Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Person Identification Using Anthropometric and Gait Data from Kinect Sensor

Virginia O. Andersson and **Ricardo M. Araujo**
PPGC - Federal University of Pelotas
Rua Gomes Carneiro, 1, Pelotas, RS, Brazil - 96010-610
vandersson@inf.ufpel.edu.br, ricardo@inf.ufpel.edu.br

Abstract

Uniquely identifying individuals using anthropometric and gait data allows for passive biometric systems, where cooperation from the subjects being identified is not required. In this paper, we report on experiments using a novel data set composed of 140 individuals walking in front of a Microsoft Kinect sensor. We provide a methodology to extract anthropometric and gait features from this data and show results of applying different machine learning algorithms on subject identification tasks. Focusing on KNN classifiers, we discuss how accuracy varies in different settings, including number of individuals in a gallery, types of attributes used and number of considered neighbors. Finally, we compare the obtained results with other results in the literature, showing that our approach has comparable accuracy for large galleries.

Introduction

Biometric systems is an important application area for artificial intelligence in general and machine learning in particular. As these systems become more popular, with applications ranging from surveillance to entertainment (Wang 2012), a better handling of lower-quality data from cheaper sensors becomes necessary.

A relatively recent trend in biometrics is performing person recognition using full-body characteristics, including anthropometric measurements of body parts and dynamic gait features from walking patterns. This is often performed with the use of video cameras and complex feature-extraction algorithms from video footages.

The Microsoft Kinect device is a set of sensors with an accompanying Software Development Kit (SDK) that is able to track movements from users by using a skeleton mapping algorithm. The first version of the device is able to track 48 skeletal points at 30 frames per second and is used in the X-Box 360 video-game console as an input for a gesture-based interface.

While this device is primarily aimed at tracking users that are standing in the same place, it has been used to track walking subjects in order to extract gait information (Araujo, Graña, and Andersson 2013). This use provides a cheaper, off-the-shelf, alternative to complex multi-part video setups.

However, several limitations, including range, lighting and occlusions degrade the accuracy of the data provided by the device and impose significant challenges to its use in a biometric system. Hence, while using the Kinect device greatly simplifies the data capturing stage, it may require improved techniques to handle these limitations.

In this paper, we provide a comparison of machine learning algorithms applied to a large data set comprised of walking subjects captured using the 2010 version of the Kinect device, aiming at person recognition. We focus on the K-Nearest Neighbors algorithm that is widely used in the literature, but also provide comparisons with Multi-Layer Perceptrons and Support Vector Machines.

Related Work

The use of body measurements aiming at subject identification can be traced back to mid-XIX century in the France police (Harrap 1956). (Godil, Grother, and Ressler 2003) performed an extensive analysis of the effectiveness of using body measurements and shape for identification purposes, using the CAESAR database, which collects accurate static anthropometric data using markers.

The use of joint positions collected using a Kinect device appears in (Munsell et al. 2012) as part of a broader set of features that included motion patterns; a SVM was used for classification along with a statistical model. In (Araujo, Graña, and Andersson 2013) body segment lengths derived from joint positions were used exclusively to train a KNN classifier. Both cases used only a very limited data set with the former using 20 individuals and the latter 8.

Using human gait as a biometric feature was motivated by evidences that individuals describe unique patterns during their gait cycles (Murray, Drought, and Kory 1964). Approaches used to extract and analyze human gait can be classified as (i) model-based, where the human gait is described by gait theory fundamentals and is reconstructed through a model, e.g. "stick figure", that fits the person in every gait sequence frame; the spatiotemporal and kinematic parameters, then, are extracted from the model during gait cycle; and (ii) model-free, where features such as silhouette or enclosing boxes are used as attributes, without explicitly considering human gait fundamentals (Ng et al. 2011).

Recent model-based approaches include (Cunado, Nixon, and Carter 2003), where gait analysis was restricted to the

use of sagital rotations of hips, knees, and ankles and a KNN classifier was used to perform person identification. KNN is also used in (Yoo and Nixon 2011), where a detailed method to extract simplified skeleton figures from videos is presented and gait features are extracted from these figures.

Model-free approaches include the concept of Gait Energy Images (Han and Bhanu 2006). In (Sivapalan et al. 2011), Gait Energy Volumes are proposed as an extension and a database containing depth information on 15 subjects walking towards a Kinect sensor is used to test the methodology, using an KNN-based approach.

In (Hofmann and Bachmann 2012), the authors proposed the use of Depth Gradient Histogram Energy Image to improve identification when many more subjects are being identified, reporting high accuracy (81%-92%) over a data set created using a Kinect sensor; however, the database used contains only very few gait cycles due to a fixed sensor placement and only depth and regular video data is used. Therefore this approach focused on generic object-motion characteristics, without considering gait signature information. Again, a KNN classifier is used.

Compared to previous approaches, the present paper makes use of a comprehensive (140 subjects) data set captured using a Kinect sensor, extends the attributes to include both body measurements and model-based gait information and provides more in-depth experiments using the data, including insights on the used machine learning classifiers.

Methodology

The Kinect Sensor

The Kinect device used in our experiments was the 2010 model for the X-Box 360 video-game console, connected through an adapter cable to a PC running the SDK version 1.0. This device is equipped with a RGB camera and a depth sensor composed of an infrared light emitter and a infrared-sensitive camera.

A software library, called NUI API (Natural User Application Programming Interface), retrieves and process data from the sensors. This API is responsible for providing detailed information about the location, position and orientation of individuals located in front of the sensor. This information is provided to the application as a set of 48 three-dimensional points called "skeleton points". These points approximate the main joints of human body and the actual position of the individual in front of the sensor.

The API provides data in the form of frames at the rate of 30 frames per second. Each frame contains an array containing all the extracted points at the moment of the capture.

Capturing Methodology

Volunteers walked in front of the sensor in a semi-circular trajectory while data was being recorded. A spinning dish was used to help move the Kinect sensor to follow the person during the walk. This combination of trajectory and Kinect's pan camera movement allows several gait cycles to be captured per individual without distortions caused by subjects moving in or out of the sensor's field of view. Each subject executed five round trip free cadence walk, starting on the left of the sensor, walking clockwise to the right of the sensor and then back.

The volunteers were recruited for the experiment at a university campus. The majority of subjects were college students, with ages between 17 and 35 years old. Each of the subjects that accepted to participate in the experiment provided gender, height and weight information. They were wearing light clothing, since the captures were conducted during summer. The captures were conducted in an empty classroom at day time with mostly artificial lighting. A total of 140 individuals were captured using the proposed methodology (95 men and 45 women). In most cases, each individual generated about 500 to 600 frames and completed between 6 and 12 gait cycles per walk.

Raw skeleton data often presents noise in the joint positions due to errors in the tracking process. In order to reduce this noise we applied an Auto Regressive Moving Average (ARMA) filter (Azimi 2012) with a window of size 8, set in an ad hoc fashion by observing a visual reconstruction of walks before and after the filter. This filter was applied to all walk samples before being used.

From the captured raw data, attributes were extracted for each walk, composing labeled examples where the label is an anonymized identifier of an individual. The attributes are divided in two sets: *gait attributes* and *anthropometric attributes*. In what follows we describe how each attribute is defined. The full data set is available at http://ricardoaraujo.net/kinect.

Gait Attributes Model-based gait analysis considers the human gait theory to help extract parameters from human walk. The angles described by the joints of the hips, knees and ankles, known as kinematic parameters, were calculated for each frame captured, using the pendulum model proposed in (Cunado, Nixon, and Carter 2003) and depicted in Figure 1 (b). Furthermore, we calculate the foot angle, described in (Murray, Drought, and Kory 1964), depicted in Figure 1 (c) and the spatiotemporal parameters described in (Yoo and Nixon 2011): the step length, stride length (or "gait cycle size"), cycle time and velocity.

As shown in Figure 1 (b), the angle θ is formed during a gait cycle between the segments of the thigh and a projection of the hip. Between the leg and the knee projection the angle γ is defined; α is the angle of the ankle rotation formed by the foot segment and the ankle projection and the foot angle β is formed by the opening of the foot in relation to the axis of the heel. These angles describe periodic curves during a walk, which can have useful characteristics for biometric recognition (Harrap 1956).

The periodic curves generated by the lower joint angles are composed by flexion and extension phases, visually noticed by peaks (flexion) and valleys (extension) (Murray, Drought, and Kory 1964). The arithmetic average and standard deviation were computed for the flexion peaks and extension valleys in order to characterize the curves of each individual. Lower and higher flexion peaks and extension valleys were treated separately, generating an arithmetic average and standard deviation for each high and low phase. Each lower joint was considered independent of the others,

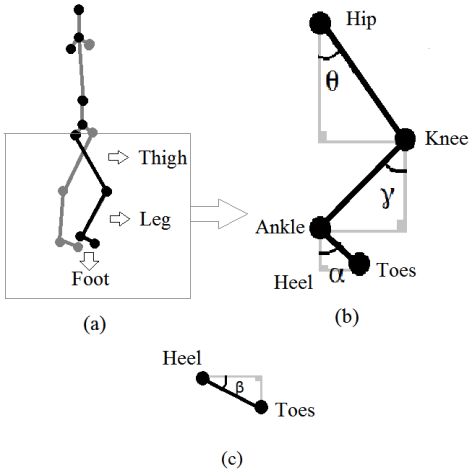


Figure 1: Angles tracked to compose gait attributes.

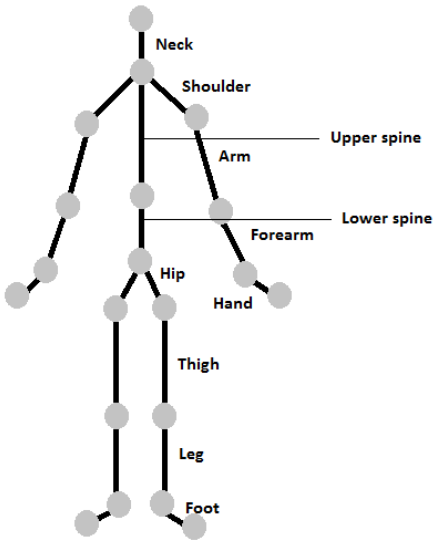


Figure 2: Tracked joints (circles) and body segments used as attributes.

generating attributes equally independent.

Spatiotemporal parameters were calculated based on the step length and frame rate of the Kinect sensor. The step length was obtained by averaging the highest values of the difference between the right and left heels. In addition, we use as attributes the stride length (Eq. 1), average stride length over all n strides (Eq. 2), cycle time (Eq. 3) and velocity (Eq. 4). A total of 60 gait attributes were defined.

$$strideLength = 2 * stepLength \tag{1}$$

$$avgStrideLength = \sum_{i=1}^{n} \frac{strideLength}{n} \tag{2}$$

$$cycleTime = \frac{avgCyclePeriod}{30} \tag{3}$$

$$velocity = \frac{avgStrideLength}{cycleTime} \tag{4}$$

Anthropometric Attributes For each frame captured the measurements of several body segments, shown in Figure 2, were calculated using the Euclidean distance between joints, in a similar fashion to the methodology employed in (Araujo, Graña, and Andersson 2013). The subject's height was defined as the sum of the neck length, upper and lower spine length and the averages lengths of the left and right hips, thighs and lower legs.

The mean and standard deviation of each body segment and height over all frames of a walk were calculated. Measurements beyond two standard deviations from the mean were discarded and attributed to noise. The recalculated means for each part were used as attributes, totaling 20 anthropometric attributes.

Classifiers

From our literature review, K-Nearest Neighbor (KNN) classifier is the most commonly used model in full-body biometrics, followed by Support Vector Machines (SVM). In this section, we compare these two models applied to our data set and, in addition, also Multi-Layer Perceptron (MLP).

Parameters for each algorithm were set by varying their values while trying to maximize the resulting accuracy using a 10-fold cross-validation (Mitchell 1997) over a smaller random validation subset composed of 20 users. The same number of evaluations was performed for each classifier. All attributes were normalized before use by mapping their values to the range $\{-1, 1\}$.

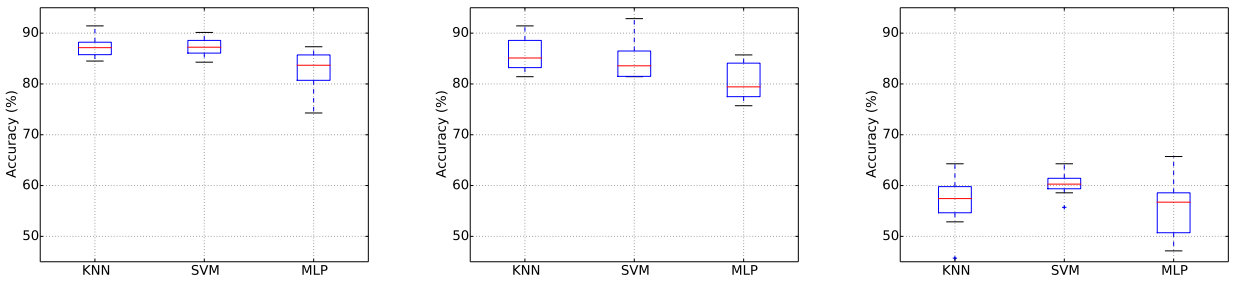
KNN was set to $K = 5$, Manhattan distance as the distance metric and distance weighting of $1/d$. For the MLP, we only considered networks with a single hidden layer and the number of hidden units was set to 40. Training was performed using the Backpropagation (Haykin 2008) algorithm with momentum set to 0.2, learning rate to 0.3 and 1000 maximum epochs. The SVM was trained using the Sequential Minimal Optimization (SMO) algorithm (Platt 1999), using a polynomial kernel and $C = 100.0$.

We use 10-fold cross-validation to validate the models with the above parameters i.e. the data set was randomly partitioned in 10 subsets and training was performed ten times, each time leaving one partition out of the training process, which was used for testing; the reported accuracies are the averages of these ten executions. When required, statistical significance tests are performed using a Wilcoxon signed-rank test (Wilcoxon 1945).

Results

Classifiers Accuracy

Figure 3 plots accuracy data over the 10 validation folds for each algorithm and data set. Table 1 shows the mean values for each case.



(a) All attributes. (b) Anthropometric attributes. (c) Gait attributes.

Figure 3: Accuracy boxplots for each classifier and different attributes.

Overall, we can observe that only using gait attributes leads to a poor performance, much worse than using only anthropometric attributes. Nonetheless, combining the two subsets allows for higher accuracy.

In all cases MLP performed consistently worse than KNN and SVM. When using all attributes, KNN and SVM performed about the same; the difference in the means is not statistically significant ($p = 0.991$).

When considering only anthropometric attributes, KNN displayed slightly better mean, but not very statistically significant ($p = 0.248$). The results for KNN are, however, more consistent, with smaller standard deviation and a higher median.

For gait attributes SVM has a small, but statistically significant ($p = 0.032$), lead over KNN. It is also more consistent and has a higher median. While MLP still performs poorly over gait attributes, the difference is not as accentuated as for the other cases.

These results lead to the conclusion that KNN and SVM perform about equally well on the problem, but the chosen MLP performs considerably and consistently worse. This is consistent with the literature, where KNN and SVM are often used. The preference for KNN may be due to it being easy and fast to train, allowing for quicker experimentation.

The comparatively small increase in accuracy when combining both types of attributes shows that gait attributes do not provide much value beyond anthropometric attributes, an evidence that the two are somewhat correlated. It is clear that the latter is responsible for most of the response, with gait attributes contributing only an average of 3.3 percentage points. Nonetheless, gait attributes do show a measurable contribution and by themselves are reasonably useful (much better than random) for person identification.

Table 1: Classifiers' mean accuracy using different attributes. Bold text highlights the best values for each column.

Classifier	Gait	Anthropometric	All
SVM	**62.9**%	84.7%	86.3%
KNN	59.5%	**85.4**%	**87.7**%
MLP	59.2%	79.7%	84.7%

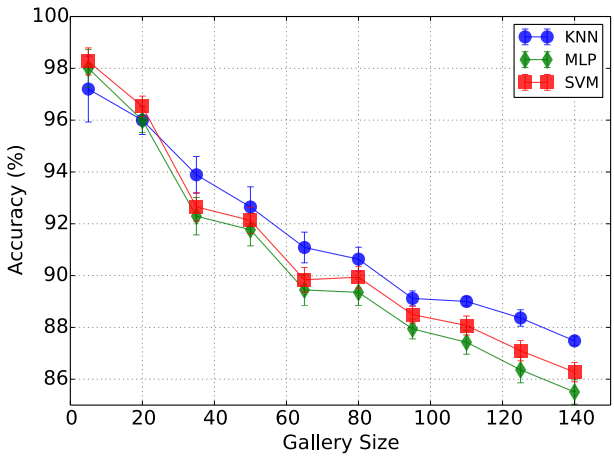


Figure 4: Classifiers' average accuracy for different gallery sizes using all attributes. Error bars are 95% confidence intervals.

Gallery Size

While the results in Table 1 are reasonable and useful for a number of application (possibly excluding strict authentication), they may seem worse than results presented in similar previous works - e.g. (Araujo, Graña, and Andersson 2013) where upwards of 98% accuracy was reported, but for only 9 subjects. One key missing aspect of these previous work is an account for how accuracy varies with the size of the gallery being identified.

Figure 4 shows how accuracy evolves when subjects in increasingly large galleries must be identified. Each point (except for 140) is the average over 10 galleries with the same size and randomly drawn from the complete data set. For very small galleries, accuracies are close to 98%, steadily converging to the results seen in Table 1.

Additionally, we can observe that KNN actually performs worse for small galleries, only becoming better than SVM and MLP when considering more than 20 individuals.

Figure 5 shows how accuracy vary for different gallery sizes and when using different attribute sets. Again we can observe that gait attributes are overall far less useful for the

task than anthropometric attributes. Nonetheless, for very small galleries ($N = 5$ in the figure), using only gait attributes provides reasonable performance.

Accuracy using gait attributes degrades much faster when increasing gallery size. Nonetheless, it is for large galleries that using gait information in addition to anthropometric attributes is useful. For galleries of size 105 or less there are no statistically significant differences in accuracy between using gait information or not (assuming significance level of 0.05) but there are significant, if rather small, differences for larger galleries.

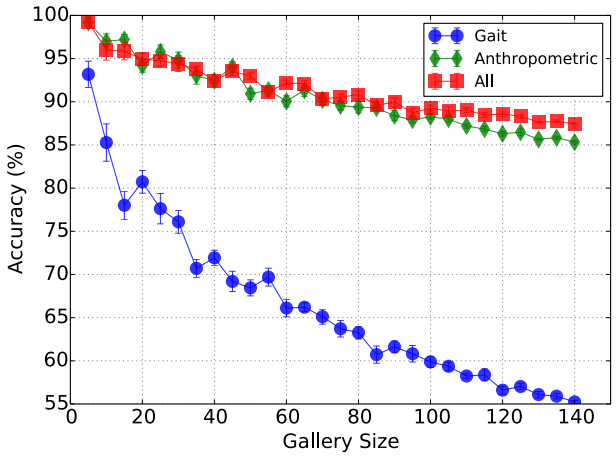


Figure 5: Average KNN accuracy for different gallery sizes and different attribute sets. Error bars represent 95% confidence intervals.

Number of Neighbors

While we used the same value of K for all gallery sizes when using KNN classifier, one may consider how this parameter affects overall performance. In order to provide an answer, for each gallery size we varied the value of K from 1 to 101 (or the maximum number of individuals in the gallery; odd values only were considered) and measured the resulting performance for each K, again as an average over 10 folds using cross-validation.

Figure 6 shows the best K for each gallery size, along with the accuracy obtained when using the best K. While the data is noisy, the general trend is clear: larger galleries benefit from larger values of K but these values do not vary considerably. Going from a gallery of 10 individuals to 140 individuals only increases the value of K from 1 to 5, a two-step change. The average improvement over a fixed $K = 5$ was of 0.6 percentage point.

Even though the best K varies little with gallery size, this parameter does have a strong effect in accuracy. Figure 7 shows how accuracy varies with K for different gallery sizes. Increasing K beyond the optimal value leads to an almost linear decrease in accuracy. This shows that accuracy is quite sensitive to K, even though the differences in optimal values for different gallery sizes do not vary significantly; hence, knowing the optimal value for some sizes allows for

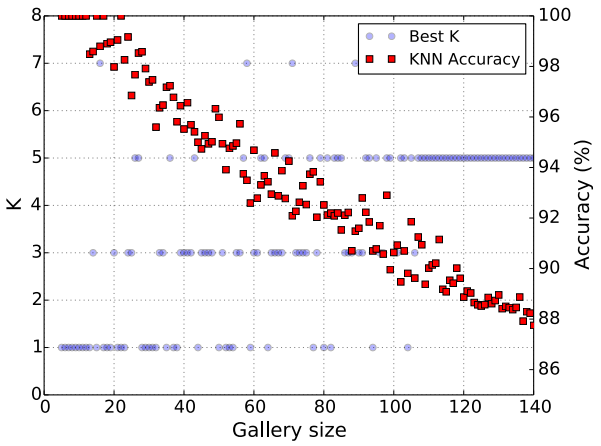


Figure 6: Best found K and KNN accuracy with best K for different gallery sizes.

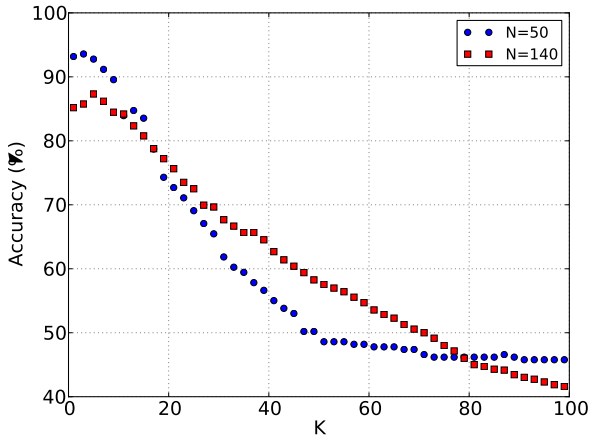


Figure 7: KNN accuracy for different values of K and two gallery sizes.

a quick search for optimal values for other sizes, since there seems to be no need for an exhaustive search over a wide range of values. A gradient search should suffice in finding this parameter.

Comparison to Other Approaches

Figure 6 allows for a rough comparison with previous similar works. For instance, (Yoo and Nixon 2011), using regular video cameras to extract 2D stick-figures information from individuals, reports an accuracy of 84.0% using KNN to classify 100 individuals, against 89.4% using our methodology and data set. However, they also report a 96.7% accuracy using 30 individuals, while ours is of 94.0%, an evidence that the adequacy of a methodology, including the chosen classifier and attributes, is conditional on the number of individuals being identified.

In (Hofmann and Bachmann 2012), only images' energies are used for identification - i.e. it does not try to reconstruct a

Table 2: Attributes selected from a Correlation-based Feature Subset Selection and how much removing each of the set affects KNN accuracy, in percentage points.

Attribute	Average Accuracy Change
Stride Length (gait)	-4.4%
Right Foot	-3.2%
Right Hand	-2.9%
Neck	-2.6%
Upper Spine	-1.9%
Right Shoulder	-1.7%
Left Hip	-1.7%
Height	-1.4%
Right Leg	-1.3%
Left Angle Peaks (gait)	-0.7%
Left Forearm	-0.6%
Right Thigh	-0.6%
Left Thigh	0.0%
Left Leg	0.0%

skeleton model from video images - and an 81.0% accuracy is reported for 176 individuals. While a more direct comparison is not possible given our more limited data set, fitting an exponential function to Figure 6, yields a projection of 84.5% for this same gallery size.

Relevant Attributes

The complete data set contains a total 80 attributes. In the provided experiments, all were used, but such large number of attributes lead to high computational requirements and may reduce accuracy if many attributes are irrelevant.

In order to better understand the role each attribute has in the obtained results, we applied a correlation-based feature subset selection (Hall 1998) to the data. The resulting subset contains only 14 attributes, shown in Table 2. While using this subset does not improve accuracy, it only performs an average of 0.1 percentage point worse when compared to the case where all attributes are provided.

It is also possible to observe that 12 of the 14 selected attributes are anthropometric and only 2 are related to gait. This reinforces our previous observation that gait features are largely correlated to anthropometric features but less reliable for identification. Nonetheless, Step Length, a gait attribute, provides the largest drop in accuracy when individually removed from this set.

Conclusions

We reported on the ability of different classifiers to discriminate walking individuals tracked by a Microsoft Kinect device, aiming at composing a full-body person identification system. In order to do so, we detailed how attributes could be extracted from the raw data and categorized these attributes as based on static body measurements (anthropometric) or motion patterns (gait). A data set composed of 140 subjects was used and three classifiers were tested (KNN,MLP and SVM).

The main contribution of this paper was to provide a set of simulations that allowed for the comparison between commonly used classifiers applied to a novel biometric data set. The experiments considered several different conditions, notably different gallery sizes and different combinations of attributes.

Our results showed that the defined gait attributes are less useful than anthropometric attributes when each is used separately; accuracies using only gait also fall faster as gallery size is increased. However, combining the two attribute types allows for higher accuracies when the number of subjects is large enough. By reducing the dimensionality of the data, we showed evidences that a very compact set of attributes is enough to ensure high accuracy and that in this smaller set, gait information is very relevant.

When comparing the classifiers, our results showed evidences that KNN and SVM display similar accuracies, while MLP perform considerably worse for all tested variations of attributes, with the exception of very small galleries. Focusing on KNN, we showed that the number of neighbors (K) does not vary considerably when very different gallery sizes are used. In addition, our best results are generally comparable to accuracies reported in the literature, but the comparison methodology does not allow for stronger claims.

In conclusion, a full-body person identification system was shown to be viable using data from a Kinect device and the proposed methodology, but the observed accuracies are not high enough for critical applications, suggesting that this approach should be used as a complement to other techniques (e.g. face recognition).

Future work include finding and testing better gait-based attributes that can be easily inferred from Kinect data and testing the trained classifiers in real-world uncontrolled scenarios.

Acknowledgments

This work is supported by CNPq (Brazilian National Research Council) through grant number 477937/2012-8.

References

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.

Cunado, D.; Nixon, M. S.; and Carter, J. N. 2003. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding* 90(1):1–41.

Godil, A.; Grother, P.; and Ressler, S. 2003. Human identification from body shape. In *Proceedings of the 4th International Conference on 3D Digital Imaging and Modeling*, 1–7. IEEE Computer Society Press.

Hall, M. A. 1998. *Correlation-based Feature Subset Selection for Machine Learning*. Ph.D. Dissertation, University of Waikato, Hamilton, New Zealand.

Han, J., and Bhanu, B. 2006. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(2):316–322.

Harrap, G. G. 1956. *Alphonse Bertillon: Father of scientific detection*. New York: Abelard-Schuman.

Haykin, S. 2008. *Neural Networks and Learning Machines (3rd Edition)*. Prentice Hall, 3 edition.

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.

Mitchell, T. 1997. *Machine Learning*. New York, NY, USA: McGraw-Hill, Inc., 1 edition.

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.

Murray, M. P.; Drought, A. B.; and Kory, R. C. 1964. Walking patterns of normal men. *The Journal of Bone Joint Surgery* 46:335–360.

Ng, H.; Tong, H.-L.; Tan, W. H.; and Abdullah, J. 2011. Improved gait classification with different smoothing techniques. *International Journal on Advanced Science, Engineering and Information Technology* 1(3):242–247.

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.

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.

Wang, L. 2012. Some issues of biometrics: technology intelligence, progress and challenges. *International Journal of Information Technology and Management* 11(1/2):72.

Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1(6):80–83.

Yoo, J.-H., and Nixon, M. S. 2011. Automated markerless analysis of human gait motion for recognition and classification. *ETRI Journal* 33(2):259–266.