Gait Based Authentication System

Vikas Kumar[1], *Associate Professor CSE*, Vatsal Agarwal[1], Vivek Kumar[1], Vijay Bhasin[1] and Yash Agarwal[1]

[1]Moradabad Institute of Technology, Ram Ganga Vihar, Phase - II,Moradabad - 244001 India

The purpose of this project was to develop a method for classifying people based on acceleration data. This paper presents a biometric user authentication based on a person's gait. Unlike most previous gait recognition approaches, which are based on machine vision techniques, in our approach gait patterns are extracted from a mobile device's accelerometer and evaluated against a machine learning model to identify person and compare various algorithms so as to obtain highest precision.

***Index Terms*—Accelerometer,Machine Learning, Gait, Azure.**

I. Introduction

THIS PAPER presents gait as a robust means for human identification which is clinically proved to be highly subjective. Biometrics is being used for long time for identification purposes because of its robustness and easiness.
The conventional biometric identification techniques as fingerprint matching, face recognition require the user to log entry into a system and wait for confirmation - there is delay involved in the process instead of the proposed system, which is *unobstructive* in nature, because mobile devices (mobile phones, PDAs etc.) have inbuilt accelerometer sensors in them so continuous verification for person is possible with intervention. This makes it hold an edge over other identification methods. Also the comparing the equal error rate(EER) the gait system achieves lower EERs than fingerprint recognition[1] or 2-dimensional face recognition[2].This can be user friendly as the identification goes in background processes.
Human gait has been introduced as a particular style as a particular style and manner of moving human feet [3]. In a more detailed level view, the mechanism of human gait involves synchronization between the skeletal, neurological and muscular system of human body [3]. Therefore, gait characteristics will vary from people to people. Gait recognition has been studied as a behavioral biometric for decades. Its techniques could be typically divided into 3 categories: Machine Vision Technology (WVT) [4, 5], Floor Sensor Technology (FST) [6], and Wearable Sensor Technology (WST) [7].

WST is recognized as the most approachable and newest of all. Sensors in WST are attached to human body in various positions, such as pockets, waist or shoes to record physical motions. WST takes advantage of mobile devices' sensing capabilities including GPS, accelerometer, and gyroscope sensor, etc. Thus, it will provide developers an edge over improving various techniques in identification.
In this paper, we propose identification through the onboard accelerometer present in smartphones.The data is collected with an app and feature extraction is performed over the data, later the data is classified using Decision Jungle algorithm : Decision Forest implementation of Microsoft Research[8].

II. Data Collection and Feature Extraction

In this section we will describe how we collect the accelerometer data and extract interesting features from the logs as to build predictive models for identification.The datasets are splitted into training and test datasets to obtain robust model. The remaining section describes how the raw accelerometer data is being collected and transformed for this process.

A. Data Collection

People with age ranging from 18 to 40 (students and instructors) participated in volunteer data collection program, both male and female participated and logged the data in a whole span of week at various time of the day, walking with a constant pace at a smooth surface.A upload utility was provide to aggregate the data at a central repository[9].

The data is being collected with the help of Accelerometer Log android application available on Google Play Store with due permission. This application logs data from device's accelerometer in 3 coordinates axes - x, y and z and stores them on device's external storage. The format used for logging accelerometer data is comma separated values(CSV).

The app starts logging upon launching and the device needs to be put in the following manner for effective logging of accelerometer data.

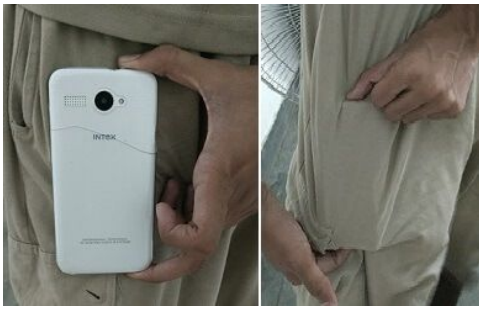

FIG. 1 Position of device

B. Feature Extraction

There are several ways to prepare the raw sensor data before using it for biometrics. Some gait-based biometric work utilizes the data within the time domain [10, 11, 12], but other successful systems map the time-series sensor data into examples using a sliding window approach, which permits the use of conventional classifier induction systems that cannot handle time-series data.

We are using sliding window approach to extract features for a dataset bounded by the time slot of the window i.e. the actual time person has logged the data through given means - android application. Several features are computed with the dataset using custom R scripts which include :

- Average : Average sensor value (each axis)
- Standard Deviation : Standard deviation (each axis)
- Average Absolute Difference : Average absolute difference between the 200 values and the mean of these values (each axis)
- Average Resultant Acceleration : For each of the sensor samples, take the square root of the sum of the squares of the x, y, and z axis values, and then average them.

Each set of data is being tagged with initials of the person logging the data ex. vtsl for Vatsal and vkku for Vivek. These extracted features are being added to the current dataset which is being passed for evaluating on the classification model.The whole dataset is being converted to a comma separated file(CSV) for evaluation. The computation model is implemented with the help of cloud based platform - Microsoft Azure Machine Learning. The tagged dataset upon evaluation will contain a field - Scored Labels which is used to decide final result.

III. Classifier Algorithm

The classifier algorithm used for this project is Microsoft Research's implementation of Decision Forest - Decision Jungles[8]. Problem with decision forest is that given enough data, the number of nodes in decision trees will grow exponentially with depth which is being addressed by Decision Jungles.This ensembles of rooted decision directed acyclic graphs (DAGs).Unlike conventional decision trees that only allow one path to every node, a DAG in a decision jungle allows multiple paths from the root to each leaf.During training, node splitting and node merging are driven by the minimization of exactly the same objective function, here the weighted sum of entropies at the leaves. Results on varied datasets show that, compared to decision forests and several other baselines, decision jungles require dramatically less memory while considerably improving generalization.

Although trees allow making predictions efficiently, learning the optimal decision tree is an NP-hard problem [13]. In his seminal work, Quinlan proposed efficient approximate methods for learning decision trees [14, 15]. Some researchers have argued that learning optimal decision trees could be harmful as it may lead to overfitting [16]. Overfitting may be reduced by controlling the model complexity, e.g. via various stopping criteria such as limiting the tree depth, and post-hoc pruning.Task of learning each DAG in a jungle is treated as an energy minimization problem. We then propose two minimization methods for learning the optimal DAG. Both methods alternate between optimizing the split functions at the nodes of the DAG and optimizing the placement of the branches emanating from the parent nodes.

Rooted binary DAGs have a different architecture compared to decision trees and were introduced by Platt et al. [17] as a way of combining binary classifier for multi-class classification tasks. More specifically a rooted binary DAG has: (i) one root node, with in-degree 0; (ii) multiple split nodes, with in-degree ≥ 1 and out-degree 2; (iii) multiple leaf nodes, with in-degree ≥ 1 and out-degree 0. Note that in contrast to [17], if we have a C-class classification problem, here we do not necessarily expect to have C DAG leaves. In fact, the leaf nodes are not necessarily pure; And each leaf remains associated with an empirical class distribution.

We train each rooted decision DAG in a jungle independently, though there is scope for merging across DAGs as future work. Our method for training DAGs works by growing the DAG one level at a time.1 At each level, the algorithm jointly learns the features and branching structure of the nodes. This is done by minimizing an objective function defined over the predictions made by the child nodes emanating from the nodes whose split features are being learned.

IV. RESULTS

The identification experiments, involve building a single predictive model to identify a specific user from a set of users. At the lowest level, our results are based on identifying each user based on sample of walking data. In order to demonstrate how this scheme works, and to provide greater insight into the results, a confusion matrix generated by evaluating dataset on classifier algorithm is presented in Table 1. The results in this table are based on an identification model generated from accelerometer data and using the Decision Jungle Algorithm.

Overall accuracy	0.873114
Average accuracy	0.936557
Micro-averaged precision	0.873114
Macro-averaged precision	0.835511
Micro-averaged recall	0.873114
Macro-averaged recall	0.812187

Table. 1 Classification Parameters

Actual Class \ Predicted Class	shwetaMam	vijay	vtsl	yash
shwetaMam	97.6%	0.1%	1.2%	1.1%
vijay	1.3%	80.2%	9.7%	8.7%
vtsl	4.2%	0.3%	94.8%	0.7%
yash	6.6%	39.9%	1.2%	52.3%

Table. 2 Prediction Results

The predicted class for shwetaMam is correct for 97.6% of the time, whereas 94.8% for vtsl. Increasing data samples for other users may increase their accuracy. The overall accuracy would be simply the total number of correct predictions divided by the total number of predictions. The overall accuracy obtained is 87.31% while 95.65% as average accuracy.

V. CONCLUSION AND FUTURE WORK

Gait as a means of identification is explored and found to be reliable in this paper. The available commercial smartphone embedded accelerometer was able to carry out identification with modest accuracy. As stated before, the advantage of this method to other biometric systems which could be implemented on mobile phones, is the unobtrusive operation which gives a high user friendliness. To make biometric gait recognition using embedded accelerometers a technology suitable for practical use, further research on feature extraction and comparison is required. However the achieved results are promising and the proposed approach contains potential for enhancement.

We plan to improve our identification and authentication systems in several ways. With respect to data collection, we intend to increase the number of users in the data set, collect more data per user. Instead of our own database we will try to evaluate our model on various online available data sources. More experimentation is needed regarding the placement of the device, instead of placing it in definite position and configuration we will try this placement to be natural so as to increase of user friendliness to a level up. The attack resistance of biometric gait recognition should also be analyzed. Studies by Gafurov [18] and Mjaaland [19] show that it is difficult for an attacker to imitate another person. This needs to be confirmed for the special scenario of mobile phones.

REFERENCES

[1] "FVC2006: the Fourth International Fingerprint Verification Competition," http://bias.csr.unibo.it/fvc2006/results/O_res_db2_a.asp, [Online; accessed 28-March-2017].
[2] P. Abeni, P. Baltatu, and R. D'Alessandro, "A face recognition system for mobile phones," ISSE 2006 — Securing Electronic Business Processes, pp. 211–217, October 2006
[3] D. J. Fish and J. Nielsen, "Clinical assessment of human gait", in Journal of Prosthetics and Orthotics 2, April 1993.
[4] L. Lee, "Gait Analysis for classification", Massachusetts Institute of Technology-artificial Intelligence Laboratory, 2003.
[5] L. Wang, T. Tan, W. Hu and H. Ning, "Automatic gait recognition based on statistical shape analysis", in IEEE Transactions on Image Processing, 2003.
[6] J. Suutala, K. Fujinami and J. Roning, "Gaussian process person identifier based on simple floor sensor", in Proceedings of the 3rd European Conference on Smart Sensing and Context, 2008.
[7] A. Annadhorai, E. Guenterberg, J. Barnes and K. Haraga, "Human identification by gait analysis", in Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments, 2008.
[8] Jamie Shotton, Toby Sharp and Pushmeet Kohli, "Decision Jungles: Compact and Rich Models for Classification".
[9] gbas data upload utility, https://script.google.com/macros/s/AKfycbxEiWElCj9-CpHwsnGP9HfrcIuaZgGOMfquRidZQmbhkJ36NdM/exec
[10] C. Nickel, H. Brandt, and C. Busch, "Classification of acceleration data for biometric gait recognition on mobile devices." BIOSIG, vol. 11, pp. 57–66, 2011.
[11] M. O. Derawi, C. Nickel, P. Bours, and C. Busch, "Unobtrusive

user-authentication on mobile phones using biometric gait recognition," in Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on. IEEE, 2010, pp. 306–311.

[12] H. M. Thang, V. Q. Viet, N. D. Thuc, and D. Choi, "Gait identification using accelerometer on mobile phone," in Control, Automation and Information Sciences (ICCAIS), 2012 International Conference on. IEEE, 2012, pp. 344–348.

[13] L. Hyafil and R. L. Rivest. Constructing optimal binary decision trees is NP-complete. Information Processing Letters, 5(1):15–17, 1976.

[14] J. R. Quinlan. Induction of decision trees. Machine Learning, 1986.

[15] J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

[16] K. V. S. Murthy and S. L. Salzberg. On growing better decision trees from data. PhD thesis, John Hopkins University, 1995.

[17] J. C. Platt, N. Cristianini, and J. Shawe-Taylor. Large margin DAGs for multiclass classification. In Proc. NIPS, pages 547–553, 2000.

[18] D. Gafurov, "Performance and security analysis of gait-based user authentication," Ph.D. dissertation, Faculty of Mathematics and Natural Sciences, University of Oslo, 2008.

[19] B. B. Mjaaland, "Gait Mimicking - Attack Resistance Testing of Gait Authentication Systems, Master Thesis, Norwegian University of Science and Technology," 2009.