Multi-modality fusion of floor and ambulatory sensors for gait classification

DOI:
10.1109/ISIE.2019.8781127

Citation for published version (APA):

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Multi-modality fusion of floor and ambulatory sensors for gait classification

Yunas, Syed Usama
Department of Electrical and Electronics Engineering,
The University of Manchester, Manchester, United Kingdom.
syd.yunas@manchester.ac.uk

Alharthi, Abdullah
Department of Electrical and Electronics Engineering,
The University of Manchester, Manchester, United Kingdom.
abdullah.alharthi@manchester.ac.uk

Ozanyan, Krikor B
Department of Electrical and Electronics Engineering,
The University of Manchester, Manchester, United Kingdom.
k.ozanyan@manchester.ac.uk

Abstract—In a case study of gait classification from floor and ambulatory sensors, we compare results with data from each modality. The automatic extraction of features is achieved by Principle Component Analysis or Canonical Correlation Analysis, the latter performing better even with a reduced number of components used. Non-linear classifiers are most efficient for fused features. With a Kernel Support Vector Machine around 94% accuracy is demonstrated, improving over the 87% and 79% accuracies obtained with separate floor and ambulatory sensor data, respectively.

Keywords—floor sensors, wearable sensors, sensor fusion, machine learning (ML), inertial measurement unit (IMU), principal component analysis (PCA), canonical correlation analysis (CCA).

I. INTRODUCTION

In this work, the case study of gait classification from two separate gait monitoring modalities is used, as an example where, due to the complexity of the object under investigation it is difficult to define and target experimentally reliable classification features. A further motivation is to assess the improvement which is possible to achieve by the fusion of data from more than one modality.

Gait in humans defines their walking behaviour. Humans can be differentiated on the basis of their walking pattern, as any other biometric entity. Walking behaviour changes on the basis of age, weight, height and gender in humans. Human gait can be classified as either natural (used by humans instinctively) or trained (used by humans not instinctively or learned through training) [1]. Abnormal gait is a specific type of gait in which humans walk in a way different than natural. Abnormal gait could be caused by ageing, physical disability or event, such as a stroke. However it could be improved through medical treatments and exercises.

Over the past few decades, the increase in public awareness about healthcare, physical activities, health and safety sensing has initiated an emerging need for smarter sensor technologies and monitoring applications which can sense, monitor and provide feedback about the human physical activities and their health status [2].

Recently, ambulatory sensors and some data mining intelligent approaches have been used to analyse human activities and gait parameters as convenient and cheaper methods. Despite the fact that these sensors are quite usable and comfortable as compared to other methods, there are many problems to use ambulatory sensors to accurately assess human gait patterns. The uses of multi-sensor approach also requires proper calibration. High accuracy is also needed while detecting and classifying different activities and gait to evaluate the quality and performance of systems.

Machine Learning (ML) as a key tool for sensor data analysis, is becoming a centric part of novel sensor design. ML is a domain of artificial intelligence which was first introduced by Arthur Samuel in 1959. ML algorithms are used to implement complex models and methods through predictions [3] which are able to learn and make decisions on data obtained from multiple sensors; thus ML has a major role to play in the field of data processing and sensor fusion in particular. These algorithms are also very useful in exploring the hidden aspects through learning background relationships and trends in data. ML has now entered everyday lives due to the reliable and repeatable results delivered, namely with Facebook facial recognition, Kinect devices, virtual reality headsets, speech or voice recognition over phones, Robot dogs, online retail such as Amazon and Netflix etc.

In this research paper, we have proposed an efficient and effective way of sensor fusion of two different modalities i.e., Plastic Optical Fiber (POF) based floor sensors and Inertial Measurement Unit (IMU) based ambulatory sensors at feature level. Different ML algorithms such as Logistic Regression, KNN, SVM(Linear), Kernel SVM, Decision Trees and Random Forest Algorithm are used to evaluate the performance of ML algorithms for gait classification. It is evident that multi-modality fusion performs more accurate analysis of effected walking behaviours of subjects whilst performing different tasks. This paper is organized as follows. Section II briefly describes modalities used to capture human gait. Section III explains our methodology of multi-modal fusion using data processing, dimension reduction, and classification. Result and findings have been further concluded and discussed in section IV.

II. GAIT MODALITIES

There are a number of modalities used to capture and analyse human activities and gait. Some of their peculiarities and characteristics are discussed as follows:
A. Ambulatory Sensors

Use of ambulatory sensors to monitor and classify human activities and gait has proven to be important [4]. A specific type of sensor, the inertial measurement unit (IMU) has been widely used due to its small size, cost, light weight and high precision characteristics. IMU sensors consist of accelerometer, gyroscope and magnetometer which give information about the acceleration, angular velocity and the heading direction respectively. These sensors could be worn on different parts of human body such as head, chest, waist, thigh, shank and foot [5] etc.

For gait analysis, mainly experiments are conducted on Heel strike and Toe-off events which determine the stance time and swing time of subject in gait cycle. Performance is checked on a dataset of information using different classification algorithms and techniques [6]. Panebianco et al. [7] reported a systematic review to assess the human gait and its temporal parameters in terms of accuracy and repeatability using 17 algorithms. 5 IMUs were used, one on the back, two on the shanks and two on the feet. It was determined that for human gait detection and estimation of stance time, algorithms based on the acceleration measurements on the shank and foot perform better than those based on lower trunk. However, the sensor position did not affect the step estimation.

IMU sensors are widely used in recognition of human motion disorders such as Parkinson’s disease (PD) and early detection of Freeze Of Gait (FOG) [8].

B. Video Cameras

Human gait is also acquired through video cameras. High quality video cameras are required for use under moderate lighting conditions. Basic video is captured using two or more cameras with a known focal length at a fixed distance from the subject. Results obtained from all cameras need calibration to give corrected results [9].

Mostly human silhouettes are used in gait recognition to increase the performance and reduce computational load. Human silhouettes are further segmented to divide the human body in sections such as head, pelvis and foot regions [10]. Image and video processing techniques are used to extract gait features such as step length and gait cycle from the data. Bei et al. discussed the kinematic aspect of gait such as leg swing motion to explain the correlation between the angles formed at the hip, knee and ankle joints with the help of a single kinetic sensor [11].

C. Floor Sensors

Floor sensors provide an unobtrusive way of acquiring gait information and are mainly installed at the front entrance of buildings or in access control areas. These systems can also be used to identify the location of subject with in a certain area. Floor sensors can be used on the factory floor to provide data needed for monitoring of the position and activity of ambulatory industrial robots and, in cases of co-occupancy with humans, provide additional information needed for health and safety. In this approach, typically a set of sensors or force plates are used on the floor. Mostly resistive, capacitive or fiber optic based floor sensors are used. Data obtained from sensors is used to calculate gait features. Middleton et al. [12] used different gait features such as stride length, stride cadence and ratio between time on toe to time on heel and achieved 80% recognition accuracy with a resistive sensor mat.

Without involving pre-determined gait features or image reconstruction for display purposes, raw values from sensors could also be used with machine learning methods and techniques. For classification purposes, Costilla-Reyes et al. [13] used raw readings from plastic optical fiber (POF) based floor sensors and achieved 93% of accuracy on large datasets.

D. Multi-Modal Approach

Multi-modal approach is used to employ different modalities such as ambulatory sensors, video cameras, floor sensors and pressure sensors etc. to achieve higher accuracy and reliability. A survey of multi-modal approaches is summarised in table I.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Modalities</th>
<th>Accuracy Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shakharovitch et al. [14], 2001</td>
<td>Video face &amp; Video Gait</td>
<td>80% &amp; 87%</td>
</tr>
<tr>
<td>Vidjounaite et al. [15], 2006</td>
<td>Voice &amp; IMU Sensors</td>
<td>2%-12% equal</td>
</tr>
<tr>
<td>Zhou et al. [16], 2007</td>
<td>Video face &amp; Video Gait</td>
<td>64.3% &amp; 85.7%</td>
</tr>
<tr>
<td>Chen et al. [17], 2015</td>
<td>Video &amp; IMU Sensors</td>
<td>2%-23% improvements</td>
</tr>
<tr>
<td>Bai et al. [18], 2017</td>
<td>EMG &amp; IMU Sensors</td>
<td>Multiple improvements</td>
</tr>
<tr>
<td>Leal-Junior et al. [19], 2018</td>
<td>POF &amp; IMU Sensors</td>
<td>1% best case - 4% worst case</td>
</tr>
</tbody>
</table>

III. METHODOLOGY OF MULTI-MODAL SYSTEM

Our multi modal system deals with feature level fusion of sensors comprising two modalities:

A. Ambulatory Sensor System

User wears the ambulatory sensor system on their body based on a portable battery operated Raspberry Pi 3 (RPi model B+) connected to the waist and two IMU sensors attached to the ankle of each leg of the subject. RPi is connected with IMU sensors through usb cables and controlled through a computer on a wireless network. The ambulatory sensors system can be seen as highlighted yellow in figure 3. Our selection and arrangement of sensors together with the wifi based RPi makes it a comfortable and portable ambulatory sensor system.

a) Data Acquisition and Preprocessing: When turned on, the RPi connects to the pointed wireless network and waits for a connection request from a designated client. IMU sensors are configured at sampling frequency of 200Hz. Only acceleration and angular velocity values obtained from each IMU sensor are used in this research. To isolate the effect of gravity on acceleration values, a low pass filter with cut-off frequency of 5Hz is applied. To remove drift and calculate the angle ($\theta$) from angular velocity ($\omega$), the following formula is used:

$$\theta = \theta + \omega \cdot \Delta t$$  \hspace{1cm} (1)

where $\Delta t$ represents the change in time. Data is recorded from two IMU sensors in a CSV format file. 20 frames of data are received in one second with the time stamp information. IMU sensors can be aligned to one of the body panels as shown in figure 1(a). We have aligned the axis of IMU sensors
according to ‘transversal plane’ as the subject will be walking across the floor sensor in a horizontal plane. Therefore whilst walking in a straight line from one end of carpet to the other end, wearing sensors; the maximum acceleration will be in x direction, moderate in z direction (between heel strike and toe-off phases of foot) and minimum along y direction, as shown in figure 1(b).

b) Data verification and testing: A window of 10 signals has been used on the 200Hz signal to calculate the mean value which is stored at 20 frames per second. Basic checks are put in place to alert if any of the ambulatory sensors are not synchronized or delivering an invalid output. The data obtained from ambulatory sensors are tested and verified before processing. Acceleration and angular velocity signals captured from one heel strike to another heel strike for both feet over time can be seen in figure 2.

B. POF based Floor Sensors (Smart Carpet)

The user walks on the floor sensors surface (carpet of size 2 x 1 m approx.) comprising a mesh of 116 fiber optic sensor elements, each terminated with an LED as a light source and a photodiode as detector. The signal acquisition from the floor sensor is managed by another RPi controlled through a computer on wireless network. The floor sensors (“Smart Carpet”) can be seen highlighted green in figure 3.

a) Data Acquisition and Preprocessing: The data acquired from floor sensor is in the form of a string of values in each frame, each digital value being output from a 12 bit ADC converter. Firstly, received data from RPi is converted into fraction of transmitted light by dividing values by 4096 and multiplying by 100, then normalised to the maximum transmittance. Secondly, the long string of values is truncated and organised into three plies as sensors are arranged on the sensor, at three projection angles as shown in figure 4.

b) Data Verification and Testing: The data obtained from the floor sensor is saved in a CSV file with time stamps at 20 frames per second. Since the set of measurements constitutes a Radon Transformation, a real-time tomography image reconstruction algorithm is used to display the movements of the subject on the carpet, as shown in figure 5.
C. Dimensionality Reduction and Sensor Fusion

It is widely accepted that data fusion can be performed at three levels: feature, score and decision [3]. Decision level and score level fusion have been researched immensely, mostly using multi classifier methods. Feature level fusion was very successful at early times, however it is still an open and challenging problem. In this paper, we have applied two approaches for dimensionality reduction and sensor fusion at feature level as follows:

a) Principal Component Analysis (PCA): Dealing with a number of independent variables creates a problem to visualize prediction regions and prediction boundaries. Therefore PCA is used as a model based approach for data analysis. PCA identifies the changing patterns in data and detects the correlation between variables. PCA was implemented by the following sequence [22]:

- Standardization of data.
- From the covariance or correlation matrix, get the eigenvectors and eigenvalues or perform the singular vector decomposition.
- Sort in descending order the eigenvalues and select the k eigenvectors that correspond to the k largest eigenvalues where k is the number of dimensions of the new feature subspace, k ≤ d.
- Create the projection matrix W from the selected k eigenvectors.
- Transform the original dataset say, X using projection matrix W to get k-dimensional feature subspace Y.

PCA has multiple applications like filtering of noise, data visualization, extraction of features, predictions in stock market, analysis of genetic data and many more.

b) Canonical Correlation Analysis (CCA): CCA is a multivariate statistical model that facilitates the study of interrelationship between the two sets of variables. As compared to multiple regression which predicts only one dependent variable, CCA predicts multiple dependent variables. The implementation of CCA can be summarized [23] as follows:

- Take two sets of variables, \( X = [X_1, X_2, \ldots, X_p] \) and \( Y = [Y_1, Y_2, \ldots, Y_q] \) such that \( p \leq q \).
- Find linear combinations \( U \) and \( V \) such that \( U \) corresponds to \( X \) and \( V \) corresponds to \( Y \). \( U \) and \( V \) can be described as:

\[
U_1 = A_{11}X_1 + A_{12}X_2 + \cdots + A_{1p}X_p \\
U_2 = A_{21}X_1 + A_{22}X_2 + \cdots + A_{2p}X_p \\
\vdots \\
U_p = A_{p1}X_1 + A_{p2}X_2 + \cdots + A_{pp}X_p
\]

And,

\[
V_1 = B_{11}Y_1 + B_{12}Y_2 + \cdots + B_{1q}Y_q \\
V_2 = B_{21}Y_1 + B_{22}Y_2 + \cdots + B_{2q}Y_q \\
\vdots \\
V_p = B_{p1}Y_1 + B_{p2}Y_2 + \cdots + B_{pq}Y_q
\]

- Find correlation coefficient vectors, \( A_1 = [A_{11}, A_{21}, \ldots, A_{p1}]^T \) and \( B_1 = [B_{11}, B_{21}, \ldots, B_{p1}]^T \) to maximize the correlation (\( \rho \)) given by the equation:

\[
\rho = \text{corr}(XA_1, YB_1)
\]

Or

\[
\rho = \text{corr}(U_1, V_1)
\]  \hspace{1cm} (2)

\((U_1, V_1)\) is the first canonical variate pair, \((U_2, V_2)\) is the second canonical variate pair and the rest can be found the same way. However, the second pair is uncorrelated with first pair and none of the canonical variate pairs are correlated to each other. In total \( p \) canonical covariates exist in number.

D. Block Diagram

Fig. 5. Image reconstructed diagram of a subject standing on carpet (Variable legend bar displaying the percentage of light stopped at POF carpet sensors)

Fig. 6. Block Diagram of Multi-Modal System
E. Experimentation and Analysis on Data

In order to test the multimodal system, an ethical approval is requested and permission has been granted to ask volunteers for experiments. For this research paper, a total 5 young and healthy volunteers, as listed in table II, were asked to perform dual task gait on the floor sensor whilst wearing the ambulatory sensors.

<table>
<thead>
<tr>
<th>User</th>
<th>Weight(kg)</th>
<th>Height(cm)</th>
<th>Gender</th>
<th>Age(yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>75</td>
<td>185</td>
<td>Male</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>79</td>
<td>176</td>
<td>Female</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>65</td>
<td>171</td>
<td>Male</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>70</td>
<td>173</td>
<td>Male</td>
<td>29</td>
</tr>
<tr>
<td>5</td>
<td>96</td>
<td>177</td>
<td>Male</td>
<td>33</td>
</tr>
</tbody>
</table>

Table II. User Profile

After calibration of the floor sensor, 106 sensors values were retained as being in the expected range. From ambulatory sensors, 3 axis’ values for acceleration and 3 axis’ values for angles were used from two sensors, yielding a total of 12 values available. Four dual task gait activities were designed to characterize the dual task gait among volunteers. All activities were recorded 10 times each. The performed dual task activities as the volunteer is walking are listed as:

- Subtraction of a prompted random number.
- Listening to a story.
- Texting on personal mobile device.
- Talking to the operator.

F. Classification and Comparison of Results

The dataset from 106 + 12 = 118 sensors acquired with two modalities, was used with linear and non-linear classifiers, as displayed in table III, in 5 cases for each classifier. The cases differ depending on the choice of extracted features used for input:

- Only floor sensors data (106 components)
- Only ambulatory sensors data (12 components)
- Combined sensors using PCA (12 components)
- Combined sensors using CCA (12 components)

Within the classification procedure, the data is split into training and testing sets. The possible outcomes from the classification procedure are: True Negatives (TN), True Positives (TP), False Negatives (FN) and False Positives (FP). These event definitions allow the calculation of the accuracy metrics:

\[
\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + FN} \tag{3}
\]

IV. Discussion and Conclusions

In table III, highlighted numerical values indicate the cases in which the accuracy achieved using sensor fusion is better than from individual single modality data, for the same classification algorithm (rows). All experiments are 10-fold cross-validated with 95% confidence interval. Below is a summary across the feature extraction cases and classifiers.

A. Feature Fusion

CCA summarizes the data correlation into fewer number of statistics while preserving the main aspects of the relationships. The motivation for CCA is very similar to PCA, however in the latter, the next new variable represents the maximum variance in the individual datasets. On the other hand in CCA, the new variable is identical for both sets of data such that the correlation between the two resulting new variables is maximized. This suggests that fewer canonical variates could be adequate to identify correlation between two datasets.

Table III highlights only a single classification accuracy measure for 12-component PCA (column C) using K-SVM higher when compared to both Floor Sensor System (column A) and Ambulatory Sensor System in (column B). Similarly, for 12-component CCA (column D) accuracy measures for K-SVM and K-NN are higher when compared to both Floor Sensor System (column A) and Ambulatory Sensor System (column B). It is evident that the overall accuracies for 12-component CCA (column D) using non-linear classifiers: K-NN and KSVM are much higher than 12-component PCA (column C) when compared with both Floor Sensor System (column A) and Ambulatory Sensor System (column B).

Indeed, results for CCA fused features reveal that in some cases it is possible to reduce the computational load and still
increase the overall classification accuracy while selecting fewer components. Highlighted is the row in table III (K-SVM), where the largest improvement in accuracy is achieved with 9-component CCA (column E) over 12-component PCA (col: C).

Further improvements in classification accuracy would be expected from data fusion at feature level by using more recently introduced Kernel CCA (KCCA) and Deep CCA.

B. Classification Algorithms

In the multi-modal cases, KNN, KSVM, Decision Trees and Random forest algorithms yield higher accuracy as compared to Logistic Regression and SVM. This is a straightforward reflection of the nature of the task and a confirmation of the non-linear character of the data; e.g., angles from IMU sensors are not linear when fed along with the linear acceleration values to the classification algorithms.

KNN outperforms the other algorithms in overall accuracy on separate and fused datasets (with the single exception of IMU data processed by Random Forest). However, from table III, there is no significant change in performance when compared to floor sensor data only, while KSVM gives advantage of PCA-fused features over single modality ambulatory sensor features and all CCA-fused cases perform better than any of the two individual modality datasets. KSVM avoids mapping of data to and from higher dimensional space by introducing a kernel; it was observed that a Gaussian kernel performed better than sigmoid or polynomial alternatives.

The Decision Trees algorithm is trained with the bagging method, which is a combination of learning models to improve the overall results. The Random Forest algorithm is an ensemble of decision trees (the default 10 trees were used), therefore yields higher accuracies as compared to the Decision Trees algorithm, as confirmed in table III. Decision Trees and Random Forest algorithms simply do not work in this method of sensor fusion. Higher accuracies have been achieved for both algorithms on Floor Sensor System (column A) as compared to the accuracy measures of sensor fusion (columns: C, D & E).

In conclusion, classifications of human gait are demonstrated by ML from data, acquired separately from floor and IMU sensors as well as from data fused at feature level by PCA and CCA. The latter performs better for most non-linear algorithms when 12 components are used. It is possible by fusion to achieve improvement over the IMU case even with 9 CCA components. Such reduction in the computational cost is essential in a system implementation with the best performing KSVM classifier, which can be demanding with larger datasets. Multi-modality fusion at feature level recommends itself for classification in a range of problems and embedding in implementations where the primitives of data variability are undefined or unknown.

ACKNOWLEDGMENT

Peter M. Green is acknowledged for his endless hardware support and Patricia Scully for her guidance on POF systems.

REFERENCES