Human Skeleton Detection, Modeling and Gesture Imitation Learning for a Social Purpose

Linda Nanan Vallée, Sao Mai Nguyen, Christophe Lohr, Ioannis Kanellos, Olivier Asseu

Vallée, L.N., Nguyen, S.M., Lohr, C., Kanellos, I. and Asseu, O. (2020) Human Skeleton Detection, Modeling and Gesture Imitation Learning for a Social Purpose. Engineering, 12, 90-98. https://doi.org/10.4236/eng.2020.122009

HUMAN SKELETON DETECTION, MODELING AND GESTURE IMITATION LEARNING FOR A SOCIAL PURPOSE

Linda Nanan VALLÉE[1*], Sao Mai NGUYEN[2], Christophe LOHR[2], Ioannis KANELLOS[2], Olivier ASSEU[1]

[1]Ecole Supérieure Africaine des TIC, Abidjan, Côte d'Ivoire

[2]Institut Mines Telecom Atlantique de Bretagne, Lab-STICC, France

*linda.vallee@esatic.edu.ci

Abstract

Gesture recognition is topical in computer science and aims at interpreting human gestures via mathematical algorithms. Among the numerous applications are physical rehabilitation and imitation games. In this work, we suggest performing human gesture recognition within the context of a serious imitation game, which would aim at improving social interactions with teenagers with autism spectrum disorders. We use an artificial intelligence algorithm to detect the skeleton of the participant, then model the human pose space and describe an imitation learning method using a Gaussian Mixture Model in the Riemannian manifold.

Keywords: Imitation learning, Artificial intelligence, Gesture recognition, Autism spectrum disorders (ASD), Gaussian Mixture Model (GMM).

I. INTRODUCTION

Autism spectrum disorders (ASD) are linked with brain development [1]. Main symptoms of ASD are difficulties with communication and social interactions, repetitive behaviors and obsessive interests.

Autistic people also have talent [2]. For instance, an exceptionally good memory, a great attention to detail, an excellent ability to respect schedules, an exceptional level of honesty. Some of them are savants [3].

According to the World Health Organization [4], the global autism prevalence is around 1 in 160 children and autism is generally more common in boys than girls. Some autism prevalence studies were made per geographical area [5].

Still today, the exact cause of ASD is not known [6].

Several scientific studies have targeted improving the way autistic children communicate or interact with others [7]. This is because these two functions are crucial. Other symptoms of autism can be seen mostly as consequences of impairments in social interactions or communication.

Additionally, the imitation process is known to be a pillar in learning, communication and social interactions. Imitation games can therefore prove useful in helping autistic people interact with others.

In the present work, the structure of a gesture imitation game is proposed, which shall improve social interactions with autistic teenagers and preteens. Furthermore, skeleton detection and imitation learning methods are described.

The following section of this paper presents existing work related to imitation learning as well as improvement of social interactions with autistic children.

Section III then describes the methodology: the main phases of the imitation game as well as the skeleton detection and human motion learning methods.

Simulation results for skeleton detection are presented in section IV.

Section V finally concludes this paper and suggests future work.

II. RELATED WORK

II.1. On imitation learning and gesture recognition

In computer science, imitation learning, also called programming by demonstration, is a technique for teaching a computer or a robot to perform new tasks, through generalization from observing multiple demonstrations [8].

Within the framework of gesture recognition, a gesture would be performed several times by a human being and then, a method used for the system to be able to later recognize the task.

Different spatial gesture models exist **(figure 1).** Some are 3D-model based and others are appearance-based. Image sequences and deformable 2D templates are part of the latter group. The former comprises of skeletal and volumetric models.

Subtypes of the volumetric model category are NURBS, primitives and super-quadrics.

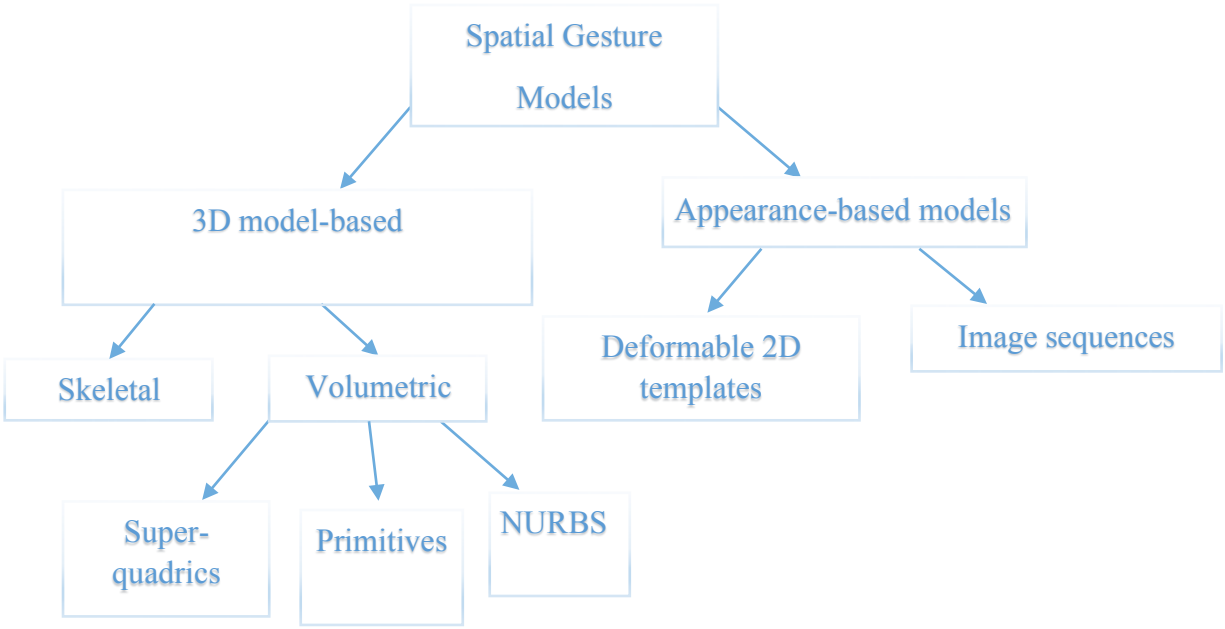


Figure 1: Spatial gesture model types

Human motion features can be limited to joint angles or extended to joint positions.

Since human movements are nonlinear, the Euclidian space is not really suitable to represent those. Human postures and motion are therefore often represented in alternative spaces such as the Riemannian one [10], which has proved useful as shown in [11].

Once the human body is modeled, gestures must actually be learned by the system by observation of several demonstrations. Probabilistic methods serve this purpose and can for instance be based on Hidden Markov Models or Gaussian mixture models.

A Gaussian mixture model (GMM) is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters.

The learned GMM model then represents a probabilistic description of the target (ideal) movement against which imitation attempts are compared.
The skeletal model, the representation of the human body in the Riemannian space as well as GMM have been used within the framework of physical rehabilitation exercises [12] [13].
In the context of our imitation game for teenagers with ASD, skeleton detection and gesture recognition methods are redirected to serve social purposes, as the aim is to improve the participants' ability to interact with others.

II.2. On imitation and autism

In [7] the author indicates that autistic children are able to imitate, whereas the general opinion previously differed.
The imitation process is complex and consists of subcomponents: induced imitation, spontaneous imitation, recognition of being imitated. This process is fundamental for learning, communicating and interacting socially.
In [8] an experimental work on imitation practice is presented in order to improve imitation abilities and reduce the autism level of 21 autistic children aged 4 to 10.
Nadel's imitation scale is used to evaluate the level of the three subcomponents of the imitation process.
These two studies are interesting on the psychological level because they show how imitation can positively impact autistic children, but they do not use modern techniques for imitation learning. The experimentation consists of human caregivers performing simple imitation games with the children.

In [14] Bernardini describes a multi-site intervention where 46 children with ASD aged 5 to 14 improved their social interactions through playing games, among which imitation games, with an intelligent agent called Andy. In most cases, the probability of the child answering Andy's requests increased. However social interactions initiated by the child were not really impacted.

In the work [15], the authors perform a review of studies using technological tools with autistic children and show that very few:

- aim at therapeutic effectiveness as well as technology usability;
- focus on teenagers;
- have a robust methodology.

III. METHODOLOGY

At the time of using artificial intelligence algorithms within our gesture imitation game for autistic teenagers, it is important to develop a robust methodology. Furthermore, the ease of use of the technology must be guaranteed while trying to improve the participants' initial condition in terms of social abilities.

Our gesture imitation game consists of two preliminary, three core and one final stages.

The two preliminary ones are the greetings and pairing stages.

Then come the three imitation modules: one based on induced imitation, another on spontaneous imitation and the third one on the recognition of being imitated.

This proposed structure follows multiple discussions with autism professionals.

In this paper we focus on potential methods for three core processes that will be useful throughout the game: skeleton detection, body representation, and finally recognition of previously learned gestures.

At game initiation, the skeleton of the participant is detected through the computer camera. This is done using the Openpose algorithm with the open source Tensorflow library, which is used to develop Machine learning and Deep learning algorithms. Tensorflow allows for solving of high complexity mathematical issues using experimental learning architectures. It is similar to a programming system in which computations are represented by graphs where nodes are mathematical operations and arrow borders are interconnected multidimensional

data called tensors. Tensorflow application programming interface (API) is Python-based but high-performance C++ is used for the execution of the applications.

Tensorflow is used to train and execute neural networks for element classification like in gesture recognition.

As for body representation, a human pose y at time t is represented by the orientation and position of all of the considered joints. The number of joints here is N. Therefore:

$$y_t = [O_1, P_1, O_2, P_2, \ldots, O_N, P_N] \quad (1)$$

where O_N are joint orientations.

Joint positions P_N are not absolute but normalized relative positions. They are computed from their absolute positions p_n relatively to the absolute position p_{ss} of the spine shoulder. Their normalization is done using the spine bone length L_{spine}:

$$P_N = (p_n - p_{ss})/L_{spine} \quad (2)$$

Unlike joint positions, joint orientations cannot be viewed in the Euclidian space but they can be represented in a 3D Riemannian manifold.

Therefore the human pose space is modeled as the Cartesian product of position and orientation of all of the human joints:

$$\mathcal{H} = \mathbb{R}^3 \times S^3 \times \mathbb{R}^3 \times S^3 \times \ldots \times \mathbb{R}^3 \times S^3 \quad (3)$$

In such space, among the various available methods, the one that was chosen was the Gaussian Mixture Models in Riemannian manifolds, as explained in [13].

Since the Riemannian space is nonlinear, tangent spaces at reference points are considered in order to be able to compute standard statistics, like mean and covariance.

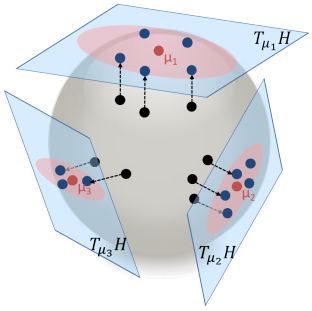


Figure 2: Illustration of the human pose space with three Gaussians computed on tangent space at means μ_k [13]

Paper [17] allows for the calculation of the mean μ of N points p_i on the human pose space:

$$\mu = \arg\min_p \sum_{i=1}^{N} d(\mu, p_i)^2 \qquad (4)$$

where d(μ, p) is the geodesic distance on the manifold which can be written using logarithmic map as d(μ, p) = ‖ $Log_\mu(p)$ ‖.

μ is also called the Riemannian center of mass.

The covariance matrix can then be computed, allowing for the learning of a Gaussian Mixture Model:

$$p(x) = \sum_{k=1}^{K} \phi_k N(x \mid \mu_k, \Sigma_k) \qquad (5)$$

where x encodes both the human pose y_t and the timestamps t, K is the number of Gaussians, ϕ_k the weight of the k-th Gaussian, μ_k the Riemannian center of mass of the k-th Gaussian computed on the manifold and Σ_k the covariance matrix of the k-th Gaussian. The parameters ϕ_k, μ_k and Σ_k are learned using Expectation-Maximization on the human pose space [18].

IV. EXPERIMENTAL RESULTS

For human pose detection, Tensorflow Pose Estimator was implemented. Below are the key command lines executed:

```
$ git clone https://www.github.com/ildoonet/tf-pose-estimation
$ cd tf-pose-estimation
$ pip3 install -r requirements.txt
$ cd tf_pose/pafprocess
$ swig -python -c++ pafprocess.i && python3 setup.py build_ext --inplace
$ python3 run_webcam.py --model=mobilenet_thin --resize=432x368 --camera=0
```

The *git clone* command copied the Tensorflow Pose Estimator file hierarchy into our work environment.

We then moved to the main directory and installed all of the required modules:

argparse
dill

fire
matplotlib
numba
psutil
pycocotools
requests
scikit-image
scipy
slidingwindow
tqdm
git+https://github.com/ppwwyyxx/tensorpack.git

Then from the *tf_pose/pafprocess* directory, we executed the *swig* command with the appropriate arguments and launched the setup file.

The *swig* command connects the C++ programs with the Python ones.

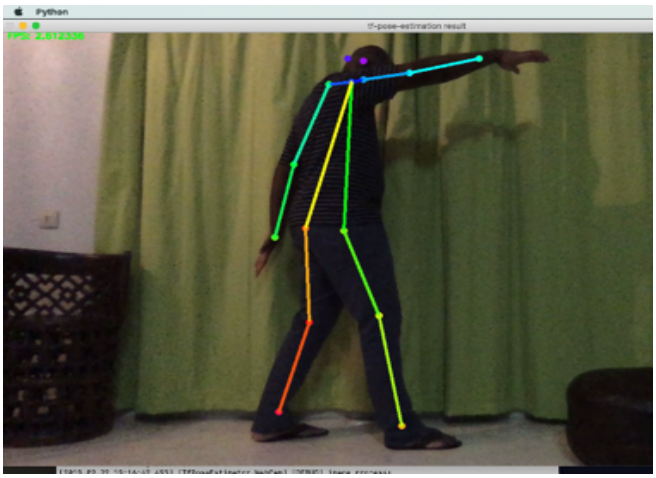
As mentioned earlier, Tensorflow API is Python-based but for the execution of the applications, high-performance C++ is used.

We finally executed *run_webcam.py* with the model and format of our choice, and obtained the following results captured through the webcam **(figure 3)**.

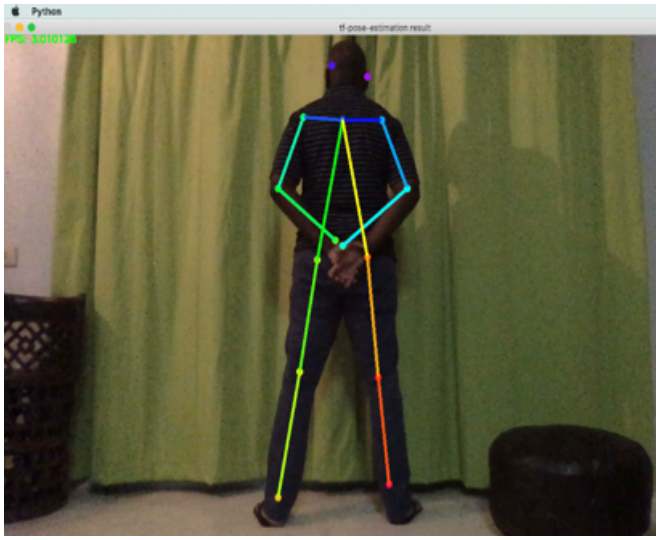
The body joints of the participant are represented by dots and connected through lines. Different colors are used to distinguish the different body parts.

Elements represented by the dots are the ankles, knees, hips, wrists, elbows, shoulders and ears.

On the first capture (fig.3a), we can see our participant from side-on, with the left arm behind him and the right arm raised in front of him, at head level.



(a)



(b)

Figure 3: Experimental results of the implementation of the Openpose algorithm

On the second capture (fig.3b), the participant appears from behind. His arms are crossed in his back and hands joined.

In spite of body occlusion (some body parts are superimposed), the body parts are correctly detected.

On a real-life background and even with low light, skeleton detection is functional using the Openpose algorithm.

This will be performed at the beginning of our gesture imitation game and useful throughout the different stages.

Each pose is represented by the position and orientation of the (fourteen) previously listed joints, as shown in formula (1). The joint positions are calculated relatively to the spine as seen in formula (2).

The human pose space is represented in the Riemannian manifold as expressed in formula (3) presented in the methodology section. Equations (4) and (5) are then used for gesture learning and recognition.

V. CONCLUSION

In this paper, the structure of a gesture imitation game was suggested and technical methods for skeleton detection and representation, as well as gesture recognition, were described.

Initial results for skeleton detection were then presented.

The full serious game will be developed in order to practice induced and spontaneous imitation with autistic teenagers and assess the improvement of their imitation and social abilities.

Later on, it would be interesting to transform the initially hard-coded game into a smart adaptive tutoring system and to use a social robot as a game partner. This could increase participants' interest and involvement.

ACKNOWLEDGEMENT

Here is the opportunity to express deep gratitude to Dr Lavenne-Collot Nathalie, child psychiatrist at the university hospital of Brest, France, as well as to Mr. Shaw Joël, caseworker with autistic children in Abidjan, Cote d'Ivoire, for their availability and interest in this work.

REFERENCES

. [1] The National Autistic Society (2014) What Is Autism? http://www.autism.org.uk/about-autism

. [2] Baron-Cohen, S., Ashwin, E., Ashwin, C., Tavassoli, T. and Chakrabarti, B. (2009) Talent in Autism: Hyper-Systemizing, Hyper-Attention to Detail and Sensory Hypersensitivity. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364, 1377-1383. https://doi.org/10.1098/rstb.2008.0337

. [3] Hughes, J.E.A., Ward, J., Gruffydd, E., et al. (2018) Savant Syndrome Has a Distinct Psychological Profile in Autism. Molecular Autism, 9, 53. https://doi.org/10.1186/s13229-018-0237-1

. [4] WHO, World Health Organization (2013) Autism Spectrum Disorders & Other Developmental Disorders. From Raising Awareness to Building Capacity. Geneva, Switzerland.

. [5] Franz, L., Chambers, N., von Isenburg, M. and de Vries, P.J. (2017) Autism Spec- trum Disorder in Sub-Saharan Africa: A Comprehensive Scoping Review. Autism Research, 10, 723-749. https://doi.org/10.1002/aur.1766

. [6] Shaw, C.A., Sheth, S., Li, D. and Tomljenovic, L. (2014) Etiology of Autism Spectrum Disorders: Genes, Environment, or Both? OA Autism, 2, 11.

. [7] Nadel, J. (2005) L'autisme, chapter Imitation et austisme.

. [8] Bendiouis, S. (2015) Imitation et communication chez le jeune enfant avec autisme. Psychologie. Université Paul Valéry-Montpellier III, Français.

. [9] Attia, A. and Dayan, S. (2018) Global Overview of Imitation Learning. ArXiv, abs/1801.06503.

. [10] J. Jost, Riemannian Geometry and Geometric Analysis, ser. Springer Universitat texts. Springer, 2005. [Online]. Available: https://books.google.fr/books?id=uVTB5c35Fx0C

. [11] Zeestraten, M., Havoutis, I., Silv'erio, J., Calinon, S. and Caldwell, D.G. (2017) An Approach for Imitation Learning on Riemannian Manifolds. IEEE Robotics and Automation Letters (RA-L), vol. 2, no. 3, 1240-1247. https://doi.org/10.1109/LRA.2017.2657001

. [12] Lin, J.F.-S. and Kulic, D. (2014) Online Segmentation of Human Motion for Automated Rehabilitation Exercise Analysis. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22, 168-180. https://doi.org/10.1109/TNSRE.2013.2259640

. [13] Devanne, M. and Nguyen, S.M. (2017) Multi-Level Motion Analysis for Physical Exercises Assessment in Kinaesthetic Rehabilitation. 2017 IEEE-RAS 17th IEEE International Conference on Humanoid Robots (Humanoids), Birmingham, 15-17 November 2017, 529-534. https://doi.org/10.1109/HUMANOIDS.2017.8246923

. [14] Sara, B., Kaska, P.-P. and Tim, S. (2014) ECHOES: An Intelligent Serious Game for Fostering Social Communication in Children with Autism. Information Sciences, 264, 41-60. https://doi.org/10.1016/j.ins.2013.10.027

. [15] Mazon, C., Fage, C. and Sauzéon, H. (2019) Effectiveness and Usability of Technology-Based in-Terventions for Children and Adolescents with ASD: A Systematic Review of Reliability, Consistency, Generalization and Durability Related to the Effects of Intervention. Computers in Human Behavior, 93, 235-251.

. [16] Calinon, S., Guenter, F. and Billard, A. (2007) On Learning, Representing, and Generalizing a Task in a Humanoid Robot. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 37, 286-298.https://doi.org/10.1109/TSMCB.2006.886952

. [17] Karcher, H. (1977) Riemannian Center of Mass and Mollifier Smoothing. Communications on Pure and Applied Mathematics, 30, 509-541. https://doi.org/10.1002/cpa.3160300502

. [18] Simo-Serra, E., Torras, C. and Moreno-Noguer, F. (2016) 3D Human Pose Tracking Priors Using Geodesic Mixture Models. International Journal of Computer Vision, 122, 1-21. https://doi.org/10.1007/s11263-016-0941-2