The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence
Artificial Intelligence Applied to Assistive Technologies and Smart Environments:
Technical Report WS-16-01

Child-Centred Motion-Based Age and Gender Estimation with Neural Network Learning

Anara Sandygulova
Department of Robotics and Mechatronics
School of Science and Technology
Nazarbayev University
Astana, Kazakhstan

Yerdaulet Absattar, Damir Doszhan
Department of Computer Science
School of Science and Technology
Nazarbayev University
Astana, Kazakhstan

German I. Parisi
Knowledge Technology Group
Department of Informatics
University of Hamburg
Hamburg, Germany

Abstract

The focus of this work is to investigate how children's perception of the robot changes with age and gender, and to enable the robot to adapt to these differences for improving human-robot interaction (HRI). We propose a neural network-based learning architecture to estimate children's age and gender based on the body motion performing a set of actions. To evaluate our system, we collected a fully annotated depth dataset of 28 children (aged between 7 and 16 years old) and applied it to a learning-based method for age and gender estimation by modeling children's 3D skeleton motion data. We discuss our results that show an average accuracy of 95.2% and 90.3% for age and gender respectively in the context of a real-world scenario.

Introduction

Socially interactive technologies are no longer reserved for adults. In particular, research and commercial robots have infiltrated homes, hospitals and schools, becoming attractive and proving impactful for children's healthcare (Belpaeme et al. 2012), therapy (Dautenhahn et al. 2009), education (Kanda et al. 2004), and other applications.

In order to establish social and bonding relationships with children in public populated environments such as hospitals or educational institutions, robots need to be able to adapt to child's developmental differences, so that child-robot interaction (cHRI) is the most effective: robot is liked and accepted, provides comfort and companionship, perceived to be a friend or a peer.

However, social robots of public environments are challenged with interactions involving previously unseen users: people of different age and gender groups that have various preferences and needs toward the social robot. The approach presented in this paper is to create such a system which is able to dynamically adapt to its users by characterizing them using some of the same cues that people use. People tend to take cognizance of the age and gender of the interlocutor to tailor the content they communicate together with their behaviors, in terms of non-verbal social cues such as body language, gestures, and gaze. Moreover, prior work on cHRI has identified the significance of interactive robots supporting child-centered adaption: children's reaction to social robots varies with their age and gender (Scheeff et al. 2002) (Kanda et al. 2004), (Fink et al. 2014), and similarly, children demonstrate varying preferences for human-like robots, their age and gender (Tung 2011). In the work described in this paper, we propose a novel methodology for child-centered gender and age estimation based on the motion data as the basis for the adaptation of robot's social and verbal behaviors.

Related work (Sandygulova, Dragone, and O'Hare 2014) utilized a set of 3D body metrics to effectively and robustly estimate age and gender of children resulting in 73% correct success rate when estimating gender and mean absolute error was 0.94 years with a standard deviation of 1.27 years for children when determining children's age. And compared with a state-of-the-art software based on face analysis namely SHORE (Ernst, Ruf, and Kueblbeck 2009) developed mainly for adults, our novel methodology based on the motion of the 3D skeleton data outperforms previous work by achieving outstanding age and gender estimation results.

In summary, the contributions of this paper are three-fold:

- It provides a fully annotated depth dataset of 28 individuals: 16 boys and 12 girls aged between 7 and 16 years old;
- It addresses a little explored issue of child-centered adaptation and user profile building.
- It provides a novel method for age and gender estimation by modeling motion 3D skeleton data, to which we subsequently apply a neural network-based classification algorithm.

Section 2 introduces the related systems deployed in public environments focused on the child-robot interaction. Section 3 details the data collection procedure we carried out to validate our method. The methodology used for age and gender estimation is discussed in Section 4, followed by the experimental results in Section 5 and the conclusion in Section 6.

Related Work

An increasing number of systems is developed with the mission to enable the design, implementation, and evaluation of robots that encourage social, emotional, and cognitive growth in children, including those with social or cognitive deficits.

Copyright © 2016, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Human-Robot Interaction

The selected work includes large-scale projects that span across a number of research and development partners across countries that aim to combine expertise from different domains in order to form interdisciplinary teams.

Research efforts towards the development of Robot Assisted Therapy (RAT) systems have produced a promising outcome for the therapy of children with autism spectrum disorders. The minimally expressive robot KASPAR (*Kinesics and Synchronization in Personal Assistant Robotics*) (Dautenhahn et al. 2009) and the huggable robot Probo (Vanderborght et al. 2012) are derived from such efforts. The European project ALIZ-E (*Adaptive Strategies for Sustainable Long-term Social Interaction*) (Belpaeme et al. 2012) focused on children with diabetes, aided by the humanoid NAO robot that provided help by offering training and entertainment in real hospital settings. During the European project LIREC (*Living with Robots and Interactive Companions*) numerous studies were conducted in schools to address the challenges of maintaining children's interest in social robots and self-validation with the aim to improve child's learning during long-term interactions in school settings (Shahid et al. 2010). The aforementioned systems developed for public environments need to account for the age and gender differences of children and attitudes towards the robot. Therefore, this type of systems need to support dynamic adaptation and estimation of children's age and gender groups. The following background work motivates our research on the importance of creating a perception module for the dynamic adaptability to children's gender and developmental differences. This work investigates whether children's gender may impact the way children perceive and interact with the robot. Scheeff et al. (2002) found that children aged 4-7 years tended to be very energetic around Sparky and kind to it regardless of their gender. Older children (from 7 years old to early teens) behaved differently according to gender: boys of that age were usually aggressive and girls were generally gentle with the robot. The study by Tung (2011) examined whether gender or age influences the social and physical attraction children feel toward humanoid robots with the results suggesting that girls are more accepting of human-like robots, especially female robots, than boys are. Whether younger and older children could share a secret with the robot, Bethel, Stevenson, and Scassellati (2011) found that children aged 4 and 5 were as likely to share a secret with a NAO robot as with an adult in contrast to older children. A recent study by Ozogul et al. (2013) investigated the choice in animated pedagogical agents of middle-school learners (11-13 years old). The findings support the similarity attraction hypothesis with significant preference ($p < 0.001$) in children's choice for the computer-based animated agent that matched their same age and gender.

This literature review on cHRI demonstrates that there is a need for an effective and robust method to dynamically estimate age and gender of children in public real-world environments to be able to accommodate to children's developmental and gender differences.

Work Exploring Age or Gender through Motion

A new anatomically-based protocol was obtained for gait analysis in children in the research conducted by Leardini et al. (2007). The proposed protocol was based on the analysis of pelvis and lower limb motion that obtained as a compromise between two aforementioned requirements. The experiment involved the attachment of 22 skin markers, 6 anatomical landmarks calibration by a pointer and hip joint centre identification by the prediction approach. Ten healthy children with a mean age of 9.7 years old participated in the research. Each child was assessed several times by different examiners. Three main results were obtained: an intra-subject variability was very small, inter-examiner variability was moderately small, and joint rotations and moments (that had been calculated from each subject) were very close to the corresponding data obtained by similar anatomical definitions (in spite of using different marker sets in the research). The protocol allows 3D anatomical-based measurement of segment and joint motion and data sharing which potentially can resolve many issues related to limitations in clinical gait analysis. This model is particularly well-suited for children, but can be used for adults.

The research conducted by Ferrari et al. (2007) also shows the importance of data collection and reduction procedures in gait analysis to make kinematic and kinetic measurements more comprehensible in clinical usage. The five worldwide protocols were chosen to compare analysis of kinematics and kinetics of the trunk, pelvis and lower limbs. It resulted in overall 60 markers that set up on a skin or hands with 16 anatomical landmark calibrations performed. One patient with knee prosthesis implanted and two healthy subjects were analyzed by five experts. Results in the research showed very small variability for kinematic and kinetic results observed for each subject. For each protocol, there have been found a similarity with the high rates of intra-protocol repeatability. Moreover, a general uniformity was found in all three subjects among the five protocols. A good consistency was observed for all joint flexion/extensions, for pelvic rotations, hip-out-of-sagittal plane rotations. And an acceptable consistency was observed for all joint moments.

Another research conducted by Manca et al. (2010) shows the importance of reliability of kinematic measurements in gait analysis. The aim of the study was in assessment of the inter-trial, inter-session and inter-examiner variability of an anatomical-based protocol. The subjects of the research were two young adult volunteers. Four examiners with different degrees analyzed these subjects. The data from different walking trials were collected. Rotations in the three anatomical planes of the ankle, hip, knee and pelvis were calculated. The results were the following: the standard deviations for the inter-trial, inter-session, and inter-examiner variability were consistent. Joint rotations in the traverse were significantly larger than in other planes in all three forms of variability. Only the small differences were observed between the examiners.

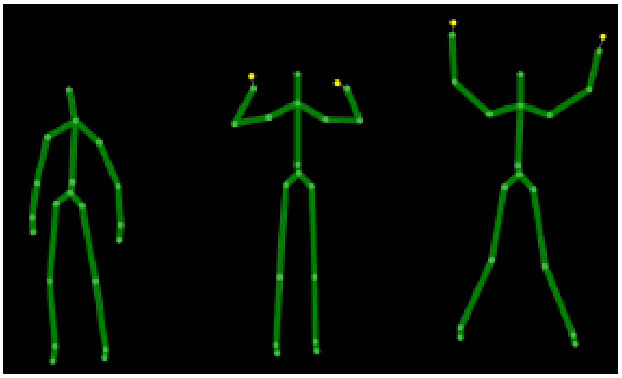

Figure 1: Example of skeleton model with body joints and limps for a correct jump movement.

Data Collection

The depth data was collected on a regular school day. Volunteers were brought to the specially allocated classroom to stand in front of the Kinect sensor in order to capture 3D body information. Upon arriving to the room, children were asked for their age and gender and had their height and weight measured. Each session involved one participant at a time.

Microsoft Kinect

A motion capture device such as Microsoft Kinect was used for human detection, tracking and for retrieving 3D body metrics that are particularly indicative of various demographics groups, i.e. age and gender. One of the core capabilities of the Kinect is the possibility to capture a depth image. An infrared (IR) emitter emits the light beams and the depth sensor reads the beams reflected back to the sensor. The reflected beams are converted into depth information measuring the distance between an object and the sensor, thus capturing the depth image. The Kinect for Windows SDK includes a number of useful features, which can be used to sense human users including skeletal and facial tracking, and voice and gesture recognition. Using the IR camera, the Kinect can detect up to six people in the field of view of the sensor. Of these, up to two people can be tracked in detail. The Kinect application can locate the joints of the tracked users in space and track their movements over time. Our implementation of the skeletal tracking enables to recognize people and save the x, y and z coordinates of every joint at each motion frame during the performance of the actions.

Experimental Setup

The Kinect was used to track and estimate 3D skeleton models from raw motion. The sensor was located in front of the subjects that had to walk and run from point A to point B five times, covering a distance of 3.20 meters 2. Weight and height of the subject were measured by scales and meters respectively. A stopwatch was used to identify the time for walking and running from point A to point B, in order to find gait velocity of the subject. The data were obtained by the method of defining coordinates of each movement of the subject, precisely the x, y, and z coordinates of each body joint.

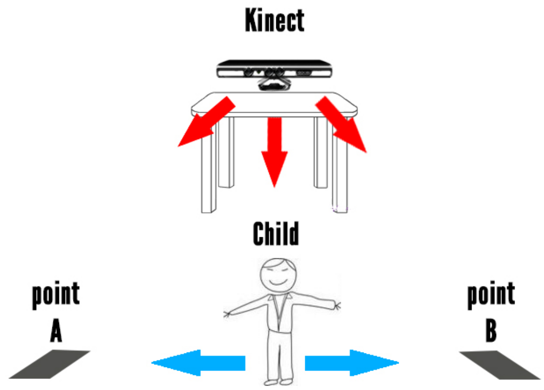

Figure 2: Experimental Setup

	1st group Females	1st group Males	2 group Females	2 group Males
Number of subjects	5	4	7	12
Age range (years)	8-10	8-10	12-15	11-16
Mean age (years)	8.80	9.00	13.57	13.91
Body mass range (kg)	20-33.6	31-39	38.5-60	25-77.5
Mean body mass (kg)	29.22	35.65	49.05	53.50
Body height range (cm)	129-142	131-144	149-167	137-189
Mean of body height (cm)	137.20	138.50	159.79	168.67
Gait velocity range (m/s)	0.89-1.24	0.97-1.41	0.68-1.09	0.88-1.26
Mean gait velocity (m/s)	1.09	1.26	0.92	1.08

Table 1: Participants' Information.

Participants

A 3D Body Model dataset was collected for 28 children, 16 boys and 12 girls aged between 7 and 16 years old. Children were divided into two groups by age category. The first group was in the age range of 7-10 years and the second group is in the age range of 11-16 years. Healthy subjects and without physical disabilities participated in the both groups. All subjects performed the same four actions: movement of the arms, walking, running and jumping. Each subject repeated each action 5 times. Moreover, parameters such as weight, height and gait velocity were measured for each child (Table 1).

First group included 5 females and 4 males. According to the measurements, the gait velocities of subjects were significantly different. It depended on their physical parameters and gender. Second group had 7 females and 13 males. The gait velocities of this group were also significantly different according to their physical parameters and gender.

Learning Architecture

Our learning architecture consists of 2 hierarchically arranged self-organizing neural networks (Figure 3). The use of hierarchical self-organization has been shown to be an efficient and effective method for recognizing human motion (Parisi, Weber, and Wermter 2015). This method is consistent with neurophysiological findings that have identified a specialized area for the visual processing of complex motion in the brain in a hierarchical fashion (Rolls and Caan 1982). More specifically, the visual system is composed of topographically arranged structures that organize according to external visual stimuli (von der Malsburg 1973). Input-driven self-organization has been shown to govern the development the connections in the visual cortex according to the distribution of the inputs. From a computational perspective, self-organization is an unsupervised learning mechanism that allows to learn representations of the input by iteratively obtaining a non-linear projection of the feature space (Kohonen 1989). Furthermore, it has been found that learning plays a crucial role in complex motion discrimination. Numerous studies have shown that the recognition speed and accuracy of humans have improved after a number of training sessions (Jastorff, Kourtzi, and Giese 2006).

Hierarchical Self-Organizing Learning

Our learning model consists of Growing When Required (GWR) networks (Marsland, Shapiro, and Nehmzow 2002) that iteratively obtain generalized representations of sensory inputs and learn inherent spatio-temporal dependencies. The GWR network is composed of a set of neurons and their associated weight vectors $\mathbf{w}_j$ linked by a set of edges. The activity of a neuron is computed as a function of the distance (usually the Euclidean distance) between the input and its weight vector. During the training, the network dynamically changes its topological structure to better match the input space following competitive Hebbian learning (Martinetz 1993). Different from other models of incremental self-organization, GWR-based learning takes into account the number of times that a neuron has fired so that neurons that have fired frequently are trained less. For this purpose, the network implements a habituation counter to express how frequently a neuron has fired based on a simplified model of how the efficacy of an habituating synapse reduces over time. This mechanism allows to create new neurons whenever it is required. The GWR algorithm will then iterate over the training set until a given stop criterion is met, in our case a maximum number of iterations. The standard procedure for GWR learning is described by Algorithm 1 (except for Steps 6.c and 7.c that are discussed in the following Section). For GWR learning, we used the following training parameters: insertion threshold $a_T = 0.70$, learning rates $\epsilon_b = 0.3$, and $\epsilon_n = 0.006$, $\kappa = 0.5$, maximum age $a_{max} = 50$, firing counter parameters $\eta_0 = 1$, $\tau_b = 0.3$, $\tau_n = 0.1$, firing threshold $\eta_T = 0.01$.

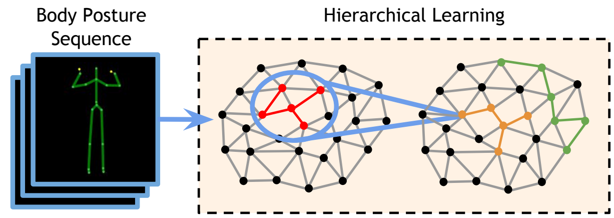

Figure 3: Hierarchical architecture with self-organizing neural learning (GWR networks). Learning is carried out by training a higher-level network with neuron activation trajectories from a lower level network trained on body posture sequences .

The motivation for using hierarchical learning is to use trajectories of neuron activations from one network as input for the training for a subsequent network. This mechanism allows to obtain specialized neurons coding spatio-temporal dependencies of the input, consistent with the assumption that the recognition must be selective for temporal order. Hierarchical learning is carried out by training a higher-level network with neuron activation trajectories from a lower level network. These trajectories are obtained by computing the best-matching neuron of the input sequence with respect to the trained network with N neurons, so that a set of trajectories of length q is given by

$$\Omega^q(\mathbf{x}_i) = \{\mathbf{w}_{b(\mathbf{x}_i)}, \mathbf{w}_{b(\mathbf{x}_{i-1})}, ..., \mathbf{w}_{b(\mathbf{x}_{i-q+1})}\} \quad (1)$$

with $b(\mathbf{x}_i) = \arg\min_{i \in N} \|\mathbf{x}_i - \mathbf{w}_j\|$ and q being the size of the temporal window.

We trained the low-level network of the hierarchy with vectors containing the 3D body information. To attenuate the effects of sensor noise, we estimated the median value for each joint every 3 vectors, i.e. resulting in 10 frames per second (instead of 30). The subsequent network was trained with activation trajectories of five ($q = 5$) neurons from the previous network using a sliding window scheme. Each network was trained for 300 epochs. This maximum number of epochs was empirically found based on the learning convergence of both networks and the final classification performance. After the training phase is completed, each high-level neuron will encode a sequence-selective action segment of 5 consecutive posture frames, i.e. half a second of motion captured at 10 frames per second.

Classification

At recognition time, our goal is to process and classify novel action sequences in terms of age and gender. For this purpose, we extended the unsupervised GWR-based learning of the higher level network to attach labels to trained neurons (Algorithm 1, steps 6.c and 7.c). In this case, the network will be trained with the motion sequences in an unsupervised fashion while using a labeling function to attach the labels of the input $\lambda(\mathbf{x}_t)$, i.e. age and gender, to best-matching neurons during the training phase. As a result of this process, each neuron in the higher level network encoding a motion segment will be associated to an input label. Different from

Algorithm 1 GWR Learning

1: Create two random neurons with weights $\mathbf{w}_1$ and $\mathbf{w}_2$
2: Initialize an empty set of connections $E = \emptyset$.
3: At each iteration t, generate an input sample $\mathbf{x}_t$
4: For each neuron n, select the-best matching node and the second-best such that:
$b_t = \arg\min_{n \in A} \|\mathbf{x}_t - \mathbf{w}_n\|$
$s_t = \arg\min_{n \in A/\{b_t\}} \|\mathbf{x}_t - \mathbf{w}_n\|$
5: Create a connection if it does not exist
5a: $E = E \cup \{(b_t, s_t)\}$ and set age of E_{b_t,s_t} to 0.
6: If $(exp(-\|\mathbf{x}_t - \mathbf{w}_{b_t}\|) < a_T)$ and $(\eta(b_t) < f_T)$ then:
6a: Add a new neuron r_t between b_t and s_t with $\mathbf{w}_{r_t} = \kappa \cdot (\mathbf{w}_{s_t} + \mathbf{x}_t)$
6b: Create edges and remove old edge: $E = E \cup \{(r_t, b_t), (r_t, s_t)\}$ and $E = E/\{(b_t, s_t)\}$
6c: Initialize label: $\lambda(r_t) = \lambda(\mathbf{x}_t)$
7: Else, i.e. no new neuron is added, update $\mathbf{w}_{b_t}$ and its neighbours i:
7a: $\Delta\mathbf{w}_{b_t} = \epsilon_b \cdot \eta(b_t) \cdot (\mathbf{x}_t - \mathbf{w}_{b_t})$ and $\Delta\mathbf{w}_i = \epsilon_n \cdot \eta(i) \cdot (\mathbf{x}_t - \mathbf{w}_i)$,
with $0 < \epsilon_n < \epsilon_b < 1$
7c: Update label: $\lambda(b_t) = \lambda(\mathbf{x}_t)$
7d: Increment the age of all edges connected to b_t.
8: Reduce the firing counters according.
9: Remove all edges with ages larger than a_{max} and remove neurons without edges.
10: If the stop criterion is not met, go to step 3.

previous approaches using GWR-based associative learning (Parisi, Weber, and Wermter 2015), in our approach each label consists of two values, age and gender, so that new samples can be processed through the hierarchy and return the label values of the best-matching sequence.

Experimental Results

The classification accuracy for each action is shown in Fig. 4. Our system achieved an average classification accuracy of 95.2% for age and for 90.3% gender estimation using 3-fold cross-validation on the training dataset. In the case of age, the standard deviation is of 2.5 years. These results suggest that although the exact age is harder to estimate than the gender, we successfully estimate the age range with high accuracy.

As reported in previous experiments, 3D body data extracted from depth information with a Kinect generally contains noisy samples that may have a negative influence on neural network learning (Parisi, Weber, and Wermter 2015). On the other hand, although the accuracy of Kinect technology is not so precise, this approach is computationally efficient and allows to extract 3D body information in real time, thereby enabling us to estimate age and gender with very low latency in a live scenario. This is in fact a very desirable property, since delays in HRI systems may have a strong negative impact in terms of the user experience and acceptability.

The learning-based approach with a self-organizing hierarchy has been shown to attenuate the negative effect of noisy samples, since for each novel observation the system will find a best-matching neuron in the trained low-level network that represents the input. This low-level network serves also as a dictionary of primitives that can be reused when learning sequences of posture samples, thereby reducing the total number of neurons required to represent all the actions. On the other hand, a limitation of this approach is that a best-matching neuron will always be found for each novel sample, even if the sample is pure sensor noise. A solution to this issue may be to introduce an embedded mechanism in the low-level network to filter out observations that are likely to be noise, e.g. values highly detached from dominating point clouds in the feature space.

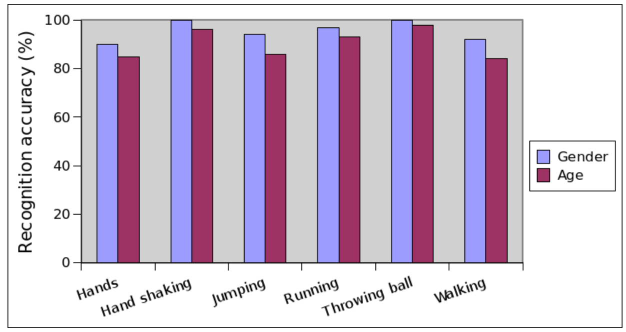

Figure 4: Experimental results: Classification accuracy for age and gender for each action.

To be pointed out is that these results indicate that neural network classification with hierarchical self-organization is an effective and efficient approach to process and learn from body attributes and action sequences. Since our experiments were conducted on a dataset of 28 subjects and 6 actions, we are unable to establish whether the recognition accuracy would be as high as our results for a higher number of children participants and a different set of actions.

Conclusion

In this work, we proposed a novel method to estimate age and gender of children based on 3D body motion information. The contribution of this work consists in a fully annotated depth dataset of 28 individuals and a learning-based method for age and gender estimation by modeling the children's motion based on 3D skeleton data. For this purpose, we extracted relevant metrics from body gait. Our reported results show that this methodology based on the learning of body metrics outperforms previous approaches on age-gender estimation in perceptually challenging environments. While this work focuses on adapting the applications of a humanoid robot to suit the preferences of children, this methodology could be applied to the design of any adaptive system to be tailored for the users.

References

Belpaeme, T.; Baxter, P. E.; Read, R.; Wood, R.; Cuayáhuitl, H.; Kiefer, B.; Racioppa, S.; Kruijff-Korbayová, I.; Athanasopoulos, G.; Enescu, V.; Looije, R.; Neerincx, M.; Demiris, Y.; Ros-Espinoza, R.; Beck, A.; Cañamero, L.; Hiolle, A.; Lewis, M.; Baroni, I.; Nalin, M.; Cosi, P.; Paci, G.; Tesser, F.; Sommavilla, G.; and Humbert, R. 2012. Multimodal Child-Robot Interaction: Building Social Bonds.

Bethel, C. L.; Stevenson, M. R.; and Scassellati, B. 2011. Secret-sharing: Interactions between a child, robot, and adult. In *Systems, man, and cybernetics (SMC), 2011 IEEE International Conference on*, 2489–2494. IEEE.

Dautenhahn, K.; Nehaniv, C. L.; Walters, M. L.; Robins, B.; Kose-Bagci, H.; Mirza, N. A.; and Blow, M. 2009. Kaspar–a minimally expressive humanoid robot for human–robot interaction research. *Applied Bionics and Biomechanics* 6(3-4):369–397.

Ernst, A.; Ruf, T.; and Kueblbeck, C. 2009. A modular framework to detect and analyze faces for audience measurement systems. In *2nd Workshop on Pervasive Advertising at Informatik*, 75–87.

Ferrari, A.; Benedetti, M. G.; Pavan, E.; Frigo, C.; Bettinelli, D.; Rabuffetti, M.; Crenna, P.; and Leardini, A. 2008. Quantitative comparison of five current protocols in gait analysis. *Gait & Posture* 28(2):207 – 216.

Fink, J.; Lemaignan, S.; Dillenbourg, P.; Rétornaz, P.; Vaussard, F.; Berthoud, A.; Mondada, F.; Wille, F.; and Franinović, K. 2014. Which robot behavior can motivate children to tidy up their toys? In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14*, 439–446. New York, New York, USA: ACM Press.

Jastorff, J.; Kourtzi, Z.; and Giese, M. A. 2006. Learning to discriminate complex movements: Biological versus artificial trajectories. *Journal of Vision* 6(8):3.

Kanda, T.; Hirano, T.; Eaton, D.; and Ishiguro, H. 2004. Interactive robots as social partners and peer tutors for children: a field trial. *Hum.-Comput. Interact.* 19(1):61–84.

Kohonen, T. 1989. *Self-organization and Associative Memory: 3rd Edition*. New York, NY, USA: Springer-Verlag New York, Inc.

Leardini, A.; Sawacha, Z.; Paolini, G.; Ingrosso, S.; Nativo, R.; and Benedetti, M. G. 2007. A new anatomically based protocol for gait analysis in children. *Gait & Posture* 26(4):560 – 571.

Manca, M.; Leardini, A.; Cavazza, S.; Ferraresi, G.; Marchi, P.; Zanaga, E.; and Benedetti, M. G. 2010. Repeatability of a new protocol for gait analysis in adult subjects. *Gait & Posture* 32(2):282 – 284.

Marsland, S.; Shapiro, J.; and Nehmzow, U. 2002. A self-organising network that grows when required. *Neural Networks* 15(8-9):1041–1058.

Martinetz, T. M. 1993. Competitive hebbian learning rule forms perfectly topology preserving maps. In *ICANN'93: International Conference on Artificial Neural Networks*, 427–434. Amsterdam: Springer.

Ozogul, G.; Johnson, A. M.; Atkinson, R. K.; and Reisslein, M. 2013. Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions. *Computers & Education* 67(0):36 – 50.

Parisi, G. I.; Weber, C.; and Wermter, S. 2015. Self-organizing neural integration of pose-motion features for human action recognition. *Frontiers in Neurorobotics* 9(3).

Rolls, E. T., and Caan, W. 1982. Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research* 47:329–342.

Sandygulova, A.; Dragone, M.; and O'Hare, G. 2014. Real-time adaptive child-robot interaction: Age and gender determination of children based on 3d body metrics. In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, 826–831.

Scheeff, M.; Pinto, J.; Rahardja, K.; Snibbe, S.; and Tow, R. 2002. Experiences with sparky, a social robot. In *Socially Intelligent Agents*. Springer. 173–180.

Shahid, S.; Krahmer, E.; Swerts, M.; and Mubin, O. 2010. Child-robot interaction during collaborative game play: effects of age and gender on emotion and experience. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, OZCHI '10, 332–335. New York, NY, USA: ACM.

Tung, F.-W. 2011. Influence of gender and age on the attitudes of children towards humanoid robots. In Jacko, J., ed., *Human-Computer Interaction. Users and Applications*, volume 6764 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 637–646.

Vanderborght, B.; Simut, R.; Saldien, J.; Pop, C.; Rusu, A. S.; Pintea, S.; Lefeber, D.; and David, D. O. 2012. Using the social robot probo as a social story telling agent for children with asd. *Interaction Studies* 13(3):348 – 372.

von der Malsburg, C. 1973. Self-organization of orientation-sensitive cells in the striate cortex. *Kybernetik* 15:85–100.