Gait Analysis Using Multiple Kinect Sensors

Gabriele Maida and Marco Morana

Abstract A gait analysis technique to model user presences in an office scenario is presented in this chapter. In contrast with other approaches, we use unobtrusive sensors, i.e., an array of Kinect devices, to detect some features of interest. In particular, the position and the spatio-temporal evolution of some skeletal joints are used to define a set of gait features, which can be either static (e.g., person height) or dynamic (e.g., gait cycle duration). Data captured by multiple Kinects is merged to detect dynamic features in a longer walk sequence. The approach proposed here was been evaluated by using three classifiers (SVM, KNN, Naive Bayes) on different feature subsets.

1 Introduction

Gait analysis is a biometric technique that aims at identifying people by their walking style. Gait recognition techniques can be categorized into three categories [9], involving the use of computer vision methods, floor sensors, or wearable sensors.

The main advantage of computer vision approaches is that the identification can be performed without physical contact between subjects and data acquisition devices. Various chapters have looked into processing the images captured by standard video cameras and analyzing the human silhouette [13, 24] in order to extrapolate characteristic features. Computer vision techniques for gait recognition can be classified as model-free and model-based ones. In model-based approaches, a set of body parameters are obtained by fitting a body model to the person captured in each frame. In contrast, model-free approaches do not utilize a model for people but the entire

G. Maida · M. Morana (✉)
University of Palermo, Viale delle Scienze, ED. 6, 90128 Palermo, Italy
e-mail: gabriele.maida@unipa.it

M. Morana
e-mail: marco.morana@unipa.it

S. Gaglio and G. Lo Re (eds.), *Advances onto the Internet of Things*,
Advances in Intelligent Systems and Computing 260, DOI: 10.1007/978-3-319-03992-3_12,

shapes of silhouettes or the whole motion of human bodies are used [11, 22]. Model-based approaches are view-invariant and scale-independent, but usually require high quality gait sequences and more computing time than model-free approaches.

The use of floor sensors and wearable sensors can provide more accurate feature detection than computer vision approaches, but the former are expensive because of the use of force plates installed on the floor [15, 17], while the main disadvantage of the latter is that they are more intrusive.

Our approach falls into the computer vision category, in which a person's gait is captured by a camera. The proposed system aims to unobtrusively identify people in an Ambient Intelligence (AmI) scenario [7] by using an array of Kinect sensors to detect gait features, and then recognize the person by his walking style.

We started from the OpenNI 2.0 APIs [2] which provide an efficient skeleton detection method that makes it possible to represent a human body as a set of connected joints. By analyzing joint positions and their spatio-temporal evolutions, it is possible to extrapolate static and dynamic features. In order to preserve the pervasiveness of the system, the Kinect sensors are coherently connected to a miniature fanless computer with reduced computation capabilities.

This chapter is organized as follows: relevant research is presented in Sect. 2, while the system architecture proposed here is described in Sect. 3. The experimental scenario is discussed in Sect. 4, followed by our conclusions in Sect. 5.

2 Related Work

Over the years, the issue of gait recognition has been addressed in different ways in various works. The first attempt at automatic gait recognition was probably the one reported in [5], where a camera was used to capture light sources mounted on selected joints of walking people.

Gait recognition techniques based on silhouette analysis have been proposed in various works. Typically, most of them involve the following phases: subject detection, silhouette extraction, feature extraction, feature selection and classification [23]. In [25], a spatio-temporal silhouette analysis is performed to detect a sequence of static body poses. For each frame of the sequence, a background subtraction procedure is used to extract the binary silhouette. Then, a principal component analysis is applied to compute the predominant components of gait signatures, and a classification technique is employed to identify the person.

Lee [12] describes a gait appearance feature vector based on moments obtained from silhouettes. The whole body is segmented into regions, and for each region, an ellipse is fitted to the visible portion of the foreground object in order to extract a set of moment-based region features. The Mahalanobis distance between feature vectors is used as a measure of similarity, and a Nearest-Neighbor approach is used to rank a set of training sequences according to their distances, by means of a query sequence.

The authors of [11] proposed a spatio-temporal gait representation called Gait Energy Image (GEI). To preserve temporal information, the motion pattern within a gait cycle is represented in a single image. GEI needs less storage space and computation time, but it is still view-dependent and performs better when a side view is used. A possible way to solve the view dependency issue is to use multiple cameras with overlapping fields of view.

The authors of [21] extended the concept of GEI by introducing the concept of Gait Energy Volume (GEV), that uses averaged reconstructed voxel volumes instead of temporally averaging segmented silhouettes. Such 3D data is reconstructed using depth images captured by a frontal viewpoint of the Microsoft Kinect sensor.

In [20], an approach that utilizes both gait and face recognition is presented. The authors demonstrate that the integration of face and gait recognition provides better performance than the use of a single technique alone. Face recognition usually works better with front-parallel images, while a gait recognition technique, based on silhouette analysis, performs more efficiently on side-view sequences.

The authors of [18] use the Kinect SDK provided by Microsoft to obtain a 3D virtual skeleton. A single Kinect sensor is placed to give a side view of the walking path in order to acquire video sequences of walking people. In each frame, the skeleton is converted into a vector containing static features (i.e., the height of the subject and the length of certain body parts). Two dynamic features are also computed along the walk (i.e., step length, and speed). Such feature vectors are then classified using different classifiers.

In [14], a module for the management of an office environment is described, using Microsoft Kinect as the primary interface between the user and the AmI system. In particular, a fuzzy classifier is trained to recognize some simple gestures (such as open/closed hands) in order to produce a set of commands, opportunely structured by means of a grammar, which are used to control the actuators of the AmI system.

3 System Overview

In recent years, the availability of an ever-increasing number of cheap and unobtrusive sensing devices has piqued the interest of the scientific community into producing novel methods for understanding what is happening in the environment, based on the raw measures acquired. Due to the heterogeneity of the data captured, an information fusion mechanism [8] is usually required to address a specific goal, such as that of understanding what the user is doing.

In this chapter, we address the issue of gait recognition using an array of Kinect sensors to unobtrusively identify people in an Ambient Intelligence scenario.

In particular, the approach proposed here relies on the analysis of certain features extracted by means of the OpenNI 2.0 APIs [2], which provide a real-time representation of the human body as a set of connected joints. Moreover, multiple Kinects are used to increase the acquisition range provided by a single Kinect in order to include a longer walk (e.g., in a hallway before the user enters the office). By analyzing joint

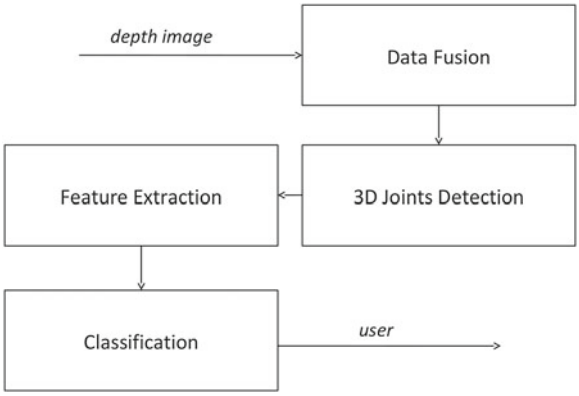

Fig. 1 System overview

positions and their spatio-temporal evolutions, it is possible to extrapolate a set of static and dynamic gait features. Finally, the classifiers are trained using the data collected and then they are used to recognize users' walks. The training set contains ten repetitions of a walk per user. If a walk is not recognized, it will be marked as "unknown".

Our system architecture (see Fig. 1) consists of four components: a *Kinect Data Fusion* step to merge data from multiple Kinects, a *3D Joints Detection* step to obtain human skeletons from depth maps, a *Feature Extraction* step to detect gait features, a *Classification* step to identify people according to the extracted features, and a *Calibration* step which is performed the first time the Kinects are placed into the environment.

3.1 Multi Kinect Architecture

As reported in [3], we found that Kinect is the most suitable device for pervasive AmI tasks, both in terms of cost and functionality, since it is equipped with ten input/output components which allow the device to perceive and interact with the surrounding environment.

The core of the Kinect is represented by the vision system, composed of an RGB camera with VGA standard resolution (i.e., 640 $\times$ 480 pixels), an IR-projector that shines a grid of infrared dots over the scene and an IR-camera for capturing the infrared light. The information obtained from projected dots is used to create three-dimensional depth maps of the observed scene (i.e., pixel values represent distances). The other components include four microphones (three on the right side and one on the left side), a led indicator that shows the state of the Kinect, a motor that allows you to control the tilt angle of the camera (30° upward or downward), and a 3-axis accelerometer that measures the position of the sensor.

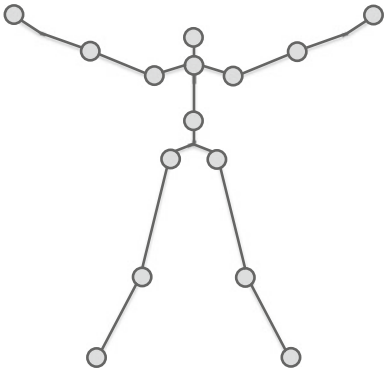

Fig. 2 The 15 joints of the human body: head, neck, torso, shoulders, elbows, hands, hips, knees, feet

A Kinect can see people standing between 0.5 and 3.5 m from the sensor, but we chose to extend this range by putting two Kinects at a distance of approximately 3 m from each other, in order to have about 0.5 m of overlapping frames acquired by both devices during a walk. In the *Calibration* step, an object is positioned at a point visible from both Kinect sensors, in order to find the roto-translation matrix representing the transformation between the two coordinate systems. Such conversion matrix is then used in the *Kinect Data Fusion* step.

Note that it is possible to add more devices by repeating the *Calibration* process, so that any Kinect addition will increase the captured walk sequence of about 3 m, thereby allowing for a better analysis of dynamic features (e.g., the walking speed). We verified that two Kinects appear to be adequate given the cyclical nature of certain features (e.g., gait cycle duration).

Kinect Data Fusion is the core of the Multi Kinect Architecture. During this step, each frame from a different device is ordered in the right position and converted into a single coordinate system, so that the entire walk can be considered as captured by a single virtual Kinect. This may be done by using the timestamp related to each frame and applying the roto-translation matrix for the conversion of the coordinate system. This step is essential, especially when the person is visible from both Kinects. Such new virtual gait frames are used in the *3D Joints Detection* step to detect the joint positions of the person and then in the *Feature Extraction* step to detect gait features.

3.2 Gait Features

In [16], a gait cycle is defined as the time interval between two successive floor contacts of the same foot. During a gait cycle, two distinct periods can be observed, namely a *swing phase*, when a leg is moving forward, and a *stance phase*, when both feet are touching the ground [26].

We chose to use both static body-shape parameters, that can not change during a walk, and dynamic features, which depend on the walk. The OpenNI 2.0 skeleton detection method is able to perform real-time detection (i.e., to find the 3D coordinates) of 15 body joints (see Fig. 2). Using these points and their evolution during a walk, we are able to extract static body-shape parameters and dynamic features from the acquired sequences.

The considered static body-shape parameters are:

- *person height*: vertical distance between head and feet;
- *torso length*: vertical distance between neck and torso joints;
- *shoulder width*: horizontal distance between shoulders;
- *arm length*: distance between shoulder and hand joints;
- *leg length*: vertical distance between hip and foot joints.

The considered dynamic features are listed below:

- *gait cycle duration*: time interval between successive floor contacts of a foot;
- *walk speed*: total walk time divided by total walk distance;
- *step width*: maximum distance between feet during the gait cycle;
- *footstep frequency*: total walk time divided by number of footsteps;
- *arm swing frequency*: total walk time divided by number of arm swings.

where *total walk time* is the difference between the first and the last frame timestamps acquired. The static body-shape parameters are extracted during the stance phase, whilst the dynamic features are extracted by analyzing the evolution of the joints during an acquired walking sequence. Once the whole walking sequence has been acquired, the features extracted are used to train a classifier in the *Classification* step.

3.3 Feature Classification

We used three different classifiers to evaluate the system's performance with different sets of selected features. In particular, the following classifiers were used: a Multi-Class Support Vector Machines, a K-Nearest Neighbor and a Naive Bayes Classifier.

Support Vector Machines (SVM) [19] are supervised learning models with associated non-probabilistic algorithms used to analyze data for binary classification and regression. A multi-class SVM is a net of SVMs able to classify instances into more than two classes.

The K-Nearest Neighbor decision rule [4] assigns to an unclassified observation the most common class amongst its k closest samples in a reference set (where k is a positive integer).

Naive Bayes [27] is a kind of probabilistic classifier based on Bayesian networks that assigns a new observation to the most probable class, assuming that the attributes are conditionally independent given the class value.

Table 1 The feature sets used for the experiments

Static features (SF)	Dynamic features (DF)	Mixed set 1 (MS1)	Mixed set 2 (MS2)
Person's height	Gait cycle duration	Person's height	Torso length
Torso length	Walk speed	Shoulder width	Arm length
Shoulder width	Step width	Leg length	Gait cycle duration
Arm length	Footstep frequency	Walk speed	Footstep frequency
Leg length	Arm swing frequency	Step width	Arm swing frequency

In order to find the most discriminating features we tested the system by using a single feature at a time and evaluating the recognition rate. The features that yielded the highest recognition rate in the test (about 25–30 %) were included in the *MS1* set, whereas the remaining features were included in the *MS2* set. The various feature sets that were tested for the evaluation of the system, reported in Table 1, include the *SF* set consisting of static features only, the *DF* set consisting of dynamic features only, and two mixed sets (i.e., *MS1* and *MS2*).

3.4 System Ontology

Ontology [10] is a formalism to share and re-use knowledge among different systems. It represents the conceptualization of the relevant entities and their relations in a common vocabulary.

Our system ontology is shown in Fig. 3. The subdivision among system modules, data types and devices is sketched. System modules are the components that manipulate data from devices. *DataType* is subdivided into *RawData*, that represents data which has not been elaborated, and *Symbolic*, representing data containing information.

There are two types of *SystemModules*, namely *TranslationModule* and *UnderstandingModule*. *TranslationModule* can be subdivided into three types called *DataFusionModule* that transforms raw data (*DepthMap*) in other raw data (new *DepthMap*), *JointExtractionModule* which transforms *DepthMap* in symbolic data (*Joint*) and *FeatureExtractionModule*, that obtains other symbolic data (*Feature*) from joints. The *UnderstandingModule* classifies symbolic features to detect the user who is performing a walk.

In our system two kinds of *Devices* are used, called, *Sensor* (Kinect), to acquire *RawData* from the environment, and *Node*, which is the device where the sensor is installed (e.g., Kinect needs a computer to work).

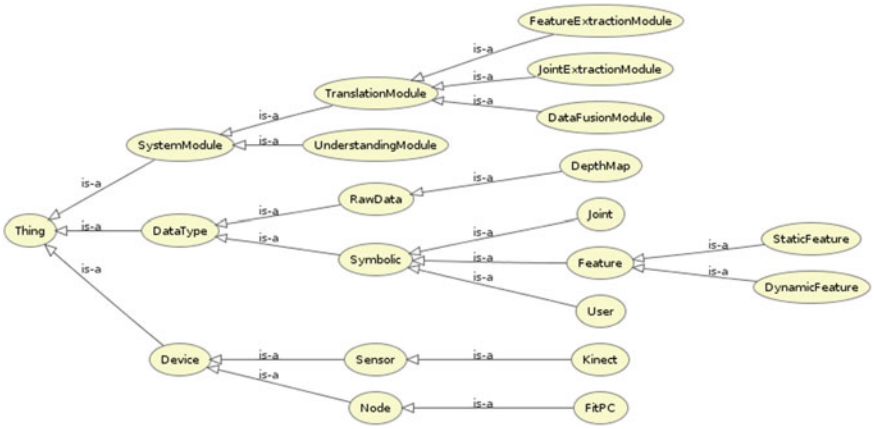

Fig. 3 System ontology

4 Experimental Results

The sensory component in our AmI architecture is implemented through a Wireless Sensor and Actuator Network (WSAN), whose nodes are equipped with off-the-shelf sensors (i.e., for outdoor temperature, relative humidity, ambient light exposure and noise level) [6].

The gait recognition module was evaluated by performing a number of tests on gait data collected at our Department.

In particular, two Kinects were placed on an office hallway at a distance of approximately 3 m from each other (see Fig. 4). Ten subjects (three women and seven men in the height range from 160 to 185 cm) were asked to walk at their normal speed down a path of approximately 8 m. The Kinect sensors captured the user's walks from a frontal-view with a frame rate of 30 fps. Each person repeated the walk ten times, so a total of 100 walking sequences were collected.

The main goal of our tests was to find the best feature set for our Gait Recognition System. Thus, we trained three classifiers with the sets described in Table 1 with all of features together. In order to evaluate the couple *classifier/feature set*, we computed the classification rate of the system by using the Leave-One-Out-Cross Validation method [1], then, for each test, 99 sequences were used for training and the remaining sequence was used for validation.

The accuracy values obtained are reported in Table 2. It is noticeable that static features (*SF*) are more relevant than the dynamic ones (*DF*), since they provided a better recognition rate for each classifier. However, increasing the number of features is not enough to improve the performance. In fact, the recognition rate decreases when all of the features were used. On the other hand, by selecting the most relevant features (i.e., the *MS1* set composed of *person's height*, *shoulder width*, *leg length*,

Fig. 4 Kinect sensors placed in the office hallway

Table 2 Accuracy obtained for each classifier and feature set

Classifier	All features (%)	SF (%)	DF (%)	MS1 (%)	MS2 (%)
SVM	54	79	38	73	47
KNN	68	79	40	**92**	58
Naive Bayes	83	75	50	86	66

walk speed and *step width*), we obtained the best results achieving a recognition rate of 92 % with the KNN classifier.

As far as classifiers are concerned, the Naive Bayes classifier was found to be more stable, yelding the highest overall classification rate for each feature set (from 50 % by using dynamic features only, to 86 % by using the *MS1* set), whereas the SVM classifier gave the worse results (from 38 % by using dynamic features only, to 79 % by using the static set). The KNN classifier provided good results with all sets, and in particular with the *MS1* set.

The overall system was tested using C language for the *Kinect Data Fusion* and *3D Joint Detection* steps and MATLAB for the *Feature Extraction* and *Classification* steps. A prototype of the activity recognition module was implemented connecting the Kinect to a miniature fanless PC (i.e., a fit-PC2i with Intel Atom Z530 1.6GHz CPU and Linux OS with kernel 2.6.32), that guarantees real-time processing of the observed scene with minimum levels of obtrusiveness and low power consumption.

5 Conclusion

The research presented analyzes a system to unobtrusively identify people by using multiple Kinect sensors.

We used an ontology to interface our gait recognition module with the entire ambient intelligence system. The use of the same concepts helped us to share information with other system modules (e.g., user presence). After an accurate analysis of the ontology, we chose to structure our system in four modules: *Kinect Data Fusion*, *3D Joints Detection*, *Feature Extraction* and *Classification*.

We collected our gait dataset by placing two Kinect sensors in a hallway in our Department and asking ten people to walk ten times.

By analyzing joint positions and their spatio-temporal evolutions, we have been able to detect several features, namely a *person's height*, *torso length*, *shoulder width*, *arm length*, *leg length*, *gait cycle duration*, *walk speed*, *step width*, *footstep frequency* and *arm swing frequency*. The features are then used to train a classifier to recognize the user performing the walk.

We carried out a number of experiments to evaluate the feature sets and classifiers by using the 100 walking sequences we captured. As a result of our tests we found that a subset of features composed by *person's height*, *shoulder width*, *leg length*, *walk speed* and *step width* was sufficient to correctly identify a person with a recognition rate of 92 %.

Acknowledgments This work has been partially supported by the PO FESR 2007/2013 grant G73F11000130004 funding the SmartBuildings project.

References

1. Arlot, S., Celisse, A.: A survey of cross-validation procedures for model selection. Stat. Surv. **4**, 40–79 (2010)
2. consortium, O.: OpenNI. http://www.openni.org/
3. Cottone, P., Lo Re, G., Maida, G., Morana, M.: Motion sensors for activity recognition in an ambient-intelligence scenario. In: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 646–651 (2013). doi:10.1109/PerComW.2013.6529573
4. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Trans. Inf. Theory **13**(1), 21–27 (1967)
5. Cutting, J.E., Kozlowski, L.T.: Recognizing friends by their walk: gait perception without familiarity cues. Bull. Psychon. Soc. **9**(5), 353–356 (1977)
6. De Paola, A., Gaglio, S., Lo Re, G., Ortolani, M.: Sensor9k : A testbed for designing and experimenting with wsn-based ambient intelligence applications. Pervasive Mob. Comput. **8**(3), 448–466 (2012). http://dx.doi.org/10.1016/j.pmcj.2011.02.006
7. De Paola, A., La Cascia, M., Lo Re, G., Morana, M., Ortolani, M.: User detection through multi-sensor fusion in an ami scenario. In: Information Fusion (FUSION), pp. 2502–2509 (2012)
8. De Paola, A., La Cascia, M., Lo Re, G., Morana, M., Ortolani, M.: Mimicking biological mechanisms for sensory information fusion. Biol. Inspired Cogn. Architectures **3**(0), 27–

38 (2013). doi:10.1016/j.bica.2012.09.002. http://www.sciencedirect.com/science/article/pii/S2212683X12000527
9. Gafurov, D.: A survey of biometric gait recognition: approaches, security and challenges. In: Annual Norwegian Computer Science Conference, pp. 19–21 (2007)
10. Gruber, T.R., et al.: A translation approach to portable ontology specifications. Knowl. Acquisition **5**(2), 199–220 (1993)
11. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2), 316–322 (2006)
12. Lee, L., Grimson, W.E.L.: Gait analysis for recognition and classification. In: Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 148–155 (2002)
13. Liu, Z., Sarkar, S.: Simplest representation yet for gait recognition: averaged silhouette. In: Proceedings of the Seventeenth International Conference on Pattern Recognition (ICPR'04), vol. 4, pp. 211–214 (2004)
14. Lo Re, G., Morana, M., Ortolani, M.: Improving user experience via motion sensors in an ambient intelligence scenario. In: Pervasive and Embedded Computing and Communication Systems (PECCS), pp. 29–34 (2013)
15. Middleton, L., Buss, A., Bazin, A., Nixon, M.: A floor sensor system for gait recognition. In: Fourth IEEE Workshop on Automatic Identification Advanced Technologies, pp. 171–176 (2005). doi:10.1109/AUTOID.2005.2
16. Murray, M.P., Drought, A.B., Kory, R.C.: Walking patterns of normal men. J. Bone Joint Surg. **46**(2), 335–360 (1964)
17. Orr, R.J., Abowd, G.D.: The smart floor: a mechanism for natural user identification and tracking. In: CHI'00 Extended Abstracts on Human Factors in Computing Systems, pp. 275–276. ACM Press (2000)
18. Preis, J., Kessel, M., Werner, M., Linnhoff-Popien, C.: Gait recognition with kinect. In: Proceedings of the First International Workshop on Kinect in Pervasive Computing (2012)
19. Scholkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge (2001)
20. Shakhnarovich, G., Lee, L., Darrell, T.: Integrated face and gait recognition from multiple views. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 436–439 (2001)
21. Sivapalan, S., Chen, D., Denman, S., Sridharan, S., Fookes, C.: Gait energy volumes and frontal gait recognition using depth images. In: IEEE International Joint Conference on Biometrics (IJCB), pp. 1–6 (2011)
22. Tao, D., Li, X., Wu, X., Maybank, S.J.: General tensor discriminant analysis and gabor features for gait recognition. IEEE Trans. Pattern Anal. Mach. Intell. **29**(10), 1700–1715 (2007)
23. Wang, J., She, M., Nahavandi, S., Kouzani, A.: A review of vision-based gait recognition methods for human identification. In: Digital Image Computing Techniques and Applications (DICTA), pp. 320–327 (2010)
24. Wang, L., Tan, T., Hu, W., Ning, H.: Automatic gait recognition based on statistical shape analysis. IEEE Trans. Image Proc. **12**(9), 1120–1131 (2003)
25. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis-based gait recognition for human identification. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12), 1505–1518 (2003)
26. Yamauchi, K., Bhanu, B., Saito, H.: Recognition of walking humans in 3d: initial results. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR), pp. 45–52 (2009)
27. Zhang, H.: The optimality of naive bayes. In: Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (2004)