Smart Annotation Tool for Multi-sensor Gait-based Daily Activity Data

Christine F. Martindale*, Nils Roth*, Julius Hannink*, Sebastijan Sprager† and Bjoern M. Eskofier*

*Machine Learning and Data Analytics Lab, Computer Science Department
Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany
Email: christine.f.martindale@fau.de, nils.roth@fau.de, julius.hannink@fau.de, bjoern.eskofier@fau.de
†Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia
Email: sebastijan.sprager@fri.uni-lj.si

Abstract—The monitoring of patients within a natural, home environment is important in order to close knowledge gaps in the treatment and care of neurodegenerative diseases, such as quantifying the daily fluctuation of Parkinson's patients' symptoms. The combination of machine learning algorithms and wearable sensors for gait analysis is becoming capable of achieving this. However, these algorithms require large, labelled, realistic datasets for training. Most systems used as a ground truth for labelling are restricted to the laboratory environment, as well as being large and expensive. We propose a study design for a realistic activity monitoring dataset, collected with inertial measurement units, pressure insoles and cameras. It is not restricted by a fixed location or capture volume and still enables the labelling of gait phases or, where non-gait movement such as jumping occur: on-the-ground, off-the-ground phases. Additionally, this paper proposes a smart annotation tool which reduces annotation cost by more than 80%. This smart annotation is based on edge detection within the pressure sensor signal. The tool also enables annotators to perform assisted correction of these labels in a post-processing step. This system enables the collection and labelling of large, fairly realistic datasets where 93% of the automatically generated labels are correct and only an additional 10% need to be inserted manually. Our tool and protocol, as a whole, will be useful for efficiently collecting the large datasets needed for training and validation of algorithms capable of cyclic human motion analysis in natural environments.

I. INTRODUCTION

Gait analysis and activity recognition in a home environment is made possible with the use of wearable sensors as they are small and unobtrusive [1], [2]. Many popular machine learning methods, such as deep learning, for activity recognition and gait analysis based on wearable sensors require large, annotated datasets for training and validation [3]. The annotation cost for producing such large datasets is high if gait phases and cycle segmentation is important, such as in the case of home monitoring of patients.

One method for producing labels for such a dataset is to use an alternative system as ground truth. Prominent examples are motion capture systems or pressurized carpets [4]. Both systems suffer from the limitations of a restricted capture volume, high cost, being immobile and laboratory based. These restrictions are problematic when home based solutions are required.

A less restrictive solution is using cameras and manual labelling, however it is time consuming and often inaccurate for gait analysis purposes. Alternatively, one can combine multiple sensors for labelling and providing an algorithm based suggestion for labels which are then manually checked and corrected, where necessary. This is often termed smart annotation or assisted labelling [5], [6]. One advantage of assisted labelling is that the annotators need not be sensor or algorithm experts as the initial labels are suggested and the labeller merely confirms if the label is correct and if not, adjusts it. This adjustment is also often suggested. The labellers can also be assisted with camera information and user information, as well as different visualisations of the sensor data [7]. This can even be crowd sourced or analysed online by domain experts [8]. This can even halve the annotation time and increase the accuracy of the subsequent labels, as shown in the smart environment context by Szewcyzk et al. [7].

Algorithms such as dynamic time warping (DTW), as well as support vector machines, are also, often used for assisted labelling [5], [8]. Within the industry based activity recognition field, cameras and inertial measurement units (IMUs) were combined to produce smart annotation tools [5], [9]. Diete et al. [5] used the initial labels from the annotator to generate a template for DTW and allowed the annotator to decide at what point to automate the remainder of the labelling based on the generated template. Whereas Barz et al. [9] focussed on integrating many commonly used commercial sensors into their tool, such as Myo [10] for electromyography.

Even with single sensors, smart annotation can suggest labels which the annotator adjusts as needed, so reducing annotation time. This was demonstrated by Orazio et al. [11] with their video based soccer annotation tool where the labeller was shown the labels generated by a computer vision based player identification algorithm. Incorrectly segmented or identified players were adjusted and the annotation tool was evaluated based on the time required to label identical datasets [11].

The best solution would be to move away from the need for fully labelled datasets with the use of semi-supervised learning algorithms. However, large, labelled datasets are still needed to validate these algorithms. While there are many such datasets for activity recognition where the labels are given for whole

activities, few exist for gait or human motion cycle analysis which are needed to validate these semi-supervised algorithms. Some datasets give gait phase labels for a limited number of patients, however, often only within a limited capture volume due to the use of motion capture as the ground truth [12]. One very large dataset exists which was manually labelled on a gait cycle level, using a single camera, for waist mounted IMUs [13]. Although it is large with over 700 participants, it is limited to under 20 steps per person. To the best of the authors' knowledge, none exists for gait analysis within a natural, unrestricted environment (e.g. not treadmill based or restricted to the capture volume of a motion capture system).

This paper aims to provide a protocol and pipeline to close this gap. We provide an initial dataset with IMUs at popular locations on the body, cameras and pressure insoles. The activities are goal orientated and as natural as possible, with a focus on cyclic activities such as gait and jumping. We also present an edge detection method to segment on-the-ground, off-the-ground phases using insole pressure data and compare this to a threshold based method and a commercial sensor. Finally, we combine these two contributions in a smart annotation tool to assist the labelling of a multi-sensor dataset to label on-the-ground, off-the-ground phases of activity data. We describe the above methods and annotation tool and then compare the methods using a manually corrected set of labels as the ground truth.

II. METHODS

A. Dataset

We propose a dataset design focusing on cyclic activities measured using IMUs, pressure sensor based insoles and video. The pressure data is used to detect on-the-ground, off-the-ground phases and the camera to detect overall activity labels. The data prepared and used to evaluate the smart annotation tool proposed in this paper was collected from 20 healthy participants with the following characteristics: 5 females and 15 males, with an average age of 28 ± 7 years, an average height of 175 ± 6 cm and weight of 74 ± 9 kg. The shoe sizes were limited to the range of 38 to 44 due to the available insole sizes.

Each participant wore IMUs, the same Bosch development platform as used in [14], mounted on the lateral side of each shoe, one on each wrist and, where possible, one in a trouser pocket as shown in Figure 1. Acceleration (± 8 g) and angular velocity (± 2000 dps) were recorded with a frequency of 200 Hz. The sensors were synchronised by a simultaneous, Bluetooth based, clock reset and start command. Each participant wore Moticon pressure sensor insoles, which have previously been compared to GaitRite under constrained conditions [15]. In this paper we further use the insoles in conditions in which other validation systems would be impractical. For this dataset, the Moticon Science Software version 01.10.00 was used [16]. Data was recorded with 5 pressure sensors and a 3-axis accelerometer at 100 Hz with the recommended automatic zeroing activated. The dataset is available at *www.activitynet.org*.
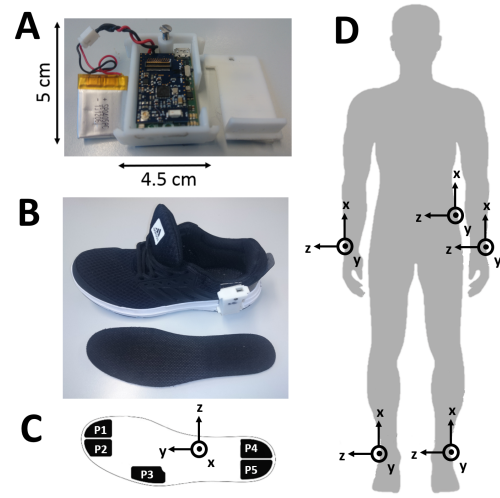

Fig. 1. Sensor locations. A. Photograph of IMU sensor system in 3D printed case. B. Picture showing sensor attachment to shoe, using industrial velcro and Moticon insole, which was used instead of the original sports shoe insole C. Pressure sensor location and approximate size and shape within the insole. Also showing axes location for the insole accelerometer. D. IMU sensor locations on the body with corresponding axes.

The participants were asked to perform a series of self-paced and task driven activities such as: "Please run as if you were late for the bus" or "Please jump twenty times". If the activity began incorrectly, it was simply repeated; if the number of iterations was roughly correct, the task was taken as complete. The protocol consisted of the following activities in a randomised order per participant:

- gait based activities including short bouts (writing on 5 posters, spaced 2 m apart), straight walking, jogging and running (2 times 20 m, including a 180 degrees turn), non-straight walking (zigzagging between three tables), standing (while sorting cards between 3 tables) and sitting (three games at different three desks spaced 3 m apart).
- continuous transitions between walking, running and jogging (6 transition possibilities, order randomised).
- cyclic motion activities: jumping, jumping jacks, skipping, stepping, hopping, side steps, jumping over a line, zigzagging between 5 cones spaced 2 m apart and high kick running.

The data was recorded as three continuous recordings, including mistakes. Between each activity there were varied amounts of walking to reach each location. The first and last sections were performed indoors between a classroom and the adjacent hallway, while the middle section was performed outdoors and therefore included stair climbing to and from the second floor. All sensors were reset and restarted between sections. The dataset consisted of about 30 minutes of data per person, totalling more than 1 000 cycles per person per foot.

Multiple cameras were used to minimise occlusion: at least one static camera per room and one hand-held camera focused on the participant's feet. As the cameras were used to label activity data and verify other labels in combination

with the other available sensors, small sections of occlusion could be tolerated. The video recordings were synchronised to the Moticon insoles using a QR code displayed within the Moticon software at the start and end of each recording section. Moticon insoles were synchronised to the IMU system using cross correlation of the y-axis accelerometer data. We assume here that the acceleration of the sensor within the insole and that of the IMU attached to the lateral side of the same shoe are similar enough that cross correlation allows sufficiently accurate synchronisation. As a final check, the participants performed 3 jumps at the start and end of each section; these were used to correct video synchronisation, if needed.

B. Algorithms for Annotation of Pressure Data

The main aim of the dataset was to provide gait data in a relatively natural setting. The gait cycles can be detected from the pressure insoles and used to label the rest of the data, assuming the systems are synchronised. While the Moticon system did provide pressure information, it did not have a validated step detection algorithm for 100 Hz data at the time of the experiment. Furthermore, commercial sensor systems often do not have open source algorithms and do not necessarily have publicly available validation for each software update. Therefore, we chose to export the raw data from the insoles and perform the step detection within the annotation tool. A simple threshold method was insufficient due to a baseline drift with time over the 30-minute data collection, even when using the recommended automatic zeroing within the Moticon Software.

In this paper, we refer to cycles being divided into on-the-ground, off-the-ground phases due to the data including other cyclic activities such as jumping as well as gait. For gait this simply refers to the stance phase and swing phase respectively. Examples of the raw data with phases labeled are given in Figure 2. Using the raw pressure data from the insoles, we compared the following methods to detect on-the-ground, off-the-ground phases of the motion data:

1) EdgeDet1: Detection of rising and falling edges of the individual pressure sensors.
2) EdgeDet2: Detection of rising and falling edges of the all 5 pressure sensors' profile (the maximum of a sensor per unit time per foot).
3) Threshold: De-drift using envelope subtraction, followed by detection of threshold crossings.
4) Moticon: Moticon algorithm based on 100 Hz data.

The main assumption here is that the start of on-the-ground phases is where the pressure on the insole rises significantly, above a baseline, causing a rising edge; and vice versa for the falling edges and the end of the pressure phase per cycle. For evaluation purposes, an expert examined the results of the first method, within the smart annotation tool, and manually adjusted any incorrect labels based on the synchronised visualisation of all the sensor data. These methods were then assessed by calculating what percentage of the generated labels required manual adjustment.

Edge Detection: The aim of this method is to detect the rising and falling edges of the insole pressure data, which represent the ground contact phases of the feet. The raw pressure data was filtered using a low pass third order Butterworth filter with a cut off frequency of 20 Hz. This was chosen to smooth the data by removing high frequency noise, as human motions fall into the lower frequency ranges. All pressure values were normalised because only the relative pressure is important when detecting the rising and falling edges. The derivative of the filtered pressure data was calculated using the finite difference equation with 5 sample points [17]. The positive peaks and the negative peaks of this derivative were detected using Matlab's *findpeaks* function, with the following restrictions: minimum prominence of 0.1, minimum height of 0.1 and minimum peak distance of 100 ms. These parameters were found empirically, and we believe they are dependant on the quality of the pressure data, i.e. the more prominent and sharp the edges are, the less the restrictions are influential and empirically dependant. The prominence of the peaks depends on noise and data quality, therefore this peak picking algorithm was also chosen empirically.

Individual sensor failure, noise and other artefacts still gave rise to falsely detected falling and rising edges detected for each pressure sensor. A rule based approach was applied, in the order given below, to filter out the unimportant edges. They are illustrated in Figure 3. If two consecutive edges are found
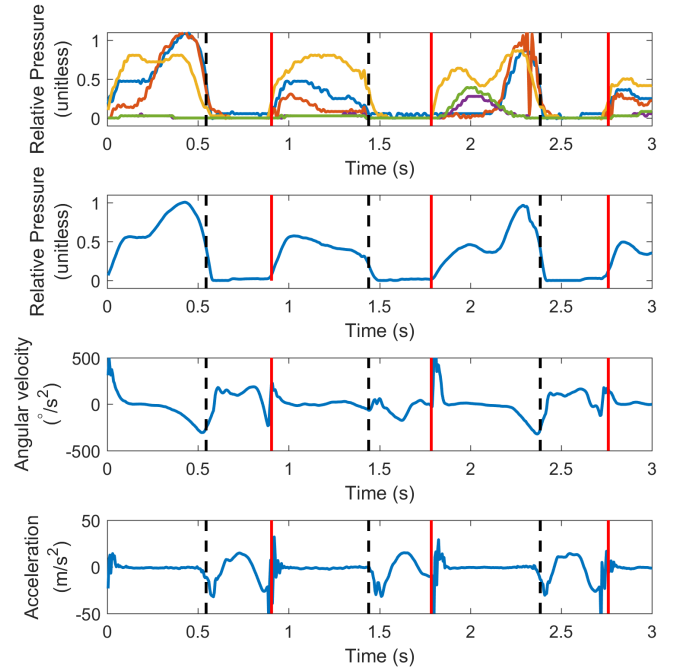

Fig. 2. Example data of stepping exercise. Top: Relative pressure data for each of the 5 pressure sensors (P1 in blue, P2 in red, P3 in yellow, P4 in purple, P5 in green). MiddleUpper: Relative pressure data as a profile of all 5 pressure sensors. MiddleLower: Angular velocity in the Z-axis Bottom: Acceleration in the Y-axis. Vertical red solid lines show the rising edge of the pressure data, i.e. start of the on-the-ground phase. Vertical dashed black lines show the falling edge of the pressure data, i.e. end of the on-the-ground phase.
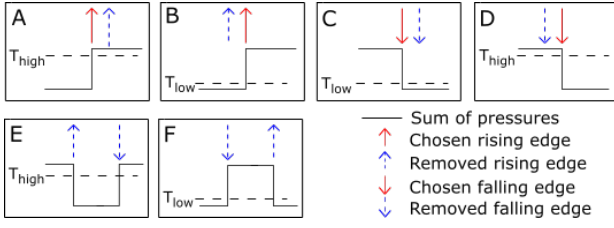
Fig. 3. Rules for post processing of the edge detection algorithm. Dashed arrows represent removed edges, solid arrows represent kept edges. The black line represents the sum of pressures.

within 2 seconds of each other then:

1) Remove the latest rising edge if it is within 50 ms of it neighbour and remove the earliest falling edge if it is within 50 ms of its neighbour.
2) When two rising edges occur consecutively, remove the second rising edge if the maximum of the sum of pressure between these edges is above the upper threshold, T_{high} (Figure 3 A).
3) When two rising edges occur consecutively, remove the first rising edge if the minimum of the sum of pressure values between these edges is below the lower threshold, T_{low} (Figure 3 B).
4) When two falling edges occur consecutively, remove the second falling edge if the minimum of the sum of pressure values between these edges is below the lower threshold, T_{low} (Figure 3 C).
5) When two falling edges occur consecutively, remove the first falling edge if the maximum of the sum of pressure values between these edges is above the upper threshold, T_{high} (Figure 3 D).

Once this is performed then loop through the edges again, comparing consecutive edges and remove:

1) Opposing edge pairs if the maximum of the sum of pressure values between rising edge and the falling edge is below the upper threshold, T_{high} (Figure 3 E).
2) Opposing edge pairs if minimum of the sum of pressure values between falling edge and rising edge is above the lower threshold, T_{low} (Figure 3 F).
3) The latest of two consecutive falling edges.
4) The earliest of two consecutive rising edges.

The thresholds used were found empirically, T_{high} was set to 0.3 and T_{low} to 0.1. This edge detection method was applied to all 5 individual pressure sensors (EdgeDet1) and to the combined profile of the pressure data (EdgeDet2).

De-drifting and Threshold: Due to the drift of the baseline of the pressure data, an envelope of the data was found by identifying the negative peaks of the profile of the pressure data per foot. These minimums were found using Matlab's *findpeaks* algorithm with the following restrictions: minimum peak distance of 250 ms, peak height of 0.04 and a minimum prominence of 5. This envelope was then set as the zero line and a threshold was chosen. The positive crossings and negative crossings of this threshold were then taken as the segmentation of the on-the-ground, off-the-ground phases.

C. Smart Annotation Tool

The aim of the proposed pipeline is to automate, as far as practically possible, the labelling process for a dataset focusing on cyclic human motions captured by wearable sensors. We fuse information from all three sensor types in order to provide assisted labelling. To achieve this, we have developed a MATLAB based tool, the interface of which is shown in Figure 4. The pipeline used within the tool can be split in the following sections:

1) Synchronise video to insole data.
2) Import protocol (participant specific task list).
3) Synchronise shoe-mounted IMU data to insole data via cross correlation of accelerometer signals.
4) Extract on-the-ground and off-the-ground phases from insole pressure data using edge detection technique.
5) Manually adjust incorrect labels using suggested possible corrections based on local minima and maxima.
6) Automatically combine phases and activity labels by including recognition of standing.

It is described as a smart annotation tool because it initially suggests all the activity and phase labels to the annotator who then corrects any erroneous labels with the help of automated suggestions. The annotator can also force a label if the suggestions are still incorrect. Each of the above stages in the annotation process is described in the following section.

Synchronise Video: All raw data (IMUs, pressure and video) were imported into the tool. The synchronisation between the pressure data and videos are extracted from the Moticon software and imported. This synchronisation was performed by recording the QR code given in the Moticon software for 5 seconds at the beginning and end of each recording. Any needed adjustments to the synchronisation between video and pressure are highlighted by inspection of the synchronisation jumps performed by the participant at the beginning and end of each section. These were also used to identify possible miscalibrations or failure to synchronise.
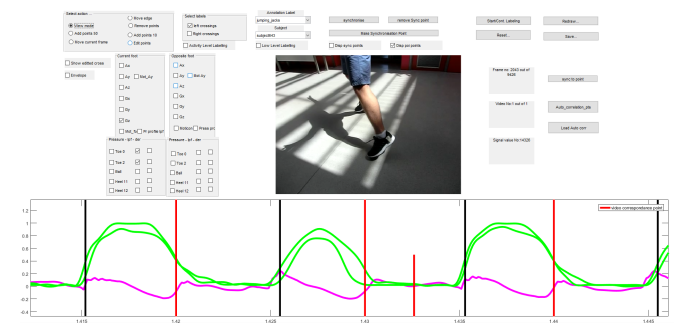

Fig. 4. Interface of smart annotation tool which allows the annotator to display, in a synchronised manner, the video frame plus all sensor signals (e.g. pressure, gyroscope and accelerometer). The annotator can also adjust synchronisation and labels. Shown here is a subject performing jumping jacks with toe pressure sensors (P1 and P2, green) and gyroscope (pink) from the left shoe. The vertical lines correspond to the start (black) and end (red) of on-the-ground, off-the-ground phases.

Import Protocol: The study protocol was imported (randomised activity order per participant) and an initial suggestion for activity labels was created by uniformly dividing the experiment time by the number of activities, including the insertion of transition periods between activities. These labels were then manually corrected using the video data.

Synchronise IMU and Insoles: IMU data is synchronised to the pressure sensor data by piecewise cross-correlation of the accelerometer signal in the y-axis. The Moticon based accelerometer data is resampled using a range of frequency ratios between 1.4 and 3.3 with a resolution of 0.001, where the expected ratio is 2.0 due to Moticon having a sampling frequency of 100 Hz and the IMU of 200 Hz. The resampling frequency which achieves the highest cross correlation value is then selected. Using this resampling frequency, the time difference between signals is then calculated. These are then used to display the synchronised pressure and IMU data. IMU to IMU synchronisation is performed based on initial clock reset and the time of the start command sent using a single packet Bluetooth command. These are again checked for any obvious errors by inspecting the synchronisation jumps.

Extract Cycle Phases: The pressure data was found, empirically, to have a baseline drift which increased over time, therefore a simple threshold method for gait phase segmentation was insufficient for accurate labelling. Two alternative methods were investigated and implemented to segment the on-the-ground, off-the-ground cycle phases: edge detection and baseline subtraction plus threshold. These were described in detail in section II-B.

Manual Adjustment: The labels generated by the chosen method were then displayed and, based on the visualisation of the combination of all sensor data, erroneous on-the-ground, off-the-ground phase labels were found and adjusted. The labeller was then able to perform the following manual corrections:

- Delete labels.
- Add labels which snapped to the local maximum or minimum of the derivatives of the pressure data, within a 250 ms window.
- Add labels which snapped to the local maximum or minimum of the derivative of the pressure data, within a 50 ms window.
- Shift the labels earlier or later within a 250 ms window of the previous label.

These manual adjustments assumed that the edges should be detected within the pressure data.

Combine Phase and Activity Labels: On-the-ground, off-the-ground phases plus the energy of the acceleration signal were used to detect standing or no movement phases and combine these with the activity labels. Any erroneous data could also be removed manually. The energy threshold was adjusted per person.

III. RESULTS & DISCUSSION

Each algorithm for assisted annotation was compared to the manually corrected version of the edge detection method applied to the 5 pressure sensors' data (EdgeDet1). A label was considered to be a false negative if there was no generated label within 50 ms of the manual label. It was considered to be a false positive if there was a generated label but no manual label within 50 ms. It was considered to be a true positive if the generated and manual labels were within 50 ms of each other. These were calculated as an average of the percentage of the total number cycles found by the algorithm per person. The combination of the percentage of false positives and false negatives can be seen as the effort required in labelling, i.e. the percentage of cycles which still needed manual input. The F1-score for each algorithm versus the manually adjusted labels was also calculated.

For the edge detection method using all 5 pressure sensors separately (EdgeDet1), 17.2% of the strides needed to be manually changed, as shown in Table I. This is much less effort than labelling 100% of the dataset. The percentage of true positives, labels correctly generated by the algorithm and so needing no manual adjustment, was 93.0%. The remaining 7% were incorrect, false positives. After this an additional 10%, relative to the total number of initially generated labels, needed to be inserted (false negatives). The manual labels refer to the labels checked and adjusted by an expert using the smart annotation tool. This is also reflected in the F1-score of 91.6%.

The second method (EdgeDet2) performed worse, requiring roughly 33 % to be manually corrected, although only 16.9% of the found labels were incorrect.

The final proposed method, involving zeroing and thresholding (Threshold), performed similarly, with the number of edits required reaching 29%. It achieved a false negative rate of 7% and a false positive rate of 21% meaning that although it detected too many labels, it only missed 7%.

These figures are influenced by the fact that the first two methods are similar to each other and find the middle of the slope of an edge, whereas the threshold method will always find the lower end of an edge. Therefore, the results for the threshold method could be considered as conservative, and are influenced by the slew rate of the change in pressure between cycle phases. Furthermore, all three proposed methods used the same peak detection algorithm. One should consider this when applying these to data collected by different sensors or conditions, where another peak picking algorithms could be advantageous.

Finally, the cycles given by the Moticon software were

TABLE I
COMPARISON OF ALL METHODS, SHOWN AS AVERAGE PERCENTAGE OF GENERATED LABELS. AVERAGE TOTAL NUMBER OF LABELS WAS: 4029, 4106, 4806, 3467 RESPECTIVELY.

Value (%)	EdgeDet1	EdgeDet2	Threshold	Moticon
Total changed	**17.2**	**33.1**	**28.5**	**73.8**
True positives	93.0	83.8	78.7	75.6
False positives	7.0	16.2	21.3	24.2
False negatives	10.1	16.9	7.2	49.6
F1-score	91.6	83.5	84.7	67.2

correct, within a 50 ms window, 75.6% of the time; although the software warned that they were not tested for 100 Hz data. The Moticon labels showed a 49.6% false negative rate. When the required accuracy of the labels was relaxed to 100 ms, then a true positive rate of 87.6% was achieved with only 12.2% which needed to be deleted (false positives). With this higher error tolerance, the results were reasonable, however, if all cycles should be detected then still 37.6% more labels would need to be added.

These results could be improved if the thresholds within the algorithms were adjusted per person or per activity, based on within which activities the cycles were best detected. The sample size used for these results is relatively small, and so difficult to draw conclusions across different populations, however it is sufficient to assess the feasibility of the smart annotation tool and experiment design.

While all the methods considerably reduced the labelling effort with an error tolerance of up to 50 ms, manually labelling 10% of the data is still a large effort. However, the proposed protocol and smart annotation tool are intended to assist the generation and labelling of validation datasets which include realistic gait and cycle data. Such a benchmark dataset would enable testing and further development of algorithms which are robust within realistic conditions. The use of cameras, IMUs and pressure sensors as well as a set protocol are needed for dataset which are to be used to validate algorithms, however not all sensors would be needed for deployable solutions. Such solutions could exploit the existence of several sensors, or test the limits of minimal sensors.

In future work, one could use this smart annotation tool and the current dataset to include sub phases of steps or cycles. The main challenge with this would be the need to define these sub phases, especially for non gait-based activities. The edges for gait sub phases are not as sharp as the on-the-ground, off-the-ground phases and so there is a need to define at what point of a rise or fall constitute the beginning of the phase.

IV. Conclusion

The proposed pipeline provides a method for data collection which allows semi-automated labelling and an efficient method of collecting daily activity data labelled on a cycle level. Although it was still in some ways a controlled setup, the same principles could be applied to more natural or specific applications. An edge detection method was proposed to detect on-the-ground, off-the-ground stride phases with only 17% manual labelling or correction required. These two contributions were combined in a smart annotation tool which is being used as the basis of an ongoing large dataset collection. This means that labelling time was reduced by 83%, versus complete manual labelling, without the assistance of the smart annotation tool.

We aim to use this methodology to label large volumes of IMU data to produce a benchmark dataset which will be useful for validation of gait events occurring within natural settings. We believe the dataset plus annotation tool will be useful for validating IMU based home monitoring algorithms or those

designed for 'in the wild' use. Furthermore, we plan to use it to train and validate semi-supervised, or even unsupervised algorithms for segmentation and classification of cyclic human motion using wearable sensors.

References

[1] J. M. Fisher, N. Y. Hammerla, T. Ploetz, P. Andras, L. Rochester, and R. W. Walker, "Unsupervised home monitoring of parkinson's disease motor symptoms using body-worn accelerometers," *Parkinsonism & Related Disorders*, vol. 33, no. Supplement C, pp. 44 – 50, 2016.

[2] S. Chen, J. Lach, B. Lo, and G.-Z. Yang, "Toward Pervasive Gait Analysis With Wearable Sensors: A Systematic Review," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 6, pp. 1521–1537, nov 2016.

[3] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in Bioinformatics*, p. bbx044, 2017.

[4] J. F.-S. Lin and D. Kulic, "Online Segmentation of Human Motion for Automated Rehabilitation Exercise Analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 168–180, jan 2014.

[5] A. Diete, T. Sztyler, and H. Stuckenschmidt, "A smart data annotation tool for multi-sensor activity recognition," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, mar 2017, pp. 111–116.

[6] C. Liu, W. T. Freeman, E. H. Adelson, and Y. Weiss, "Human-assisted motion annotation," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.

[7] S. Szewcyzk, K. Dwan, B. Minor, B. Swedlove, and D. Cook, "Annotating smart environment sensor data for activity learning," *Technology and Health Care*, vol. 17, no. 3, pp. 161–169, jan 2009.

[8] Z. Palotai, M. Lang, A. Sarkany, Z. Toser, D. Sonntag, T. Toyama, and A. Lorincz, "LabelMovie: Semi-supervised machine annotation tool with quality assurance and crowd-sourcing options for videos," in *2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI)*. IEEE, jun 2014, pp. 1–4.

[9] M. Barz, M. M. Moniri, M. Weber, and D. Sonntag, "Multimodal multisensor activity annotation tool," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp '16*. New York, New York, USA: ACM Press, 2016, pp. 17–20.

[10] Myo Gesture Control Armband. [Online]. Available: https://www.myo.com/

[11] T. D'Orazio, M. Leo, N. Mosca, P. Spagnolo, and P. Mazzeo, "A Semi-automatic System for Ground Truth Generation of Soccer Video Sequences," in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, sep 2009, pp. 559–564.

[12] F. Kluge, H. Gaßner, J. Hannink, C. Pasluosta, J. Klucken, and B. Eskofier, "Towards Mobile Gait Analysis: Concurrent Validity and Test-Retest Reliability of an Inertial Measurement System for the Assessment of Spatio-Temporal Gait Parameters," *Sensors*, vol. 17, no. 7, p. 1522, jun 2017.

[13] T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, "Similar gait action recognition using an inertial sensor," *Pattern Recognition*, vol. 48, no. 4, pp. 1289 – 1301, 2015.

[14] C. Martindale, F. Hoenig, C. Strohrmann, and B. Eskofier, "Smart Annotation of Cyclic Data Using Hierarchical Hidden Markov Models," *Sensors*, vol. 17, no. 10, p. 2328, oct 2017.

[15] B. J. Braun, N. T. Veith, R. Hell, S. Döbele, M. Roland, M. Rollmann, J. Holstein, and T. Pohlemann, "Validation and reliability testing of a new, fully integrated gait analysis insole." *Journal of foot and ankle research*, vol. 8, p. 54, 2015.

[16] Moticon science. [Online]. Available: http://www.moticon.de/science/

[17] Finite Difference Coefficients Calculator. [Online]. Available: http://web.media.mit.edu/ crtaylor/calculator.html