



HAL
open science

Gestures Imitation with a Mobile Robot in the Context of Human-Robot Interaction (HRI) for Children with Autism

Amir Aly, Adriana Tapus

► **To cite this version:**

Amir Aly, Adriana Tapus. Gestures Imitation with a Mobile Robot in the Context of Human-Robot Interaction (HRI) for Children with Autism. Proceedings of the 3rd Workshop for Young Researchers on Human-Friendly Robotics (HFR), Oct 2010, Tübingen, Germany. hal-01215275

HAL Id: hal-01215275

<https://hal.science/hal-01215275>

Submitted on 13 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Gestures Imitation with a Mobile Robot in the Context of Human-Robot Interaction (HRI) for Children with Autism

Amir Aly and Adriana Tapus

Abstract—In human-robot interaction, gestures play an important role in transferring messages from the human to the robot and vice versa. The robot has to be able to properly understand the gestures in order to respond in an appropriate manner. This research work focuses on dynamic gestures' recognition using robot's cameras, which capture the performed gestures in a certain interactional scenario. The gestures are processed and classified to a pre-defined category of gestures. Our gesture recognition system is currently part of a robot behavior control architecture developed for children with autism.

I. INTRODUCTION

Gestures are a form of non verbal communication that can be used not only for human-human interaction but also for human-machine interaction. In this work, dynamic gestures are classified into two categories: small and large amplitude gestures due to differences in speed between persons performing the gestures. Each of these categories has a specific classification method. These classification approaches are linked together using the optical flow so as to directly determine the nature of a gesture as part of one of the two predefined gesture categories.

II. SMALL AMPLITUDE GESTURES

A. Key frames Extraction

Key frames are important frames in a video stream, which indicate more details than the other neighbouring frames. These key frames could be specified by comparing the histogram of each frame to the histograms of its neighboring frames (figure 1) [1].

B. Difference image

The extracted key frames from a video stream are, usually, about 6 or 7 frames (for a video of duration 4 seconds and 140 frames). As long as the amplitude of this category of gestures is small (e.g., like waving Bye-Bye using one or two hands), it is required to choose just two key frames that well present the gesture performed. Thereafter, it is useful to subtract the extracted final two key frames (the less correlated) in order to eliminate unnecessary details like background and static parts of the body (figure 2) [2].

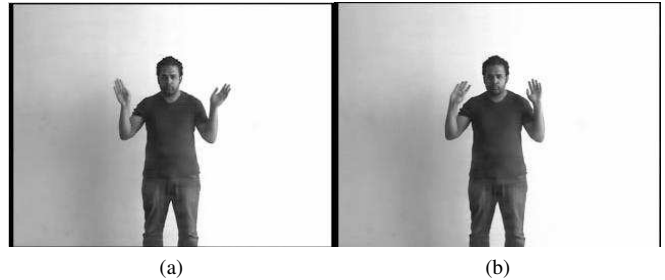


Fig. 1: Final extracted two key frames



Fig. 2: Difference image with the characteristic vector

C. Characteristic vector and Classification

The characteristic vector could be presented by an ellipse occupying the span between the initial and the final hand positions (figure 2) [2].

D. Results

The database of the small amplitude gestures is composed of 6 gestures for 9 persons which are in turn: Bye-Bye 2 hands, Bye-Bye right hand, Bye-Bye left hand, Stop, Yes, and No. The classification system used is the K-nearest neighbor (KNN) supported by the cross validation. The total average recognition score is about 95 % (see figure 3).

III. LARGE AMPLITUDE GESTURES

The processing procedures of large amplitude gestures are the same as in the small amplitude gestures. However, these gestures are characterized in terms of consequent difference images (4 or 5 according to the total number of frames in the video) to compensate the effect of speed differences between persons performing gestures. Each difference image will be characterized as in the small amplitude gestures [2].

The database of the large amplitude gestures is composed of 9 gestures for 9 persons, which are in turn: kowtow, walk

Amir Aly is a PhD student in the Cognitive Robotics Lab, Ecole Nationale Supérieure de Techniques Avancées (ENSTA-ParisTech), France (email: amir.aly@ensta-paristech.fr)

Prof. Adriana Tapus is with Cognitive Robotics Lab, Ecole Nationale Supérieure de Techniques Avancées (ENSTA-ParisTech), France (email: adriana.tapus@ensta-paristech.fr)

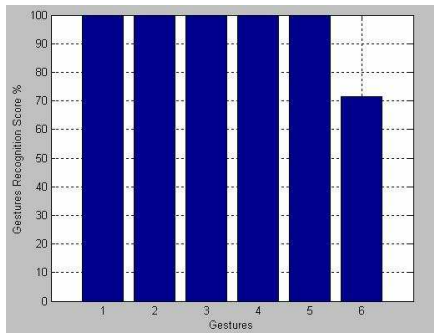


Fig. 3: Small amplitude gestures' recognition score

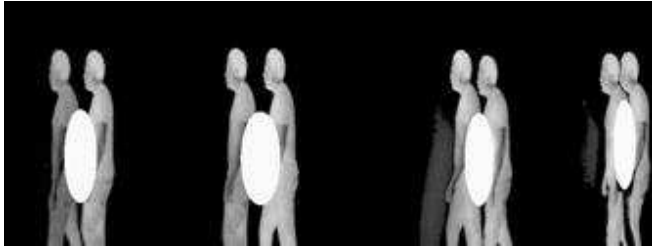


Fig. 4: Motion consequent states characterization

from left to right, walk from right to left, Right arm up, Right arm down, Right arm clockwise rotation, Right arm counterclockwise rotation, Right arm pointing to right, and left arm pointing to left. The classification system used is the same K-nearest neighbor (KNN) supported by the cross validation, but it has a different approach in processing the database by using 4 or 5 characteristic vectors for each gesture instead of one characteristic vector as in the small amplitude gestures system. The total average recognition score is about 93 % (see figure 5).

IV. OPTICAL FLOW

It is necessary for a gesture recognition system to respond in real time in order to precisely decide online the category of a performed gesture. Optical flow could be useful in this concern by tracking the dynamic part of the body using velocity fields that wrap consequent images in a video. Figure 6, indicates the tracked dynamic part of the body surrounded by the green rectangle. The main idea here is to

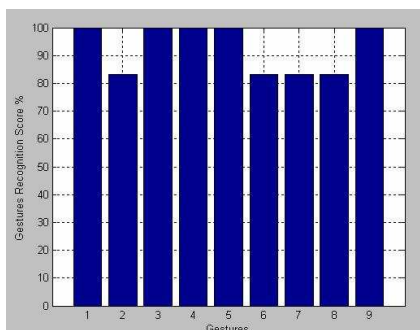


Fig. 5: Large amplitude gestures' recognition score

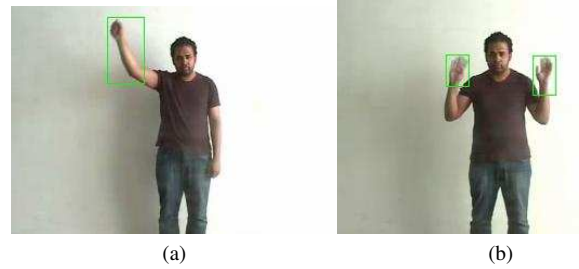


Fig. 6: Dynamic body parts tracking

calculate the centroid of the rectangle in each acquired video frame in real time, then calculating the Euclidian distance between centroid's coordinates of the first and final frames. This distance could be a real indicator for the test gesture as being of low or large amplitude [3].

V. SOCIAL ROBOTS AND AUTISM

Social robots play an important role in ameliorating the social interactional behavior of autistic children. In this study, we try to observe whether the robot can induce imitation in the user, and sustain an imitation game over time. Mimicry and imitation are considered powerful tools of human engagement and we are currently testing them in the HRI context.

Motor imitation problems are common in autism. Imitation is a combination of cognitive-representational and visual perceptual-motor components. Motor imitation can be divided into meaningful transitive and intransitive gestures versus non-meaningful single and sequential hand postures. Meaningful gestures are symbolic and language related. In this study, we focus only on intransitive gestures, e.g. wave goodbye, which are communicative gestures. Imitation of meaningful gestures requires a proper visual perceptual discrimination of the action, comprehension of the meaning of the action, and linkage to previous information about the 'timespace' formulas of a to-be-performed action. The imitation of meaningful gestures requires cognitive-representational as well as perceptual-motor processes [4].

Teaching imitation, and demonstrating the behavior of an autistic child through imitation, have both led to improved social responsiveness. Use of language and gestures within the zone of proximal development (i.e., made possible in collaboration with others) may encourage the child to use the same form themselves.

The purpose of this study is to use the triadic interaction between child, robot and therapist/family member so as to help and improve childrens performance on the imitation on meaningful gestures and to be more precise on intransitive gestures which are known as communication gestures (see Figure 7). In our work, the Nao humanoid robot is used. A video of the imitation system is available at: http://www.youtube.com/watch?v=2b8CkK_ghvU.

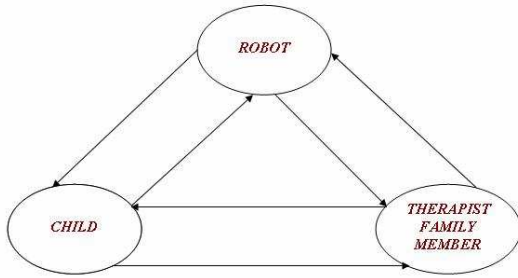


Fig. 7: Triadic interaction: Child, Robot, Therapist

VI. CONCLUSION

This research work presents a new methodology for gesture recognition that is used in a scenario designed for children with autism. The robot has to be able to understand and classify the perceived gestures in order to perform an appropriate interactional reaction. Two dynamic gesture categories are involved in this study according to their amplitude as being small or large. Moreover, tracking the dynamic part of the body by the optical flow is an important element in differentiating between the two categories for any test gesture. The total average recognition score for the whole system of small and large amplitude gestures is about 94%. More results will be available by the time of the workshop, as this paper reports ongoing work in progress.

VII. ACKNOWLEDGMENTS

This work is supported by French National Research Agency (ANR) through Chaire d'Excellence program 2009 (project HRIAA - Human-Robot Interaction for Assistive Applications).

REFERENCES

- [1] L. Shao and L. Ji, "Motion histogram analysis based key frame extraction for human action/activity representation," in *Canadian Conference on Computer and Robot Vision*, 2009.
- [2] G. Rigoll, A. Kosmala, and S. Eickeler, "High performance real time gesture recognition using hidden markov models," in *IEEE int.conf.on Acoustics, Speech and Signal Processing (ICASSP)*, 1997.
- [3] M. Proesmans, L. Van Gool, E. Pauwels, and A. Oosterlinck, "Determination of optical flow and its discontinuities using non-linear diffusion," in *3rd European Conference on Computer Vision (ECCV)*, 1994.
- [4] S. Rogers and B. Pennington, "A theoretical approach to the deficits in infantile autism," *Development and Psychopathology*, vol. 3, pp. 137–162, 1991.