10. Choudhury, D.S., Tjahjadi, T.: Gait recognition based on shape and motion analysis of silhouette contours. Comput. Vision Image Underst. **117**(12), 1770–1785 (2013)

11. Dempster, W., Gaughran, G.: Properties of body segments based on size and weight. Am. J. Anat. **120**(1), 33–54 (1967)

12. Drillis, R., Contini, R.: Body segment parameters. Office of Vocational Rehabilitation, Department of Health, Education and Welfare, New York (Report No. 1163.03) (1966)

13. Gafurov, D.: A survey of biometric gait recognition: approaches, security and challenges. In: Nik Conference (2007)

14. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2), 316–322 (2006)

15. Hofmann, M., Bachmann, S., Rigoll, G.: 2.5D gait biometrics using the depth gradient histogram energy image. In: IEEE International Conference on Biometrics: Theory, Applications and Systems, pp. 399–403 (2012)

16. Hofmann, M., Geiger, J., Bachmann, S., Schuller, B., Rigoll, G.: The tum gait from audio, image and depth (GAID) database: multimodal recognition of subjects and traits. J. Visual Commun. Image Represent. **25**(1), 195–206 (2014). (Special issue on visual understanding and applications with RGB-D Cameras)

17. Huang, X., Boulgouris, N.: Gait recognition with shifted energy image and structural feature extraction. IEEE Trans. Image Process. **21**(4), 2256–2268 (2012)

18. Lee, L., Grimson, W.: Gait analysis for recognition and classification. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, pp. 148–155 (2002)

19. Li, N., Xu, Y., Yang, X.: Part-based human gait identification under clothing and carrying condition variations. Int. Conf. Mach. Learn. Cybern. **1**, 268–273 (2010)

20. Li, X., Chen, Y.: Gait recognition based on structural gait energy image. J. Comput. Inf. Syst. **9**(1), 121–126 (2013)

21. Li, X., Maybank, S., Yan, S., Tao, D., Xu, D.: Gait components and their application to gender recognition. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **38**(2), 145–155 (2008)

22. Lin, H.W., Hu, M.C., Wu, J.L.: Gait-based action recognition via accelerated minimum incremental coding length classifier. Lecture Notes in Computer Science. Adv. Multimed. Model. **7131**, 266–276 (2012)

23. van der Maaten, L.: Matlab Toolbox for Dimensionality Reduction

24. Makihara, Y., Okumura, M., Iwama, H., Yagi, Y.: Gait-based age estimation using a whole-generation gait database. In: International Joint Conference on Biometrics, pp. 1–6 (2011)
25. Martín-Félez, R., Xiang, T.: Gait recognition by ranking. Lecture Notes in Computer Science. Comput. Vision ECCV **7572**, 328–341 (2012)
26. Matovski, D., Nixon, M., Carter, J.: Encyclopedia of computer vision, chap. Gait recognition. Springer, Berlin (in press) (2013)
27. Murray, M., Drought, A., Kory, R.: Walking patterns of normal men. J. Bone Joint Surg. **46**(2), 335–360 (1964)
28. Niyogi, S., Adelson, E.: Analyzing gait with spatiotemporal surfaces. In: Proceedings of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects, pp. 64–69 (1994)
29. Singh, S., Biswas, K.: Spatio-temporal energy based gait recognition. In: IEEE International Conference on Data Mining, pp. 998–1003 (2009)
30. Wang, C., Zhang, J., Wang, L., Pu, J., Yuan, X.: Human identification using temporal information preserving gait template. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2164–2176 (2012)
31. Whytock, T., Belyaev, A., Robertson, N.: Robust gait recognition via covariate factor mitigation. In: International Conference on Imaging for Crime Detection and Prevention (2013)
32. Whytock, T., Belyaev, A., Robertson, N.: Towards robust gait recognition. In: Advances in Visual Computing, Lecture Notes in Computer Science, Part II, vol. 8034, pp. 523–531. Springer, Berlin (International Symposium on Visual Computing) (2013)
33. Whytock, T., Belyaev, A., Robertson, N.: Dynamic distance-based shape features for gait recognition. J. Math. Imaging Vision (2014)
34. Yam, C., Nixon, M., Carter, J.: Gait recognition by walking and running: a model-based approach. In: Proceedings of the Asian Conference on Computer Vision, pp. 1–6 (2002)
35. Yogarajah, P., Condell, J., Prasad, G.: P_{RW} GEI: Poisson Random Walk based Gait Recognition. In: International Symposium on Image and Signal Processing and Analysis, pp. 662–667 (2011)
36. Yoo, J., Nixon, M.: Automated markerless analysis of human gait motion for recognition and classification. ETRI J. **33**(3), 259–266 (2011)
37. Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. Int. Conf. Pattern Recognit. **4**, 441–444 (2006)
38. Zhang, E., Zhao, Y., Xiong, W.: Active energy image plus 2DLPP for gait recognition. Signal Process. **90**(7), 2295–2302 (2010)
39. Zheng, S., Zhang, J., Huang, K., He, R., Tan, T.: Robust view transformation model for gait recognition. In: IEEE International Conference on Image Processing, pp. 2073–2076 (2011)

Alexander Belyaev is currently a Reader in Vision, Image & Signal Processing in the Institute of Sensors, Signals and Systems at School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK. His main research topics are digital geometry processing, mathematical image analysis, and applied partial differential equations, on which he published more than one hundred articles in international conferences and journals. Prior to joining Heriot-Watt University Belyaev worked at Max Planck Institute for Informatics (Saarbruecken, Germany, 2004-2007) where he led a geometric modeling group. Belyaev's previous employments include Visiting Researcher and Associate Professor at the University of Aizu (Aizu-Wakamatsu, Japan, 1993-2004) and Junior Research Fellow at Moscow State University (Moscow, Russia, 1989-1993). Belyaev received his PhD on asymptotic analysis of partial differential equations from Lomonosov Moscow State University in 1990.

Neil M. Robertson received the M.Sci. degree from Glasgow University, Glasgow, U.K. in 2000, and the D.Phil. degree from Oxford University, Oxford, U.K. in 2006. He is principle investigator of the Vision Lab at Heriot-Watt University and an Honorary Fellow of the School of Engineering, University of Edinburgh. He works on projects in collaborative robotics, human behaviour recognition, multi-modal registration, and sensor fusion with his research team. From 2000 to 2007, he worked in the U.K. Scientific Civil Service with DERA, and then QinetiQ. He held a prestigious 1851 Royal Commission Fellowship at Oxford University (2003-2006) in the Robotics Research Group.

Tenika Whytock received her M.Eng degree in Robotics and Cybertronics in 2010 and is now a Ph.D student in the Institute of Sensors, Signals and Systems at the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK. Her doctoral studies focus on robust model-free gait recognition approaches.