The Attentive Machine: be Different!

Julien Leroy, Nicolas Riche, François Zajega, Matei Mancas, Joelle Tilmanne, Bernard Gosselin, Thierry Dutoit

University of Mons (UMONS), Engineering Faculty, IT Research Unit
20, Place du Parc, 7000, Mons
{julien.leroy, nicolas.riche, françois.zajega, matei.mancas, joelle.tilmanne, bernard.gosselin, thierry.dutoit}@umons.ac.be

Abstract. We will demonstrate an intelligent Machine which is capable to choose within a small group of people (typically 3 people) the one it will interact with. Depending on people behavior, this person may change. The participants can thus compete to be chosen by the Machine. We use the Kinect sensor to capture both classical 2D video and depth map of the participants. Video-projection and audio feedback are provided to the participants.

1. Social feature extraction

The main feature which is extracted is the personal space of the participants. Social studies [1] showed that humans have different "ego-spaces": the public space (from around 3.5 meters in green on Figure 1, left image), the social space where interaction is possible (from around 1.2 meters in blue on Figure 1, left image), the personal space for close interaction (from around 0.45 meters in yellow on Figure 1, left image) and the intimate space (in red on Figure 1, left image). Those measures vary of course depending on cultural and personal contexts. This space is extracted in 3D [2] by using OpenGL and a Microsoft Kinect sensor (Figure 1, right image).

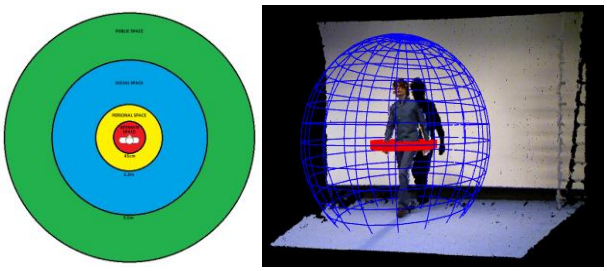

Figure 1: Right: ego-spaces, Left: 3D extraction of intimate space (red cylinder) and personal space (blue sphere)

The real-time 3D extraction of the inter-personal distances (Figure 2: example with two participants), along with the height of the people and their position and velocity provides a set of both dynamic and static, low-level and mid-level features.

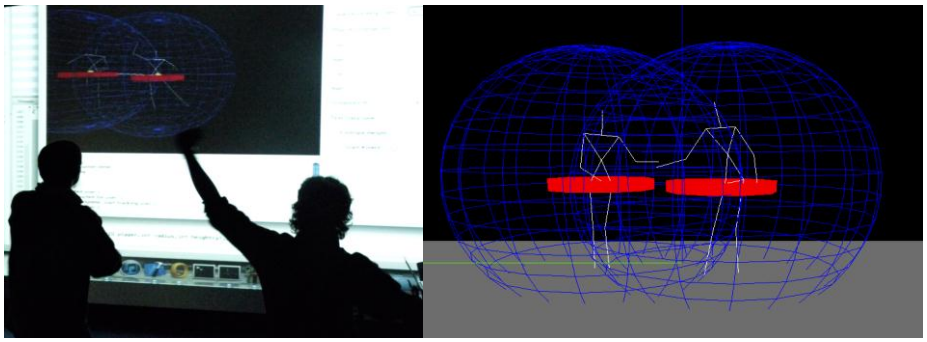

Figure 2: Two participants' real-time interaction and extraction of social cues

2. Find the most interesting people

Those features are than sent to an attention algorithm which will compute feature's relative contrast [3]. As only 3 people are taken into account here, the rarity approach of attention is not relevant, thus we use the global contrast from [3] only. If the features of a person compared with the others are contrasted enough, this person is potentially the most "worthy of interest" for the Machine which will give him the possibility to interact with it. In order to cope with several features in the same time, empirical weights are given and the person with the higher overall weighted contrast is selected.

As a result, while two out of the three people are together, the one who is alone will be selected, if the three people have the same distances, the one in the middle will be selected. If one out of three is sitting on the ground, he will be selected, while if two persons are sitting on the ground, the one standing will be selected. Concerning dynamical features like the 3D speed, they highly attract the system when one person is different (faster, slower), but those dynamic features should be perpetuated by a static feature contrast in order to provide a stable selection of the person.

3. User feedback

The video feedback which is a wall of HAL's eyes (Figure 3) will change, and all the eyes will focus on the selected person. The idea of several identical objects focusing on a person comes from [4] where people are very excited in trying to be in the mirrors' focus. The selected person will be able to create sounds by using simple gestures. This should push people to try to be different in order to be able to have this interaction. If another person performs an interesting behavior, the Machine will focus some of the HAL eyes on this second person which is a sign that the interaction could change to the second person. If the first person does not act in order to increase his own interestingness within a given time, he will lose the selection and only the second person will be able to interact and create sounds.

This set-up is of course interesting for its natural user selection in a multi-user scenario, but also because it let us observe the social behavior of the users while interacting with such a machine and the other users.

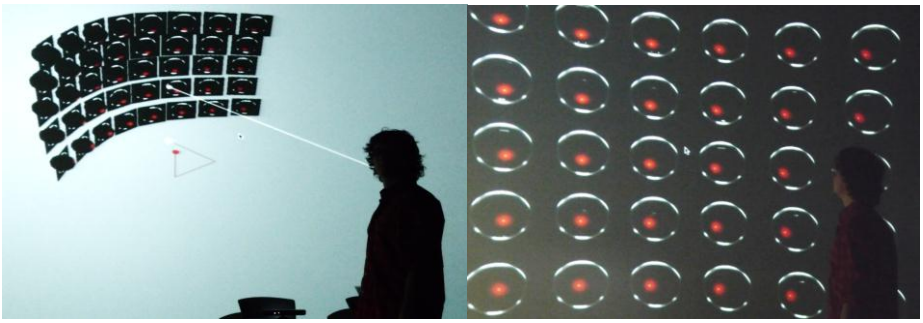

Figure 3: Left image: 3D structure of the HAL's eyes, Right image: all the eyes focus on the person in front of the screen.

4. Requirements

This demo requires a 6x6 meters space and electrical power. If possible, a video projector, a projection screen and a loudspeaker would be appreciated. Half a day should be enough to set up the demo.

5. Acknowledgments

This demo was created within the NUMEDIART Institute (www.numediart.org) funded by the Walloon region, Belgium.

6. References

[1] E.T. Hall, "The Hidden Dimension", Anchor Books, 1966.
[2] J. Leroy, M. Mancas, B. Gosselin, "Personal Space Augmented Reality Tool", First joint WIC/IEEE SP Symposium on Information Theory and Signal Processing in the Benelux, Bruxelles, Belgium, 2011.
[3] M. Mancas, B. Gosselin, B. Macq, "A Three-Level Computational Attention Model", Proceedings of ICVS Workshop on Computational Attention & Applications (WCAA-2007), Bielefeld, Germany, 2007.
[4] Audience set-up - http://www.chrisoshea.org/audience