Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference

Gender and Body Mass Index Classification Using a Microsoft Kinect Sensor

Virginia O. Andersson, Livia S. Amaral, Aline R. Tonini and **Ricardo M. Araujo**
PPGC - Federal University of Pelotas
Rua Gomes Carneiro, 1
Pelotas, RS, Brazil - 96010-610

Abstract

This paper shows results on gender and body mass index (BMI) classification using anthropometric and gait information gathered from subjects using a Microsoft Kinect sensor. We show that it's possible to obtain high accuracies when identifying gender, but that classifying BMI is a harder task. We also compare the performance of different machine learning algorithms and different combination of attributes, showing that Multilayer Perceptron outperforms Support Vector Machine and K-Nearest Neighbours in the proposed tasks.

Introduction

Person identification systems using full-body information from subjects (such as anthropometry and gait) is a recent trend in biometric studies. These systems may not require individuals' contact with sensors and can be performed at distance. The effort applied in video processing of walking subjects can be substantially simplified when using new sensor technologies, such as the Microsoft Kinect. The Microsoft Kinect sensor is a human motion tracker that do not require markers, special clothing or contact with the individual. It was developed as a companion for the Microsoft X-Box 360 video game console and used to power a gesture-based interface. The sensor is able to segment and extract three-dimensional representations of the major human joints, allowing the reconstruction of a simplified skeleton.

Previous work (Andersson and Araujo 2015) attempted to perform person identification using attributes extracted from Kinect sensors. The present paper seeks to answer what other useful features can be identified using anthropometric information and human gait, obtained from the Kinect sensor. We explore the possibility of classifying gender and body mass index (BMI) of subjects and compare the results of applying three machine learning algorithms: k-Nearest Neighbour (KNN), Multilayer Perceptron (MLP) and Support Vector Machine (SVM). We also report on a prototype system based on the ideas in this paper, deployed for a gender classification application in an e-commerce context.

Related Work

Several previous works analyse the feasibility of performing person identification using Microsoft Kinect sensors. (Sivapalan et al. 2011) proposes the use of Gait Energy Volumes to extract gait patterns from walking subjects and used data from a Kinect sensor to validate the proposal. In (Munsell et al. 2012) the positions and motion patterns of the different joints provided by the sensor are used as attributes to train both a SVM and a statistical model. (Preis, Kessel, and Werner 2012) compares the accuracy of Decision Trees (C4.5) and Naive Bayes classifiers. In (Araujo, Graña, and Andersson 2013) body segments lengths derived from joint positions were used to train a KNN classifier. Gender classification was recently explored by (Collins, Miller, and Zhang 2014), where the authors presented a modified framework that previously classified human behaviour using sparse spatiotemporal features from video clips of 101 pedestrians walking in a treadmill. They used Support vector machine algorithm to classify the videos and achieved 87% of accuracy.

All of these previous work are based on very small data sets, ranging from 8 to 25 individuals. (Hofmann and Bachmann 2012) uses a more extensive data set composed of 176 subjects but, although data was captured using a Kinect sensor, there was no attempt at inferring or using skeletal information and only depth data and regular video were used. In (Andersson and Araujo 2015) a data set composed of skeletal data from 140 subjects is used and KNN is shown to be better at the task when compared to SVM and MLP.

Methodology

In this work, we use the data set described in (Andersson and Araujo 2015). This data set contains information from 140 subjects (95 men and 45 women) walking in front of a Kinect sensor (2010 version). Each subjected executed 5 walks each with 6 to 12 gait cycles and 400 to 800 frames. Each frame contains three-dimensional data for each tracked joint. We filtered the data set and included only subjects that provided information on their gender, height and weight. For gender classification, the resulting data set totalled 112 subjects (68 men and 44 women). For BMI classification, the data set totalled 106 individuals with height and weight information.

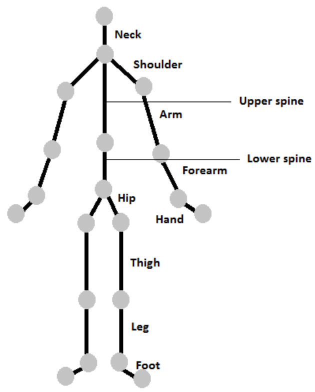

Figure 1: Main joints tracked by Kinect sensor (shown as circles) and the body segments obtained by the Euclidean distance between two tracked joints.

In order to reduce noise present in raw skeleton data provided by Kinect sensor, we applied an Auto Regressive Moving Average (ARMA) filter (Azimi 2012) with a 8 frame size window, in all walk examples selected, as proposed in (Andersson and Araujo 2015).

Anthropometric and Gait attributes

The measurement of several body segments were calculated for each frame by the Euclidean distance between two tracked joints, as proposed in (Araujo, Graña, and Andersson 2013) (Fig. 1). This resulted in a total of 20 attributes per frame. Finally, all frames from each single walk were grouped by taking the arithmetic mean over all frames. Hence, each example represents a single walk of a single subject.

Gait attributes were separated into spatiotemporal and kinematic parameters following (Andersson and Araujo 2015) methodology. Spatiotemporal parameters were calculated based on the average step length of gait cycles present in a walk example. Kinematic parameters were obtained from the angles' curves generated during the entire period of walking. For each frame, the angle between two lower segments described the hip, knee, ankle and foot angle rotation during a captured instant.

A total of 56 kinematic attributes were extracted from the angles' curves for right and left hips, knees and ankles. For the foot angle, it was only used the extension (valleys) information. Spatiotemporal parameters totalled 4 attributes. Together with anthropometric attributes, we obtained 80 gait attributes to compose our dataset.

After all anthropometric and gait attributes were obtained, we separated the dataset with gender and BMI classes by type of attributes: (i) Gait only, with all spatiotemporal and kinematic extracted attributes, (ii) anthropometry only, with the measurements of body segments and height and (iii) all attributes.

Training and Testing

For gender classification, we randomly selected 44 men to have a balanced data set. The examples belonging to each individual were divided into two disjunct sets of equal sizes for training and testing (each containing 44 persons). We made sure that no subject's examples were present in both training and testing sets - a random split over all examples would allow for examples from the same subject to be present in both sets, making it possible for the classifier to identify the subject instead of the metrics of interest.

The BMI classification is a simple index of body mass used to classify adults as being below, above or well above an ideal body weight (WHO 2014), as presented in Eq. 1:

$$BMI = \frac{weight(kg)}{(height(m))^2} \tag{1}$$

We calculated the BMI index for each subject, grouping them as healthy, above the recommended BMI (overweight) and below the recommended BMI (underweight). We randomly chosen 20 examples per class from the height and weight data set used, to balance with 40 examples of underweight class, for train and test sets.

Machine learning algorithms

Three machine learning algorithms were chosen to be applied in our datasets, as they are often used in the biometric literature: Support Vector Machine (SVM), K-Nearest Neighbour (KNN) and Multilayer Perceptron (MLP). The implementation present in the Weka library (Hall et al. 2009) was used.

For the KNN algorithm, $K = 3$ was set, and Manhattan distance as the distance metric. Neighbours' distances d_i were weighted with a weight of $1/d_i$.

The MLP was set to have a single hidden layer with 4 hidden units. Sigmoidal activation function was used for all units. Training was performed using the Backpropagation algorithm with momentum set to 0.2, learning rate to 0.3 and 120 epochs (set by early stopping).

The SVM was trained using the Sequential Minimal Optimization (SMO) algorithm (Platt 1999), using a polynomial kernel and $C = 100.0$.

When evaluating the algorithms, we use the Area Under the Curve (AUC) of a ROC diagram to compare the classifiers in the gender classification task (as it is binary) and accuracy in the BMI classification task.

Real-World Application

The used data set contains data from subjects walking in relatively homogeneous and controlled conditions. In order to test the ability of the trained classifiers to identify gender in a real world setting, we implemented a prototype application of gender classification to be used in a conference demonstration. The demonstration involved a digital showcase, where subjects in front of the sensor were shown customized advertisements on a display based on their gender. Since subjects were expected to stand in front of the sensor, embedded in the showcase, only anthropometric measurements were used.

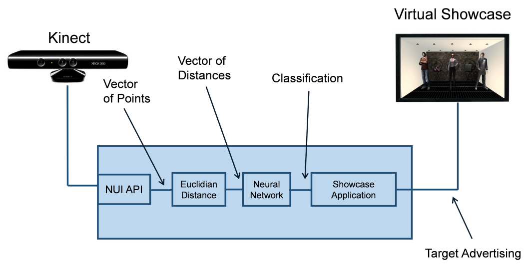

Figure 2: Customizable showcase application with the main modules implemented.

The showcase was implemented according to the diagram shown in Fig. 2, where the Kinect sensor captures an individual and the NUI Skeleton API returns referring joints 3D points tracked by the sensor during the time that the same is stopped. A module is responsible for calculating the attributes and feeding the attribute vector to the classifier. As the classifier we used the best-performing classifier (MLP) trained with all examples from the data set.

Results

Results Using the Data set

This section presents results from applying the classifiers in the examples from the data set. Initial results from the real-world application are presented in the next section.

Gender classification results using SVM, KNN and MLP algorithms are shown in Table 1, where AUC values are presented. The datasets were separated by type of attributes used. It is possible to observe that all classifiers did reasonably well in the task but the MLP attained a good score, performing consistently well across different attributes (87.7% accuracy using all attributes). It is noteworthy that MLP performed considerably better than SVM and KNN, as these latter two are often preferred in the biometric literature. In (Andersson and Araujo 2015) it is shown that MLP performs much worse than the other two classifiers in a person identification task using the same data set. Since the input attributes are similar for both applications, one explanation for our results is that there are much fewer classes to be discriminated in the present task, whereas person identification demands hundreds of subjects to be identified. Also, in opposition to person identification task, where gait data from this dataset was shown to be only marginally useful (Andersson and Araujo 2015), our results show that gait information alone has a reasonable discriminative power (81.4% accuracy using MLP) compared to when using anthropometric information only.

Table 2 shows that both Kinematic and Spatiotemporal gait information perform about the same when used separately, but both underperformed the case where they are used together.

For BMI classification, the main results are shown in Table 3. Overall performance is poor but, with the exception of KNN, still better than random. SVM and MLP perform about the same, but KNN displayed much worse accuracies.

Table 1: Classifiers' ROC area under the curve on gender classification with different attributes.

Classifier	Gait	Anthropometry	All
SVM	0.837	0.796	0.837
KNN	0.845	0.843	0.882
MLP	**0.902**	**0.931**	**0.957**

Table 2: Classifiers' ROC area under the curve on gender with different combinations of gait attributes.

Classifier	Kinematic	Space-temporal	Gait
SVM	0.702	0.762	0.837
KNN	0.788	0.770	0.845
MLP	**0.816**	**0.854**	**0.902**

More importantly, we can see that, contrary to the previous results on gender, including anthropometry *reduces* performance and gait only data shows the most discriminative power. This is even more evident in Table 4, where we can observe that using only Spatiotemporal attributes provides the best results across all tested combinations.

Anthropometric data is somewhat expected to be a poor predictor of BMI, as BMI puts greater emphasis on weight and Kinect data can only provide approximate height information. This explains why KNN performed much worse when using all attributes, since it is more susceptible to the "curse of dimensionality" (Schuh, Wylie, and Angryk 2014); when including anthropometric attributes we are effectively including many non-informative attributes that end up misguiding the distance metric.

These results are evidence that gait information can be useful to estimate BMI and that spatiotemporal attributes are the most useful set of gait attributes in this task.

Results from Real-World Application

A total of 13 subjects (6 female and 7 males) participated in the demonstration. The accuracy obtained was of 61%, much lower than the results obtained with the more controlled data set. Classifying women yielded a false positive rate of 33% and classifying men a false positive rate of 42%.

Three factors are likely to have had an impact in the lower accuracy obtained. First and most importantly, much fewer frames were captured from the subjects when compared to the controlled experiment, since ads had to be shown quickly. Around 90 frames were captured per subject before delivering a classification, whereas up to 10 times that was used in the original data set. Fewer frames leads to a higher standard deviation of the measurements and consequently to

Table 3: Classifiers' accuracy on BMI classification with different attributes.

Classifier	Gait	Anthropometry	All
SVM	53.3%	**40.0**%	50.0%
KNN	43.3%	25.0%	23.3%
MLP	**55.0**%	35.0%	**51.6**%

Table 4: Classifiers' accuracy on BMI with different combinations of attributes, compared with all gait attributes.

Classifier	Kinematic	Spatiotemporal	All
SVM	48.3%	53.3%	53.3%
KNN	40.0%	51.6%	43.3%
MLP	**53.3**%	**60.0**%	**55.0**%

less reliability. Second, and adding to the previous problem, the ARMA filter was not implemented in the prototype to further reduce the response time. Finally, subjects were captured facing towards the sensor, whereas the training data contains mostly frames of subjects on their sides.

Despite the low accuracy, the obtained results are still better than random and encouraging. Allowing for more frames to be captured and training the classifier under a more diversified set of conditions can likely improve considerably the results in real-world scenarios.

Conclusions

This paper reported on results of training classifiers to identify gender and body-mass index of subjects, using data collected from a Microsoft Kinect (2010 version) sensor. We used a previously published data set (Andersson and Araujo 2015) containing skeleton models, used originally for person identification, that also contained information on height, weight and gender for most of the subjects.

Our results provide evidences that gender identification is possible using Kinect data and that both anthropometric measurements and gait information are valuable for this task, while previous works showed that gait information on this data set is only marginally useful for person identification. However, the trained classifiers, while displaying good accuracies in controlled settings (up to 87%), were not able to fully generalize the task when deployed in uncontrolled conditions, showing much lower accuracies (an average of 61%). We argued that this was due to the constraints the system was subjected to.

Classifying BMI, on the other hand, proved to be a harder problem. Using all available attributes yielded an accuracy close to 50% (comparatively, a random choice would yield 33% since there were three balanced classes). In this case, anthropometric attributes were barely useful to train the classifiers; gait attributes, in particular spatiotemporal ones, were responsible for almost the totality of the response and, when used alone, allowed the classifiers to perform significantly better (up to 60% accuracy). These results show that there are gender and BMI signatures in data extracted from the Kinect sensor and that these can be useful in a number of scenarios ranging from automatically customized interfaces and advertisements to fitness software. Finally, by comparing different classifiers, and again in contrast to person identification tasks, we showed that MLP outperforms both SVM and KNN in these two tasks.

The main contribution of this paper is to provide initial results on gender and BMI classification using Kinect data and comparing the performance of different classifiers and attributes combinations in these tasks. Future venues of work include testing the methodology using other data sets and further improving the generated prototype to be able to attain accuracies comparable to what is obtained in more controlled settings.

References

Andersson, V., and Araujo, R. 2015. Person identification using anthropometric and gait data from kinect sensor. In *Proceedings of the Twenty-Ninth Association for the Advancement of Artificial Intelligence Conference*, AAAI 2015. "to appear".

Araujo, R. M.; Graña, G.; and Andersson, V. 2013. Towards skeleton biometric identification using the microsoft kinect sensor. In *the 28th Annual ACM Symposium*, 21–26. New York, New York, USA: ACM Press.

Azimi, M. 2012. Skeleton joint smoothing white paper. Technical report, Microsoft Inc.

Collins, M.; Miller, P.; and Zhang, J. 2014. Gait based gender recognition using sparse spatio temporal features. In Gurrin, C.; Hopfgartner, F.; Hurst, W.; Johansen, H.; Lee, H.; and OConnor, N., eds., *MultiMedia Modeling*, volume 8326 of *Lecture Notes in Computer Science*. Springer International Publishing. 80–91.

Hall, M.; Frank, E.; Holmes, G.; and Pfahringer, B. 2009. The weka data mining software: an update. *ACM Conference on Knowledge Discovery and Data Mining*.

Hofmann, M., and Bachmann, S. 2012. 2.5d gait biometrics using the depth gradient histogram energy image. In *Proceedings of the IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 399 – 403.

Munsell, B. C.; Temlyakov, A.; Qu, C.; and Wang, S. 2012. Person identification using full-body motion and anthropometric biometrics from kinect videos. In Fusiello, A.; Murino, V.; and Cucchiara, R., eds., *ECCV Workshops (3)*, volume 7585 of *Lecture Notes in Computer Science*, 91–100. Springer.

Platt, J. C. 1999. Advances in kernel methods. Cambridge, MA, USA: MIT Press. chapter Fast Training of Support Vector Machines Using Sequential Minimal Optimization, 185–208.

Preis, J.; Kessel, M.; and Werner, M. 2012. Gait recognition with kinect. *Workshop on Kinect*.

Schuh, M. A.; Wylie, T.; and Angryk, R. A. 2014. Mitigating the curse of dimensionality for exact knn retrieval. In *Proceedings of the Twenty-Seventh International Flairs Conference*, 363–368.

Sivapalan, S.; Chen, D.; Denman, S.; Sridharan, S.; and Fookes, C. 2011. Gait energy volumes and frontal gait recognition using depth images. In Jain, A. K.; Ross, A.; Prabhakar, S.; and Kim, J., eds., *IJCB*, 1–6. IEEE.

WHO. 2014. Bmi classification. $http://apps.who.int/bmi/$. World Health Organization. Last access: August 2014.