Gait Analysis of Pedestrians with the Aim of Detecting Disabled People

Amir Mukhtar[a*], Michael J. Cree[b], Jonathan B. Scott[c] and Lee Streeter[d]

School of Engineering, University of Waikato, Hamilton, New Zealand

[a]am301@students.waikato.ac.nz, [b]michael.cree@waikato.ac.nz, [c]jonathan.scott@waikato.ac.nz, [d]lee.streeter@waikato.ac.nz

Abstract. Gait classification is an effective and non-intrusive method for human identification and it has received significant attention in the recent years due to its applications in visual surveillance and monitoring systems. In this project, we analysed gait signatures using spatio-temporal motion characteristics of a person to answer the question "is there a discriminating feature in gait signal that can help to categorise disable person from healthy?". The procedure has three steps. Detection of a pedestrian using YOLO followed by the silhouette extraction using the Gaussian Mixture Model (GMM). Finally, skeletonization from the silhouette image to estimate head and torso locations and their angles with the vertical axis. Furthermore, velocity and acceleration signal were recorded to look for accelerating behaviour of person walking with a limp.

Introduction

This research aims to investigate computer vision and pattern recognition techniques for identifying physically disabled people appearing in surveillance footage. This is a challenging task since outdoor scenarios with background clutter, illumination conditions, camera viewing angle and camera resolution affect the quality of information sensed. Scenes obtained from a surveillance video are usually low resolution which adds to the complexity of the problem. Video surveillance technology is gaining popularity these days due to its low cost, ease of installation, freedom from interference, and its ability to capture detailed information in image system. Developments in computer vision, pattern recognition and machine learning have further increased the usefulness of optical sensors. Advancements in optical sensors and rapidly falling costs of high resolution cameras make them a preferred choice for an automated detection system design. We focus on gait signatures of a person walking to investigate whether gait can usefully distinguish healthy and disable persons. This involves pedestrian detection followed by extraction of the gait signal and its analysis. Skeletonization approach is computationally cheap and summarises the internal motion/gait of moving pedestrian without a prior human model.

Related Work

Vision based disable person detection involves extraction of moving objects (silhouettes) and their gait analysis for categorizing them into healthy or disable. Advances in computer vision lack such a system due to its limited applications. However, a great deal of research has been conducted on vision based pedestrian detection [1], human behaviour detection [2, 3] and gait recognition [4]. Therefore, the related literature (which may contribute to develop a gait based disable person detector) is divided into two parts,

Motion Detection and Silhouette Extraction

Background subtraction is a quick way of localizing moving objects in videos by a static camera and it is often the first step of a multi-stage computer vision system [5] (car tracking, pedestrian detection, wild-life monitoring, etc.). It assumes video sequence comprising of a static background in front of which moving objects (in distinct colours) are extracted. Some commonly used background modelling techniques are inter-frame change [6], GMM [7] and optical flow [8]. Each approach has

its advantages and disadvantages [9]. We picked GMM based for the system because it is good in handling multi-modal background scenarios.

Background modelling techniques result in silhouettes estimation of moving objects in a video sequence. Most gait analysis procedures [10, 11] rely on pre-detection of silhouettes to extract motion signals. While some approaches [12] take bounding boxes containing pedestrians. Pedestrian detection can be performed in two sequential steps: extracting candidate regions that are potentially covered by human objects and classifying the regions as human or non-human. Background subtraction is the simplest way to segregate moving objects and then declaring those as candidate regions. Another common approach is to use a sliding window method in which windows are extracted at various scales and positions without any prior knowledge of the size and location of the human object. A human classifier trained on the features extracted from the image database, labels the regions as humans or otherwise.

Convolutional Neural Networks (CNNs) are currently the most popular machine learning algorithms in computer vision for human detection [13, 14]. You Only Look Once (YOLO) [14], presented a new approach to object detection in which a single CNN predicts bounding boxes and class probabilities directly from full images in one evaluation. Being a single network, it can be optimized resulting in real-time processing of input images at 45 frames per second. At this detection rate (speed), it outperformed other detection methods, including Deformable Parts Model (DPM) and R-CNN. We employed this network for the outdoor set of experiments.

Gait Detection and Analysis

Model free set of approaches [15, 16, 17] have also been proposed in which gait is comprised of static components (size and shape) and a dynamic component of a person which reflects the actual movement. In gait recognition the static features are height, stride length and silhouette bounding box lengths. Dynamic features used are legs and head angles with centroid, velocity, acceleration and frequency of body parts contributing to the overall motion. Sarkar et al. [18] created the HumanID project which is considered a significant landmark in gait recognition. It was reported that the lower 20% of silhouette, that is, from the knee down, accounted for 80% of the recognition. BenAbdelkader et al. [17] extracted the bounding box of the extent of motion after tracking the subject in video steam. These scaled boxes, known as templates, were subtracted from each other pixel by pixel (spatially) for all given pairs of time instances (temporally). In their later study [19], stride length and cadence (walking frequency) was proposed to illustrate the human motion by segmenting and erecting a bounding box around it. Frequency of walk was estimated by noticing the change in size of the bounding box over time. Fujiyoshi and Lipton [20] analysed motion of a human by detecting its silhouette and boundaries to produce a star skeleton. Cyclic motion of skeletal segments was extracted to determine human activities such as walking or running. Although their technique is computationally inexpensive and applicable to real-world surveillance videos, it is restricted to static background and shadow less scenarios.

Temporal representation approaches model gait signatures by an XYT three dimensional space by accommodating time as the third dimension complementing the XY axes in the image plane. Niyogi and Adelson [21] formed an image cube by stacking images of a walking sequence (XY) over the time axis. They analysed the XT plane and traced out a unique braided pattern walker's ankle as compared to the linear moving pattern of head. This information was helpful to construct the stick model and to identify a walker in image sequence by gait. In another study [12], gait sequence was decomposed into XT slices to produce a periodic pattern termed Double Helical Signature through an iterative curve embedding algorithm. This representation highlighted body parts for gait recognition, typically in the context of surveillance. This work relies on availability of gait sequences extracted from the surveillance videos before estimating double helical signatures. In an automated approach, getting gait sequence from silhouette mask is a difficult task specially where shadow is accompanied by human movement.

Motivation

A competent observer can pick a disable person by appearance (if a visible mobility aid is in use) or by observing gait (in case of limping walk). Apart from the 3d information, he/she can also identify a disable person by monitoring the subject's movement in a monocular video of human walk. This study is intended for observing a distinguished feature in human walk which enabled automated detection of disable person from videos. There is a good chance that a time series of body part movement will depict abnormality sensed by human cognition. Following the approach presented by Fujiyoshi & Lipton [20], we examine temporal behaviour of head and leg positions.

This paper makes two contributions. First, it separates the joint gait signal produced by legs into two each corresponding a leg. In prior work [20, 11] a single gait signal was used but this obfuscates the individual action of legs. This new set of gait signal better illustrates the contribution of legs in human walking style. Second, we compare gait signals acquired by manual and computer assisted approaches for discerning features to detect a disable person.

Automated Approach

The topic of disability being sensitive, has less training and testing video data available on internet. We collected surveillance videos in August 2017, which resulted in fewer number of participants since majority of these people tend to stay at home and approximately 0.4% [22] appear on public places (in NZ). Collected videos involved healthy and disable person movements with the same camera angle and background. The automated task was divided into silhouette extraction, skeletonization and Gait analysis phases which are explained below.

Silhouette Extraction

It can be difficult to analyse the nature of human motion without extracting the silhouette. Since we target system deployment for outdoor environments, so a static background assumption may not hold true. Furthermore, motion detection techniques [7, 8] incorporate shadows as part of the moving objects which leads to incorrectly determined silhouette mask, therefore, we started with YOLO to localize pedestrians appearing in the footage. YOLO takes a video frame as input and after processing, provides bounding boxes with five predictions: centre of object (x,y), width (w), height (h), and confidence value. Inside YOLO, CNN performed the detection task with its initial convolutional layers extracting features from the image while fully connected layers predicted output probabilities and coordinates. An example detection by YOLO is shown in Fig. 1. The resulting bounding box serves as the region of interest (ROI) to be segmented into human silhouette. Inside this ROI, foreground subtraction was performed using GMM resulting in silhouette mask. Advantage of using YOLO is that it filters out a big part of shadow linked with the moving silhouette and reduces the data size for GMM computations. (Refer to Fig. 2 for GMM outcome.)

Fig. 1: Pedestrian detection using YOLO   Fig. 2: Pedestrian mask after GMM operation

Skeletonization and Gait Signal

To extract the gait signal, we used the approach suggested by Fujiyoshi & Lipton [20] which assumes a static background with simplified conditions. The difference in our experiment is that YOLO implementation preceded the GMM operation to get the pedestrian mask. After cleaning this mask by morphological operations, its outline was extracted using a border following algorithm (see Fig. 3(a)). The skeleton with two legs and a torso, was constructed by connecting these extremal points to the centroid of silhouette. To analyse the gait, legs and torso angles (Fig. 3(b)) with vertical axis were computed and analysed as time series data. The following gait features in temporal domain were observed in this research:

- Legs velocity and acceleration

- Head velocity and acceleration

- Legs and Head angle

Signal Split

It has been observed that the leg angle signal is contributed by both legs in an alternate fashion based on leading and trailing status. This makes it less reliable for examining individual leg movements over time. Moreover, locations/positions of legs can be estimated once we fragment the leg signal into its individual signals. Since monocular camera lacks stereo information, the legs are referred to as leg 1 and leg 2 instead of left and right. These splintered signals have improved indicators for anomalous walking behaviour. The process of signal split (for the target moving from right to left of the screen) is summarised as,

1. Extract leg angle data θ_1 and θ_2 [20]. Here, θ_1 and θ_2 refer to leading and trailing leg angles, respectively, both measured from the vertical axis.

2. Identify time instants (for $t = t_n$, where $n = 1, 2, 3, ...$) when both legs have angles closest to vertical axis. This occurs at the local minima of θ_1 so, with $S(t)$ as a logical array namely,

$$S(t) = \begin{cases} true & \text{if n is odd} \\ false & \text{otherwise} \end{cases}$$

3. Calculate new signals as,

$$\alpha_1 = \theta_1 S(t) + \theta_2(\neg S(t)) \tag{1}$$
$$\alpha_2 = \theta_2 S(t) + \theta_1(\neg S(t)) \tag{2}$$

4. Track extremal points for legs from newly generated signals to form the gait vector.

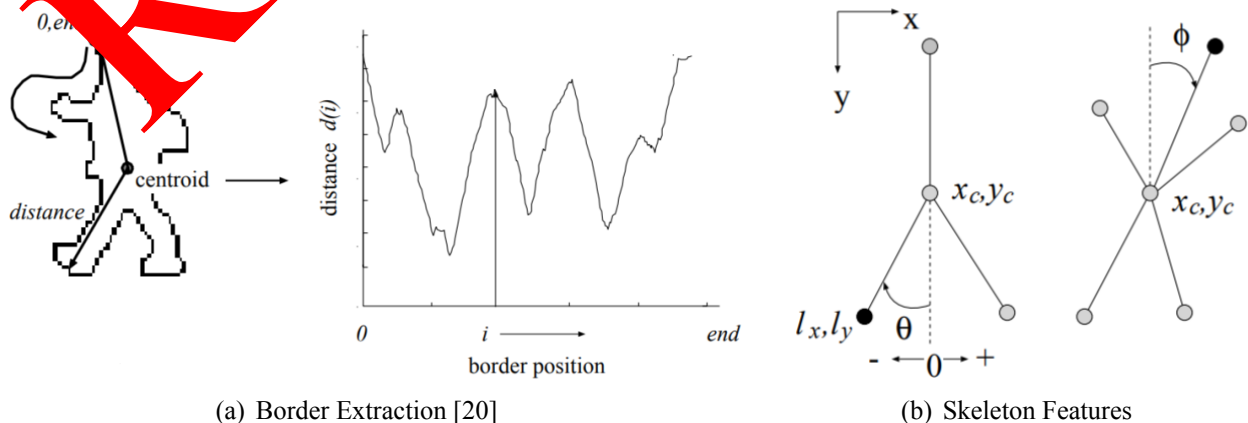(a) Border Extraction [20]   (b) Skeleton Features

Fig. 3: Star Skeletonization

Experimental Results and Discussion

Signal Extraction Results

Signal extraction was tested on indoor data acquired from The Institute of Automation, Chinese Academy of Sciences (CASIA-B) [23] gait dataset. Results (Fig. 4) show the successful operation for retrieving leg signals of a healthy person and has correlation of -0.9556 depicting highly symmetrical movement. Later on, the same experiment was repeated on videos for healthy and disable person movements in an outdoor environment to obtain leg positions with respect to time. Symmetry values of -0.9274 and -0.9174 were obtained for healthy and disabled person movements respectively. The negative sign in correlation is due to movement of legs in opposite directions. Output plots were not as smooth as those of indoor case and fluctuations were noted due to inaccuracies in pedestrian localization by YOLO and presence of shadows. This makes leg angles less reliable for detecting abnormal walking behaviour. According to YOLO authors, localization errors account for more of YOLO's errors than all other sources combined. These errors result in inaccurate centroid estimations and error propagates in head/legs angle computations.

Automated Results

In skeletonization scheme, leg angles were estimated using trigonometric functions applied on leg and centroid positions predicted by YOLO and GMM foreground mask. Although angular information lacked distinguishing features, it can help us analyse other motion features like velocity and acceleration of legs. The fact that lower 20% portion of human body contributes 80% in human recognition by gait [18], leg velocity/acceleration may contain useful gait information. Head angle and position could also be investigated for gait signal but our experiment revealed that head signal is too sensitive to fluctuations caused by rotations and minor head movements. Furthermore, head is a signal source for its gait signal while, leg data is contributed 2 sources (Leg 1 and Leg 2 in our case) making it less prone to inaccurate readings compared to the head data.

Results for velocity/acceleration data of healthy and disabled person were generated with automated approach. It was observed that both sets of graphs were noisy and it was hard to identify motion patterns for disabled person from healthy person. The question that, "is there a discriminating feature in gait signal that can help to categorise disable person from healthy?," remains unanswered. If yes, what caused the failure of automated approach for identifying gait feature of a disabled person? This lead us to manually segment the location of legs and head locations in video frames over time, and results are reported in Figs. 5 and 6.
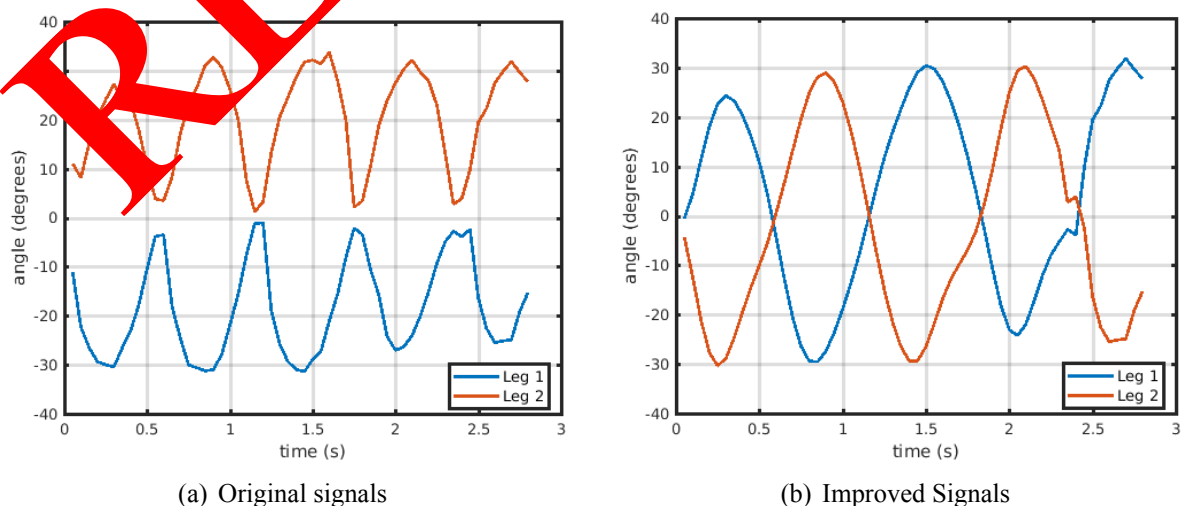

(a) Original signals   (b) Improved Signals

Fig. 4: Leg signals of Healthy Person in indoor environment

Manual Results

We marked locations of legs and head of moving person in video frames and by looking at their plots, there was a clear difference in the graph of the disable person. Graphs for disabled part (leg 2 in figs. 5(b) and 6(b)) clearly produces a signal for leg 2 which is different to that of leg 1. We considered average values of five peaks shown in Fig. 5 for these calculations. Legs of a healthy person have 1.45% difference (w.r.t larger value) in peak values while that of disabled person is 47.02%. Similar behaviour is observable in acceleration signals. This shows that Leg 2 of the disabled person is behaving differently than Leg 1 while walking, which means either of these legs is affected and not normal. Head signal was also exploited but it failed to give us encouraging results while centroid of the body is difficult to pick since it requires accurate localization and is sensitive to noise.

Conclusion

Gait recognition procedures are affected by the variation in environmental constraints, viewing angle, occlusions, shadows, imperfection in foreground modelling, object segmentation and silhouette
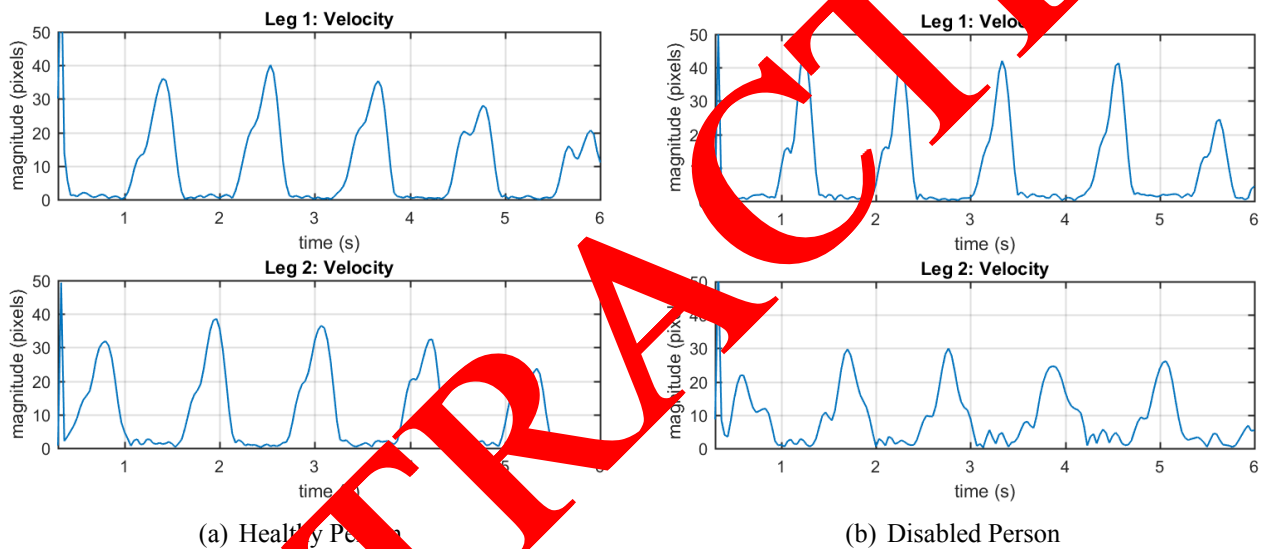

(a) Healthy Person   (b) Disabled Person

Fig. 5: Manual results for velocities

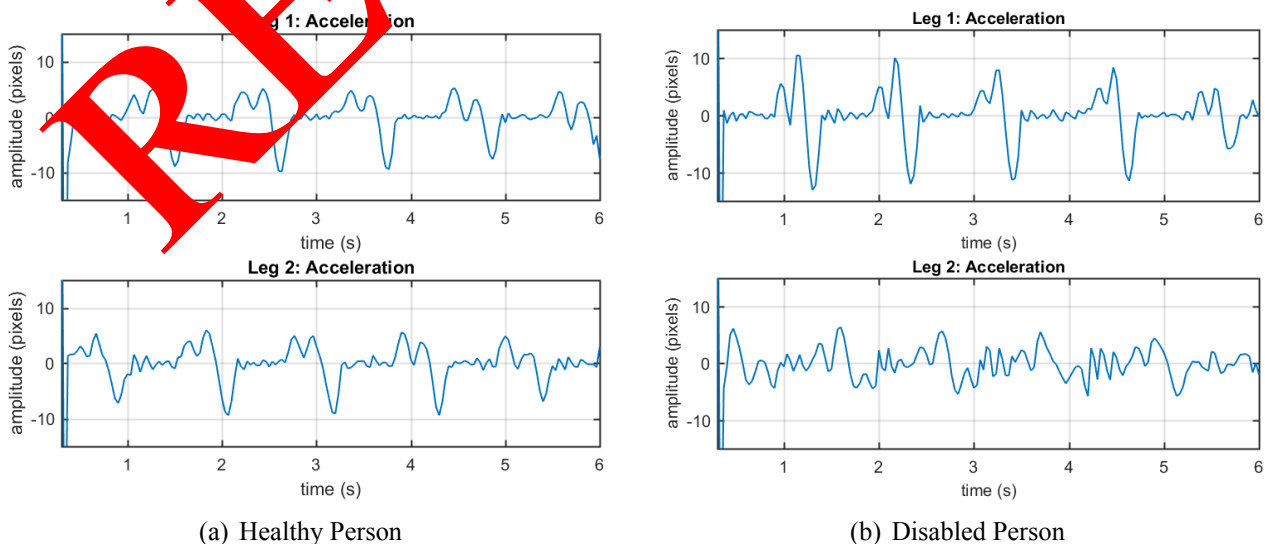(a) Healthy Person   (b) Disabled Person

Fig. 6: Manual results for acceleration

extraction. We experimented on video data of moving pedestrians to investigate presence of disabled person by automated and manual analysis of head and leg data. The automated system failed to recognise a disabled person from its gait due to imperfection in segmentation, but results from the manual localization suggest that there is enough information in the gait signal to characterize a healthy motion given a set of gait signatures. YOLO's localization error restricts the automated system to show impressive results and problem might be addressed by replacing YOLO with better localized pedestrian detection systems since exact locations of legs and centroid play a key role in shaping a gait signal. Furthermore, work on human joints detection can significantly improve gait signal which may lead to autonomous detection of disabled person. CNN based systems [24] have shown encouraging results in identifying joint locations and connecting them to other body parts of the same person. Such system have tendency to get better results than traditional gait techniques. We shall also be focusing CNN based solution to this problem along with identifying people using various mobility aids.

Acknowledgement

References

[1] Manoranjan Paul, Shah ME Haque, and Subrata Chakraborty. Human detection in surveillance videos and its applications-a review. *EURASIP Journal on Advances in Signal Processing*, 2013(1):1–16, 2013.

[2] Palwasha Afsar, Paulo Cortez, and Henrique Santos. Automatic visual detection of human behavior: A review from 2000 to 2014. *Expert Systems with Applications*, 42(20):6935–6956, 2015.

[3] Oluwatoyin P Popoola and Kejun Wang. Video based abnormal human behavior recognition—a review. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 42(6):865–878, 2012.

[4] Tracey K. M. Lee, Mohammed Belkhatir, and Saeid Sanei. A comprehensive review of past and present vision-based techniques for gait recognition. *Multimedia Tools and Applications*, 72(3):2833–2869, 2013.

[5] Dimitios Makris and Tim Ellis. Learning semantic scene models from observing activity in visual surveillance. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 35(3):397–408, 2005.

[6] Saed Shiry Ghidary, Yasushi Nakata, Toshi Takamori, and Motofumi Hattori. Human detection and localization at indoor environment by home robot. In *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, volume 2, pages 1360–1365. IEEE.

[7] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2. IEEE.

[8] Berthold K.P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1):185 – 203, 1981.

[9] Yannick Benezeth, Pierre-Marc Jodoin, Bruno Emile, Hélene Laurent, and Christophe Rosenberger. Comparative study of background subtraction algorithms. *Journal of Electronic Imaging*, 19(3):033003–033003–12, 2010.

[10] Imad Rida, Ahmed Bouridane, Gian Luca Marcialis, and Pierluigi Tuveri. Improved human gait recognition. In *International Conference on Image Analysis and Processing*, pages 119–129. Springer, 2015.

[11] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, 2003.

[12] Y. Ran, Q. Zheng, R. Chellappa, and T. M. Strat. Applications of a simple characterization of human gait in surveillance. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(4):1009–1020, 2010.

[13] Xingyu Zeng, Wanli Ouyang, Meng Wang, and Xiaogang Wang. Deep learning of scene-specific classifier for pedestrian detection. In *European Conference on Computer Vision*, pages 472–487. Springer, 2014.

[14] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788, 2016.

[15] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Kruger, and R. Chellappa. Identification of humans using gait. *IEEE Transactions on Image Processing*, 13(9):1163–1173, 2004.

[16] Ibrahim Venkat and Philippe De Wilde. Robust gait recognition by learning and exploiting sub-gait characteristics. *International Journal of Computer Vision*, 91(1):7–23, 2011.

[17] Chiraz BenAbdelkader, Ross Cutler, Harsh Nanda, and Larry Davis. Eigengait: Motion-based recognition of people using image self-similarity. In *Audio-and Video-Based Biometric Person Authentication*, pages 284–294. Springer.

[18] S. Sarkar, P. J. Phillips, Z. Liu, R. Vega, P. Grother, and K. W. Bowyer. The humanid gait challenge problem: datasets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005.

[19] C. BenAbdelkader, R. Cutler, and L. Davis. Stride and cadence as a biometric in automatic person identification and verification. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 372–377.

[20] Hironobu Fujiyoshi and Alan J Lipton. Real-time human motion analysis by image skeletonization. In *Proceedings of the 4th IEEE Workshop on Applications of Computer Vision (WACV'98)*, page 15. IEEE Computer Society, 1998.

[21] S. A. Niyogi and E. H. Adelson. Analyzing and recognizing walking figures in xyt. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 469–474.

[22] Bridget Burdett. Measuring accessible journeys. Report, Traffic Design Group, July 2013.

[23] Casia gait database, http://www.cbsr.ia.ac.cn/english/gait%20databases.asp.

[24] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1611.08050*, 2016.