

This item is the archived peer-reviewed author-version of:

Outlier detection for foot complaint diagnosis : modeling confounding factors using metric learning

Reference:

Booth Brian, Keijsers Noel L.W., Sijbers Jan.- Outlier detection for foot complaint diagnosis : modeling confounding factors using metric learning
IEEE intelligent systems / Institute of Electrical and Electronics Engineers [New York, N.Y.] - ISSN 1541-1672 - 36:3(2021), p. 41-49
Full text (Publisher's DOI): <https://doi.org/10.1109/MIS.2020.3046431>
To cite this reference: <https://hdl.handle.net/10067/1753670151162165141>

Outlier detection for foot complaint diagnosis: modeling confounding factors using metric learning

B. G. Booth (brian.booth@uantwerpen.be)

imec-Vision Lab, Department of Physics, University of Antwerp, Antwerp, Belgium

Noël L. W. Keijsers

Department of Research, Sint Maartenskliniek, Nijmegen, The Netherlands

Jan Sijbers

imec-Vision Lab, Department of Physics, University of Antwerp, Antwerp, Belgium

Abstract—Diagnosing foot complaints using plantar pressure videos is complicated by the presence of confounding factors (e.g. age, weight). Outlier detection could help with diagnosis, but these confounding factors result in data that is not independent and identically distributed (IID) with respect to a specific patient. To address this non-IID problem, we propose the modeling of confounding factors using metric learning. A distance metric is learned on the confounding factors in order to model their impact on the plantar pressures. This metric is then employed to weight plantar pressures from healthy controls when generating a patient-specific statistical baseline. Statistical parametric mapping is then used to compare the patient to this statistical baseline. We show that using metric learning reduces variance in these statistical baselines, which then improves the sensitivity of the outlier detection. These improvements in outlier detection get us one step closer to accurate computer-aided diagnosis of foot complaints.

Keywords— computer vision, medicine and science, machine learning, computer-aided diagnosis, plantar pressure.

■ **IN DIAGNOSTIC MEDICINE**, it is well-established that a health condition can often produce different symptoms in different people.

Therefore, it can be challenging or inappropriate to model a patient population as a single homogeneous group. Instead, an outlier detection

approach is preferred. This is equally the case when assessing foot complaints [2]. When assessing foot complaints, a gait analysis is performed with pressure-sensing plates [11], wearable sensors [1], or cameras [13] recording the patient's walk. In this work, we consider videos of the plantar pressures measured from the bottom of a person's foot as they walk over a pressure-sensing plate. While previous works have described anomalies from these plantar pressures at the group level, it is also known that different patients have different abilities to cope with foot complaints [11], suggesting that plantar pressure anomalies are likely patient-specific. Outlier detection is therefore desired to identify whether anomalies exist in an individual's plantar pressure video, and if so, when and where. By localizing these abnormalities, it is hoped that a better diagnosis of foot complaints can be achieved [2].

One popular approach to outlier detection in the medical domain is to statistically model healthy controls, then use this model as a baseline to which individual patients can be compared [6]. Patients that are significantly different from this baseline are identified as having an anomaly. For medical imaging applications like our plantar pressure video analysis, outlier detection is regularly combined with statistical parametric mapping (SPM) to localize anomalies in both space and time [2], [4]. SPM-style outlier detection typically follows an established workflow. First, the plantar pressure videos brought into spatiotemporal alignment [3]. At each pixel in each frame, the plantar pressures are then statistically modeled using Normal distributions, resulting in a statistical baseline defined by a mean plantar pressure video and a standard deviation video. Subsequently, a patient's plantar pressure video is spatiotemporally aligned to the baseline's mean video, and single-sample t-tests are then computed as the outlier scores at each pixel. Finally, Random Field Theory is used to identify whether the patient's plantar pressures are statistically significant outliers of the baseline distributions [10]. This SPM-style outlier detection allows for the localization of anomalies to specific anatomical structures and specific time points in the footstep.

While this approach to outlier detection has shown promising results in other applications [4], one property that limits its application to plan-

tar pressure videos is that the computation of mean and standard deviation videos assumes each healthy control is independently and identically distributed (IID) from the same population as the patient under examination. Unfortunately, this is not the case for plantar pressures. It has been well established that demographic factors like weight, age, sex, height, and shoe size, impact plantar pressures in ways that are unrelated to known foot complaints [7]. As a result, we have a contextual outlier detection problem where the demographic features define the context under which plantar pressures normality should be judged. While multiple contextual outlier detection techniques exist [2], [8], [15], they are either limited to linear regression models [2], or have yet to show compatibility with the SPM framework. Specifically, SPM requires that the contextual outlier detection algorithm produce a statistical measure (e.g. t-statistic, F-measure) as its outlier score. That way, Random Field Theory can still be used to establish the threshold at which a pixel becomes an outlier.

In this paper, we propose the integration of contextual outlier detection into SPM for the purpose of detecting outliers in plantar pressure videos. We base our approach on the idea that the plantar pressures from our non-IID healthy controls should be weighted based on the demographic similarities that those control subjects have with the patient under examination. As in [15], these similarities are modeled using metric learning in order to manage the relative importance of each demographic factor on the statistical baseline. We hypothesize that this metric learning approach will produce more numerically accurate statistical baselines to which patients can be compared, resulting in more reliable anomaly detection.

METHODS

Consider a set of plantar pressure videos $\{\mathbf{V}_1, \dots, \mathbf{V}_N\}$ sampled non-IID from N healthy controls. For each healthy control, we also assume that we have measured the confounding factors of age, sex, weight, height, and shoe size. Let $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ be column vectors containing these demographics. Similarly, let \mathbf{V}_{test} and \mathbf{y}_{test} be the plantar pressure video and demographics for the patient to be evaluated. We will

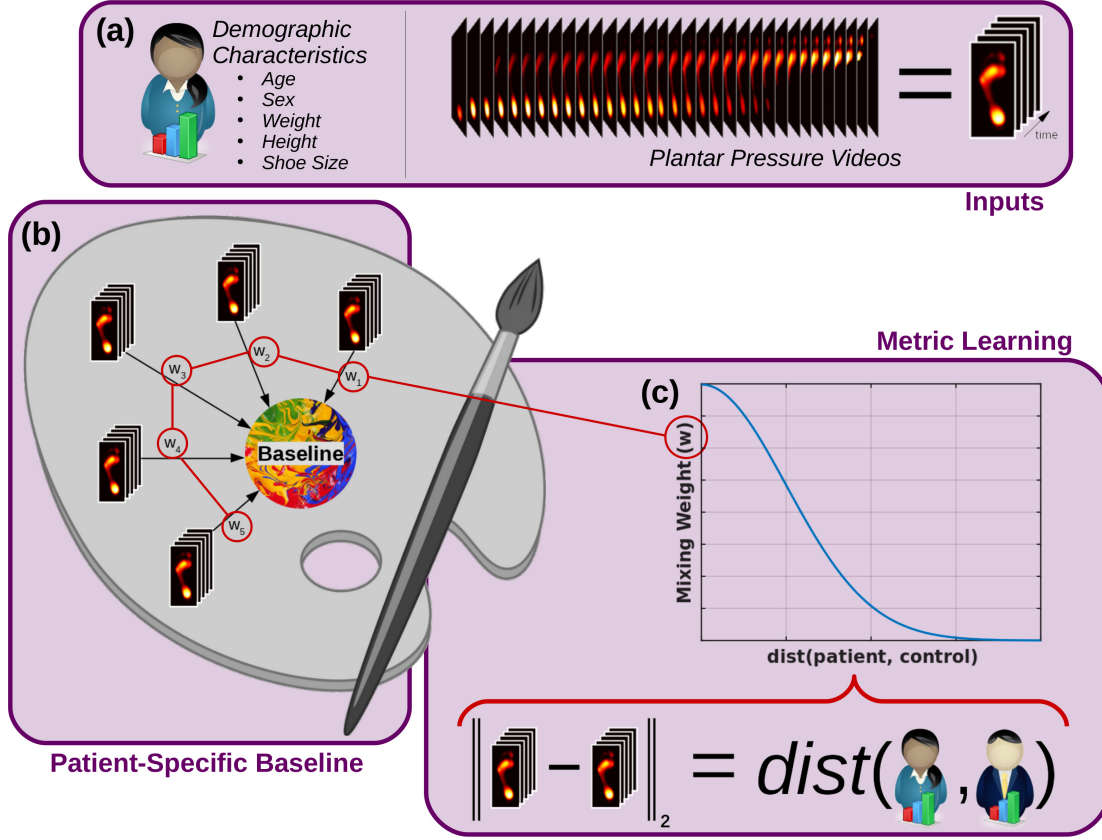


Figure 1: The creation of a statistical baseline for a patient’s plantar pressure video. Much like mixing different paints to produce a new color, previously collected plantar pressure videos are also mixed to produce a statistical baseline. The inclusion of a video in the baseline is controlled by the demographic characteristics that are contextual factors for plantar pressure measurements. Metric learning is used to estimate how much each demographic factor should impact the creation of the statistical baseline.

assume that these plantar pressure videos have already been spatially and temporally aligned using STAPP [3].

Patient-Specific Baselines

In previous works [4], healthy controls are assumed to be sampled IID from the same population of the patient. As a result, the baseline mean, \mathbf{M} , and standard deviation, \mathbf{S} , videos are constant for all patients and given by:

$$\mathbf{M}(\mathbf{x}, t) = \frac{1}{N} \sum_{i=1}^N \mathbf{V}_i(\mathbf{x}, t), \quad (1)$$

$$\mathbf{S}(\mathbf{x}, t) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N [\mathbf{V}_i(\mathbf{x}, t) - \mathbf{M}(\mathbf{x}, t)]^2}, \quad (2)$$

where \mathbf{x} is a pixel location, and t a time frame, in the video. In practice, however, each healthy control is not an IID sample. Instead, we know that a person’s age, weight, height, sex, and shoe size all influence their plantar pressures. Therefore, we propose to model this problem as a contextual outlier detection problem where these demographic factors provide the necessary context.

To address these concerns, we propose to weight each healthy control differently in the creation of the statistical baseline. The demographic factors will be used to define these weights in order to address the non-IID sampled healthy controls. This approach takes inspiration from how an artist would mix different amounts of different paints in order to produce a new color (Figure 1). It should be noted that Serag et al. pro-

posed a similar idea where a single demographic measurement (age) was used [12], but their use of a single demographic factor results in a trivial way of defining similarity between individuals.

In contrast to the IID outlier detection approach in Equations (1) and (2), we construct the baseline pixel-by-pixel normal distributions using a global weighted kernel regression, resulting in mean and standard deviation videos defined as,

$$\mathbf{M}_{test}(\mathbf{x}, t) = \frac{\sum_{i=1}^N w(\mathbf{y}_i, \mathbf{y}_{test}) \mathbf{V}_i(\mathbf{x}, t)}{\sum_{i=1}^N w(\mathbf{y}_i, \mathbf{y}_{test})}, \quad (3)$$

$$\mathbf{S}_{test}(\mathbf{x}, t) = \sqrt{\frac{\sum_{i=1}^N w(\mathbf{y}_i, \mathbf{y}_{test}) [\mathbf{V}_i(\mathbf{x}, t) - \mathbf{M}(\mathbf{x}, t)]^2}{\sum_{i=1}^N w(\mathbf{y}_i, \mathbf{y}_{test}) - \frac{\sum_{i=1}^N w(\mathbf{y}_i, \mathbf{y}_{test})^2}{\sum_{i=1}^N w(\mathbf{y}_i, \mathbf{y}_{test})}}}, \quad (4)$$

where the weights are obtained from a Gaussian kernel over the distances between the contextual demographic factors:

$$w(\mathbf{y}_i, \mathbf{y}_j) = \exp(-dist(\mathbf{y}_i, \mathbf{y}_j)). \quad (5)$$

This kernel regression ensures that healthy controls with similar demographics to the patient receive higher weights for their plantar pressure videos than other less similar healthy controls. Unfortunately, this regression alone does not explain how distances should be measured between demographics factors in order to define an accurate statistical baseline for an individual. For example, a person's weight has a greater impact on plantar pressures than a person's age [7] and this should be captured in the distance metric, but to what degree? How do we place a number on this?

Metric Learning

To address this question, we developed a high-dimensional extension of the metric learning for kernel regression (MLKR) algorithm initially proposed by Weinberger and Tesauro [14]. MLKR has been used for contextual outlier detection before [15], but here we combine it with our statistical baseline above in order to integrate it within the SPM framework. Conceptually, MLKR aims to define a distance metric that optimally

reconstructs the plantar pressure videos. This objective is captured mathematically through a leave-one-out regression loss function

$$\mathcal{L} = \sum_{i=1}^N \sum_{\mathbf{x}} \sum_t \left\| \mathbf{V}_i(\mathbf{x}, t) - \frac{\sum_{j \neq i} w(\mathbf{y}_i, \mathbf{y}_j) \mathbf{V}_j(\mathbf{x}, t)}{\sum_{j \neq i} w(\mathbf{y}_i, \mathbf{y}_j)} \right\|, \quad (6)$$

incorporating $w(\cdot)$, the kernel over demographic distances defined in Equation (5). Effectively, this approach aims to learn a distance metric that emphasizes each demographic factor based on the degree to which each factor influences the plantar pressure videos. A Euclidean norm is used here as each plantar pressure dataset is assumed to be an equally reliable estimate of the gait of a healthy individual.

While MLKR allows for a variety of distance metrics, we follow the approach in [14] and model the distances using a Mahalanobis-style metric,

$$dist(\mathbf{y}_i, \mathbf{y}_j) = (\mathbf{y}_i - \mathbf{y}_j)^T \mathbf{P} (\mathbf{y}_i - \mathbf{y}_j) \quad (7)$$

where the positive semi-definite matrix \mathbf{P} captures the relative influences of each demographic factor to both the plantar pressure videos and to each other. To estimate \mathbf{P} , we first replace it by its Cholesky decomposition $\mathbf{P} = \mathbf{L}\mathbf{L}^T$ in order to preserve its positive semi-definite structure. Subsequently, the decomposed matrix \mathbf{L} is then solved for by inserting it into Equation (7), inserting Equation (7) into Equation (5), inserting Equation (5) into Equation (6), and minimizing the resulting regression loss function \mathcal{L} with respect to \mathbf{L} . We note that this loss function is convex and can be minimized using, among other techniques, gradient descent. The gradient of \mathcal{L} with respect to \mathbf{L} is,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}} = 4\mathbf{L} \sum_{\mathbf{x}} \sum_t \sum_{i=1}^N \left([\mathbf{M}_i(\mathbf{x}, t) - \mathbf{V}_i(\mathbf{x}, t)] \sum_{j \neq i} [(\mathbf{M}_j(\mathbf{x}, t) - \mathbf{V}_j(\mathbf{x}, t)) w(\mathbf{y}_i, \mathbf{y}_j)(\mathbf{y}_i - \mathbf{y}_j)^T(\mathbf{y}_i - \mathbf{y}_j)] \right) \quad (8)$$

Using the plantar pressures and demographics from our healthy controls, we employ gradient descent to solve for the decomposition matrix \mathbf{L} . This matrix can then be used to generate patient-specific baseline plantar pressures using (3) and (4). Note that this optimization has no hyperparameters that need to be fine tuned [14].

Outlier Detection

Once a distance metric is learned over the demographics, patient-specific statistical baselines are computed at each pixel of the plantar pressure video and the SPM framework can then be used to identify outliers. To do so, the plantar pressure video of the patient under evaluation is checked for anomalies using z-scores [4]:

$$\mathbf{z}_{test}(\mathbf{x}, t) = \frac{\mathbf{V}_{test}(\mathbf{x}, t) - \mathbf{M}_{test}(\mathbf{x}, t)}{\mathbf{S}_{test}(\mathbf{x}, t)}. \quad (9)$$

Pixel-by-pixel anomalies are then defined as those whose z-scores are statistically significant, at $\alpha = 0.05$, following multiple comparison correction using Random Field Theory [10].

EXPERIMENTAL SETUP

To evaluate the proposed outlier detection technique, we employ three datasets of plantar pressure videos. First, an internal dataset of 430 healthy controls is used as the training set to learn the distance metrics over the demographic factors. This dataset is also used to generate the baseline mean and standard distribution videos. Second, the CAD WALK healthy controls database (<http://doi.org/10.5281/zenodo.1265420>) contains plantar pressure videos from 55 healthy individuals. These measurements are used to validate how well the estimated baselines match real plantar pressure videos from healthy individuals. Finally, the CAD WALK hallux valgus database (<http://doi.org/10.5281/zenodo.2598496>) is used to evaluate how capable the proposed technique is

at identifying the plantar pressure anomalies that are known to exist for this patient population [2]. This dataset consists of 69 hallux valgus cases measured from the feet of 50 patients (19 of these 50 patients have both feet affected). All plantar pressure videos were collected using calibrated rs scan footscan[®] pressure-sensing plates (rs scan, Paal, Belgium). The CAD WALK healthy controls were measured at 500 Hz while the other two groups were measured at 200 Hz. Also, participants in both CAD WALK datasets were measured using a 3-step protocol (i.e. the third step of the walk is measured), while the internal dataset was collected using an 8-step protocol [5]. The study was approved by the internal review committee of the Sint Maartenskliniek and met the requirements for exemption from the Medical Ethics Committee review under the Dutch Medical Research Involving Human Subjects. The study was performed in accordance to the declaration of Helsinki.

To test our proposed outlier detection technique (subsequently labeled FULL for full metric learning), we compared our technique to five similar competing approaches. The first competing technique was a one-dimensional version of the metric (subsequently labeled 1D) where $\mathbf{P} = \sigma \mathbf{I}$ (σ is the unknown scaling parameter and \mathbf{I} is the identity matrix). The one-dimensional version is conceptually equivalent to the approach of Serag et al. [12]. The second technique was the IID approach described in (1) and (2) and used in [4] (subsequently labeled IID). The third technique (subsequently labeled DIAG) uses a diagonal \mathbf{L} matrix for the metric. This approach serves as a middle ground between our proposed FULL metric approach and the 1D approach of Serag et al. [12]. The fourth competing approach is PAPPI [2], which uses linear regression to model the impact the demographics have on the plantar pressures. Finally, we also compare to Robust Contextual Outlier Detection (ROCOD) [8], which combines the linear regression baseline of PAPPI with an IID baseline calculated over a local neighborhood defined by the demographics. For ROCOD, the local demographic neighborhoods were empirically defined to contain samples whose max-min normalized demographics are within a Euclidean distance of $\tau = 0.25$ of

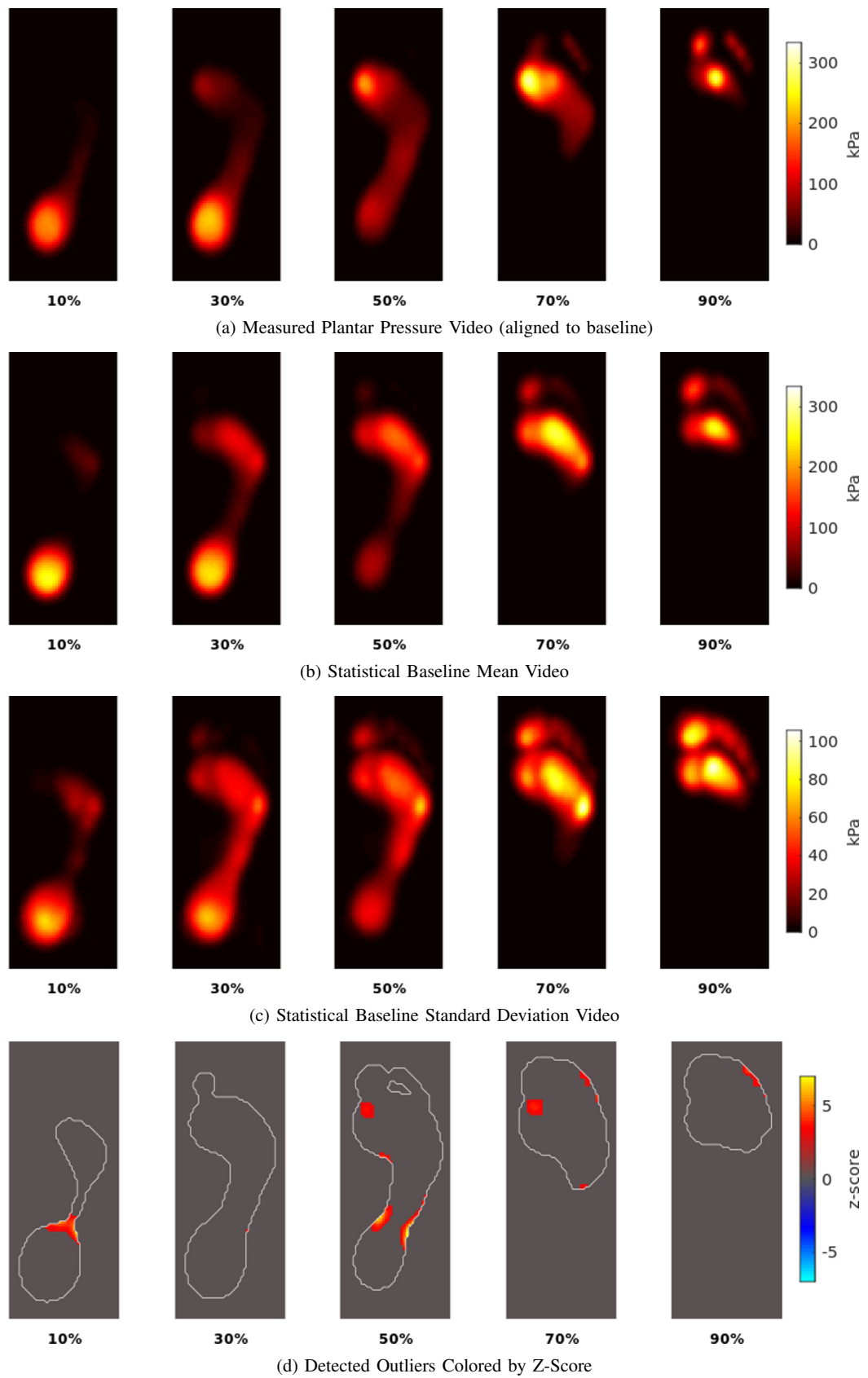


Figure 2: Example of proposed outlier detection algorithm at different percentages into the stance phase.

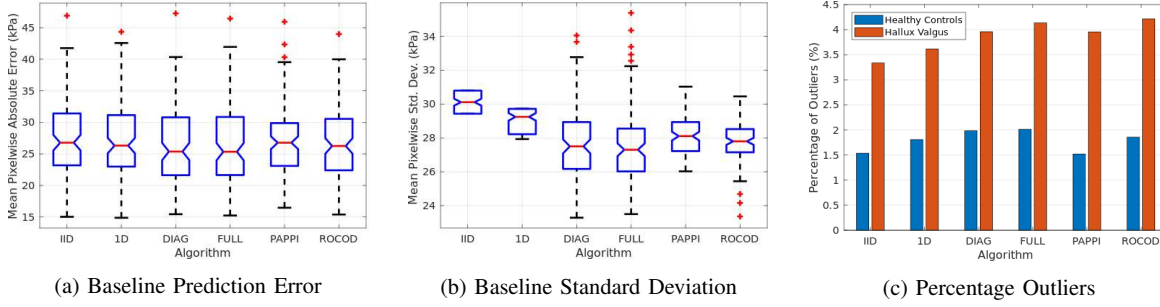


Figure 3: Quantitative results on statistical baseline estimations. The proposed FULL metric learning approach produced the most accurate baseline prediction with the lowest variance, resulting in among the largest number of outliers detected. See text for further details.

the test patient.

To quantify the performance of each outlier detection technique, we perform 3 experiments. First, we compare the estimated baseline mean plantar pressure videos to real measured plantar pressures from healthy controls. Differences between the two videos are quantified using the average per pixel absolute error. We hypothesize that our FULL metric learning approach will achieve the lowest absolute errors. Second, we measure baseline sensitivity using the average per pixel magnitude of the baseline standard deviation videos. We hypothesize that our FULL metric learning approach will show the lowest variability compared to the competing algorithms. Third, we perform outlier detection on both the hallux valgus patients and healthy controls. We hypothesize that our FULL metric learning approach will be more sensitive to outliers than competing algorithms. Additionally, we hypothesize that the Hallux Valgus patients examined with our FULL metric learning approach will have plantar pressures outliers that show better agreement with previous Hallux Valgus studies than competing approaches. Specifically, we expect to see outliers around the hallux (i.e. big toe) where the foot condition is present [2].

RESULTS

Figure 2 shows an example of our ML outlier detection technique on a plantar pressure video from one of our 69 Hallux Valgus cases. Note that the estimated baseline plantar pressure video shows a realistic pressure pattern and that the detected outliers under the midfoot, toes, and metatarsal 1 agree with what is commonly seen

in Hallux Valgus patients [2]. Similar figures for all plantar pressure datasets and all algorithms are provided as supplementary material.

Figure 3a shows the average per pixel absolute error between the measured plantar pressures from healthy controls and the baseline mean images from each statistical baseline. Both the DIAG and FULL metric learning approaches produced baseline mean images that were closest to the real measured plantar pressures, each having an average error of 26.56 kPa. These errors were 1.53% - 3.15% lower than competing algorithms. Paired t-test showed that the FULL metric learning approach produced errors that were statistically lower than the IID and 1D approaches ($p = 0.004$ vs. IID, $p = 0.009$ vs. 1D).

Figure 3b shows the average per pixel standard deviations for each statistical baseline estimation technique. The inclusion of the demographic factors significantly reduced the variability in the baseline, with our FULL metric learning approach showing the lowest variability on average. These decreases in variability were between 0.67% - 8.66% in magnitude. Paired t-tests show that FULL metric learning produced significantly lower baseline variability than IID ($p < 1e^{-10}$), 1D ($p < 1e^{-10}$), DIAG ($p < 4e^{-4}$), and PAPPI ($p = 0.030$).

For each dataset and outlier detection technique, the average percentage of pixels in the plantar pressure videos that were identified as outliers are shown in Figure 3c. As expected, the Hallux Valgus patients showed more pressure outliers than the healthy controls. Also, the number of outliers increased as the statistical baselines became more accurate and contained less vari-

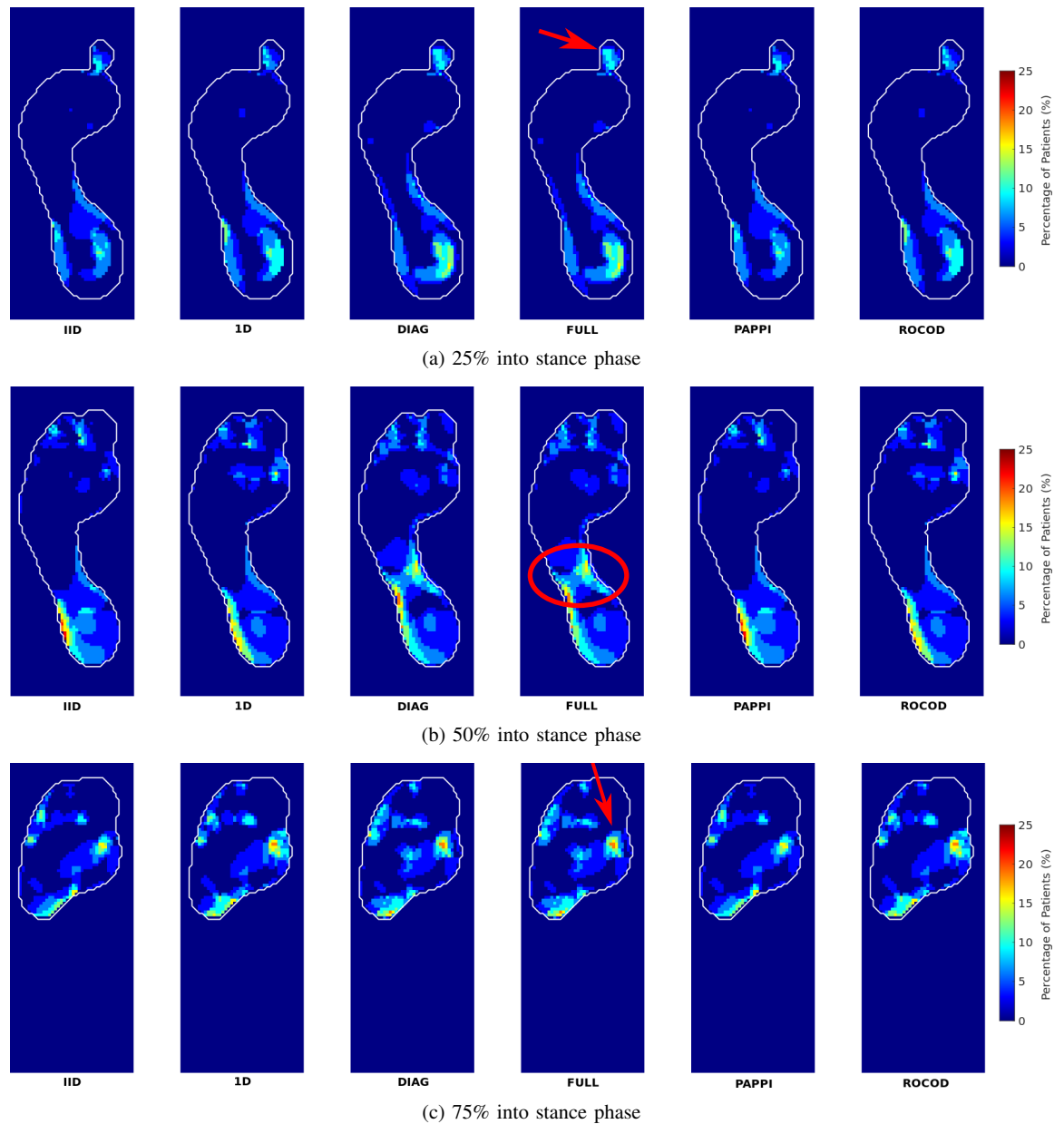


Figure 4: Histograms of detected outliers for the Hallux Valgus patient dataset (left foot). Note that the FULL and DIAG metric learning approaches show more outliers in the midfoot (at 50% into stance phase), hallux (at 25% into stance phase), and metatarsal 1 (at 75% into stance phase), results which agree best with previous Hallux Valgus studies.

ability, suggesting improved outlier sensitivity. In this respect, both FULL and ROCOD show the greatest number of outliers.

Figure 4 shows spatiotemporal histograms (i.e. pixel-by-pixel counts) of the detected outliers for each of the three outlier detection approaches on the left feet of hallux valgus patients. Note that for Hallux Valgus patients, abnormal plantar

pressures are expected at the location of the foot condition: under the hallux (i.e. big toe) and metatarsal 1 (i.e. at the base of the big toe). The FULL and DIAG metric learning approaches equally show the highest number of outliers in these areas as shown by the red arrows in Figures 4a and 4c. Additionally, there is evidence that Hallux Valgus patients are more likely to

have flat feet [2]. The FULL and DIAG metric learning approaches also equally identify outliers in the midfoot in Figure 4b, showing that it is identifying these flat feet. It is also worth noting that the non-IID outlier detection algorithms were better able to identify outliers under the expected foot regions than IID.

Finally, Figure 5 shows the precision matrices, \mathbf{P} , learned for the 1D, DIAG, and FULL metric learning approaches. We noticed that a person's weight has the largest impact on their plantar pressures, and that notable interactions are present between all demographic factors. These results further emphasize the need for non-IID outlier detection.

DISCUSSION

Overall, the estimations of statistical baselines improved as the impact of the demographic factors were increasingly and more flexibly modeled. Our FULL metric learning approach also integrates well into the SPM framework in order to perform outlier detection across whole plantar pressure videos. As a result of these improved statistical baselines, our approach was able to identify more outlier plantar pressures and, more importantly, have those outliers match better with previously-reported results on Hallux Valgus patients. These results suggest that our proposed outlier detection technique has improved the sensitivity and reliability of outlier detection in plantar pressure videos.

Nevertheless, there are results in our study that suggest caution regarding the interpretation of the detected outliers. First, the number of outliers detected for the healthy controls was higher than we would expect. When we qualitatively evaluated the outliers for this dataset (see supplementary materials), we observed that these outliers are generally caused by two effects: errors in spatiotemporal alignment between the baseline and measured plantar pressure videos, and harder heel strikes in the CAD WALK healthy control dataset than in our internal healthy controls dataset. The former of these effects is a limitation of all SPM-style outlier detection algorithms and is something that should be manually checked each time an algorithm like this is used [2], [3]. The latter of these effects may be related to a difference in data collection between the two datasets [5].

The CAD WALK dataset was collected using a 3-step protocol while our internal dataset used an 8-step protocol. As a result, the CAD WALK participants may have been in a less natural walking rhythm at the time of measurement than the participants in our model. Whether this is indeed the case is something that we intend to investigate.

Additionally, Figure 3b shows a number of large outliers in our proposed estimation of baseline standard deviations. We hypothesize that the increased baseline variability for these individuals may be related to their demographic similarity to the training database. Note that the proposed algorithm creates statistical baselines through the interpolation of plantar pressure videos in the database, and that this interpolation is based on demographic factors. If an patient with notably different demographic measurements presents themselves, we do not have similar individuals in the training database from whom we can interpolate a good statistical baseline. In such cases, extrapolation is required. Our future work will look into whether recent work in machine learning can provide this extrapolation effect.

Finally, it is worth noting that the proposed technique – incorporating contextual outlier detection with SPM – has applications beyond the analysis of plantar pressure videos. SPM has seen extensive use in the field of biomechanics [10] and the proposed technique could be used to identify form breaks in athletes. Medical imaging also makes extensive use of SPM [4], and the proposed technique for computer-aided diagnosis. The geosciences also employ SPM for the study of hyperspectral images [9]. The proposed technique can also be used there to detect, for example, groundwater contamination. These areas remain as future work to be explored.

CONCLUSION

We have proposed herein an algorithm for the statistical outlier detection, incorporated it into the SPM framework, and applied the result to the outlier detection in plantar pressure videos. Due to the contextual effects of multiple demographic factors, we employed metric learning to generate patient-specific plantar pressures benchmarks for each pixel of each frame of the plantar pressure

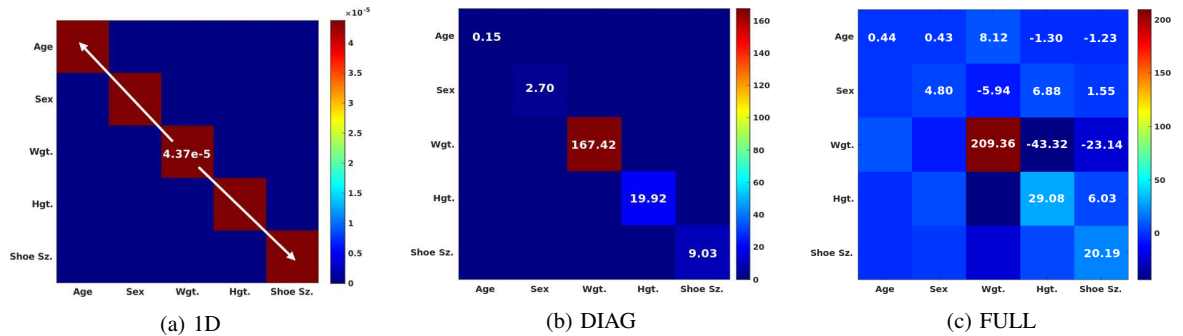


Figure 5: Precision matrices, \mathbf{P} , for the metrics learned in this study. Note that one's weight has the largest impact on plantar pressures and that interactions do exist between the demographic factors. Our FULL metric learning approach appears the best-suited of the three to model these interactions.

videos. We observed that, for healthy individuals, our proposed patient-specific benchmarks were more comparable to measured plantar pressures than an IID approach or linear models can provide. This leads to an outlier detection technique that is more sensitive to pressure outliers and appears to detect pressure outliers that show better agreement with clinical literature. While more study is needed, the introduction of this non-IID outlier detection approach may ultimately improve the ability to diagnose foot complaints from plantar pressure videos.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 746614.

REFERENCES

1. M. A. R. Ahad, T. T. Ngo, A. D. Antar, M. Ahmed, T. Hossain, D. Muramatsu, Y. Makihara, S. Inoue, and Y. Yagi. Wearable sensor-based gait analysis for age and gender estimation. *Sensors*, 20:2424, 2020.
2. B. G. Booth, E. Hoefnagels, T. Huysmans, J. Sijbers, and N. L. W. Keijsers. PAPPI: Personalized analysis of plantar pressure images using statistical modelling and parametric mapping. *PLoS ONE*, 15(2):e0229685, 2020.
3. B. G. Booth, N. L. W. Keijsers, J. Sijbers, and T. Huysmans. STAPP: spatiotemporal analysis of plantar pressure measurements using statistical parametric mapping. *Gait & Posture*, 63:268–275, 2018.
4. B. G. Booth, S. P. Miller, C. J. Brown, K. J. Poskitt, V. Chau, R. E. Grunau, A. R. Synnes, and G. Hamarneh. STEAM - statistical template estimation for abnormality mapping: A personalized DTI analysis technique with applications to the screening of preterm infants. *NeuroImage*, 125:705–723, 2016.
5. S. A. Bus and A. de Lange. A comparison of the 1-step, 2-step, and 3-step protocols for obtaining barefoot plantar pressure data in the diabetic neuropathic foot. *Clinical Biomechanics*, 20(9):892–899, 2005.
6. M. Hauskrecht, I. Batal, M. Valko, S. Visweswaren, G. F. Cooper, and G. Clermont. Outlier detection for patient monitoring and alerting. *Journal of Biomedical Informatics*, 46(1):47–55, 2013.
7. N. L. W. Keijsers, N. M. Stolwijk, and J.-W. K. Louwerens. The effect of various subject characteristics on plantar pressure pattern. *Journal of Foot and Ankle Research*, 7(1):A40, 2014.
8. J. Liang and S. Parthasarathy. Robust contextual outlier detection: Where context meets sparsity. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 2167–2172, 2016.
9. S. A. McKenna. *Handbook of Mathematical Geosciences*, chapter Statistical Parametric Mapping for Geoscience Applications, pages 277–297. Springer, 2018.
10. T. C. Pataky. Assessing the significance of pedobarographic signals using random field theory. *Journal of Biomechanics*, 41:2465–2473, 2008.
11. J. A. Ramirez-Bautista, A. Hernández-Zavala, S. L. Chaparro-Cárdenas, and J. A. Huerta-Ruelas. Review on plantar data analysis for disease diagnosis. *Biocybernetics and Biomedical Engineering*, 38(2):342–361, 2018.
12. A. Serag, P. Aljabar, G. Ball, S. J. Counsell, J. P. Boardman, M. A. Rutherford, A. D. Edwards, J. V.

- Hajnal, and D. Rueckert. Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. *NeuroImage*, 59(3):2255–2265, 2012.
13. J. P. Singh, S. Jain, S. Arora, and U. P. Singh. Vision-based gait recognition: A survey. *IEEE Access*, 6:70497–70527, 2018.
 14. K. Q. Weinberger and G. Tesauro. Metric learning for kernel regression. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, pages 612–619, 2007.
 15. G. Zheng, S. Brantley, T. Lauvaux, and Z. Li. Contextual spatial outlier detection with metric learning. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2161–2170, 2017.

Brian G. Booth is currently a postdoctoral researcher and Marie Curie Fellow at the University of Antwerp’s Vision Lab, Antwerp, Belgium. His research interests are focused towards medical image analysis, specifically on the use of machine learning techniques to identify patterns of abnormality in medical images. Booth received his Ph.D. in computing science in 2016 from the Medical Image Analysis Lab at Simon Fraser University. Contact him at brian.booth@uantwerpen.be.

Noël L. W. Keijsers is a senior researcher at Sint Maartenskliniek, Nijmegen, The Netherlands and at the department of Rehabilitation, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands. His field of expertise can be summarized as experimental research on biomechanics and motor control in patients with motor disabilities. Keijsers received his Ph.D. at the department of biomedical physics of the Radboud University medical Center. Contact him at n.keijsers@maartenskliniek.nl.

Jan Sijbers graduated in Physics in 2003. He received a PhD in Physics from the University of Antwerp in 2008. In 2010, he was appointed as a senior lecturer and became a full professor at the University of Antwerp in 2014. Jan Sijbers is the head of imec-Vision Lab and co-founder of IcoMetrix and Deltaray. His main interests are in image reconstruction, processing, and analysis with focus on Magnetic Resonance Imaging and X-ray Computed Tomography. Contact him at jan.sijbers@uantwerpen.be.