KinAct: the attentive social game demonstration

F. Zajéga, M. Mancas, R. Ben Madhkour, J. Leroy, N. Riche, F. Rocca, Y. P. Rybarczyk, and T. Dutoit

***Abstract*—In this paper we present the demonstration of an intelligent game system capable of selecting the players who exhibit the most outstanding behavior from groups of people playing on a network. The system uses both static and dynamic features extracted from the upper part of the players' bodies such as symmetry, contraction index, motion index or height. Those features are extracted using the RGB-D Kinect sensor and their relative contrast and time evolution enable an adaptive selection of the most salient or different behavior without any complex rules.**

***Index Terms*—saliency, attention, game, kinect, bottom-up, top-down, behavioral features, eye tracking, social interaction**

I. Introduction

EFFICIENT and low-cost devices as the Kinect sensor opened new highways in gaming community. People detection and skeleton extraction in real life light conditions became a reality and more and more complex interactions are possible. Despite those new possibilities, still the interactions are achieved with pre-learnt and pre-programmed gestures. This means that if game scenarios are different from what is expected, the system cannot cope with the novelty.

In this paper, we demonstrate a game which is able to select the active player in an attentive way. The system only uses players' behavior during their observation.

Beyond the ability to choose the active user, the system could be used to study the social organization of the player teams or the influence of the players' localization.

The characteristics we use are related to 3D motion and expressive features as those developed by Camurri et al. [1].

This paper is organized as follows: section II discusses the game idea. Section III describes features extraction while section IV deals with the attention mechanism. A conclusion is provided in section V.

II. Game Description

We based our game on "World of Goo" [2], a game developed by 2D Boy. The purpose proposed to players is to build a structure made out of clouds interconnected by semi-rigid wires.

For this demonstration, two teams of two players each are playing trying to build a cloud structure and rich the sky before the other team to win. Figure 1 shows a screenshot of the game where the right player has been selected by the system to drop a cloud and build the structure.

F. Zajega, M. Mancas, R. Ben Madhkour, J. Leroy, N. Riche, F. Rocca and T. Dutoit are from the TCTS Lab (http://www.tcts.fpms.ac.be), University of Mons (UMons), Belgium.

Y. P. Rybarczyk is from the Departamento de Engenharia Electrotécnica, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal.

Fig. 1. *Right player selected to drop a cloud.*

To be selected by the system the player has to surprise the machine by behaving in a different way compared to the 3 other players.

An additional difficulty is in the cloud structure evolution. Poor choices while droping the clouds can make it collapse. Indeed a physical engine was implemented and clouds with too few links result in week and unstable structures)

The demonstration uses one video projector for player feedback and 2 Kinects (one working as a slave and the other as a master) communicating together on a local network. Each team of two people is analyzed by one Kinect and the results are fused in a single playground of four players.

III. Features Extraction

The first step for an interactive machine is to extract features from the observed people. For that purpose, we use the Kinect sensor for its ability to extract smooth depth maps in complex illumination conditions. Libraries as OpenNI [3] are available to detect human silhouettes and extract anatomical features from skeletons tracking (Figure 2).

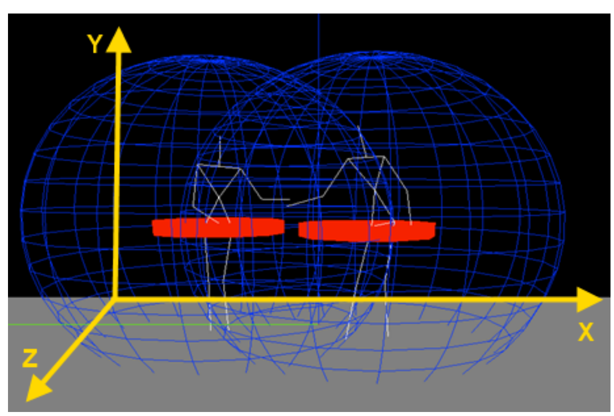

Fig. 2. *(X, Y, Z) axes and real-time skeleton extraction.*

Only the points of the upper body are taken into account. Four features based on [1] are extracted from the upper body part which is more stable than the lower one.

One of the four features is dynamic, namely the motion index. It is computed as the mean variation of the same skeleton points between two frames in 3D (on X, Y and Z). This feature will describe an excitement degree or movement transition of the body.

A second feature extracted from the upper body part is a static feature, namely the asymmetry index. This feature is only computed on the X axis by differencing the distances between the barycenter point and the right shoulder, elbow and hand points with the left ones. This index provides information about the symmetry of the upper body.

The third feature is the contraction index. This index is the ratio between the maximal distance between skeleton points on X axis and the maximal distance on the Y axis. This index tells us if the person is more or less contracted.

The fourth and final feature is the player height. That one is simply computed by measuring the player barycenter Y coordinate. It provides information about the player position (standing, sitting).

After normalization, those four features provide a quite complete description about the level of excitement, and the upper body configuration of each player. They describe classical human behaviour like waving a hand (assymmetric and moving), two hands (symmetric and moving), standing or sitting, etc...

IV. Computational Attention: Who Is Different?

The aim of computational attention is to provide algorithms to automatically predict human attention. The term of attention refers to the whole attentional process that allows one to focus on some stimuli at the expense of others.

Much research has already addressed the computational visual attention, but these works were devoted almost exclusively to 2D image analysis. Up to now little has been done on 3D data [4]. The arrival on the market of low cost 3D cameras opens up new prospects for developing algorithms for visual attention, which should make the move towards more realistic ambient intelligence systems.

As stated in [5] a feature does not attract attention by itself: bright and dark, locally contrasted areas or not, red or blue can equally attract human attention depending on their context. The main cue, which involves bottom-up attention, is the contrast and rarity of a feature in a given context.

The approach here uses the feature contrast between players. As explained in Figure 3, the player who is the most different compared to the others will be selected.

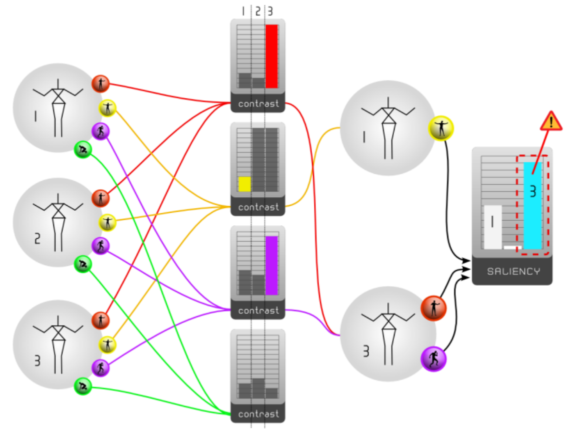

Fig. 3. *Example of bottom-up saliency computation for 3 players. For each of the four behavioral features, a contrast is computed between the players. A threshold will eliminate features which are not contrasted enough between players (here the forth feature in green is eliminated). The player having more contrasted features with higher weights will be selected as the most salient (here the third player)*

V. Conclusion

In addition to the eye-tracking tests that we performed and which confirmed the fact that that humans also look at people behaving differently, several tests were done with players which did not know the system. Even if at the beginning the learning curve is less fast than for a classical game where each player plays when its turn arrives, the players enjoy much more this game which never becomes boring. After the first game round they understand the rules and beginning from the second game round they have fun.

Moreover, from the second game round, team strategies are found by players (e.g. one of the two players of one team imitates the others to let his co-player be more different and selected).

We performed tests where each player drops one cloud in a simple sequential strategy, and the players understand much faster the rules, but they interact much less (sometimes they just wait for their turn without talking with the others). Generally, this game when using the attention strategy and after a longer adaptation period really needs people to observe each others and communicate and becomes a social mediator between them.

Acknowledgment

Numediart is a long-term research program centered on Digital Media Arts, funded by Région Wallonne, Belgium (grant N°716631).

References

[1] A. Camurri, C. Canepa, P. Coletta, N. Ferrari, B. Mazzarino, and G. Volpe, “The interactive piece the bow is bent and drawn,” in *Proc. of 3rd ACM Intl. Conf. on Digital Interactive Media in Entertainment and Arts (DIMEA 2008)*, 2008, pp. pp. 376–383.

[2] 2dboy, “World of goo is a physics-based puzzle game developped by 2dboy.” [Online]. Available: http://www.worldofgoo.com

[3] W. Garage, “Openni,” multi-platform library for skeleton extraction from depth maps. [Online]. Available: http://www.openni.org

[4] N. Riche, M. Mancas, B. Gosselin, and T. Dutoit, “3d saliency for abnormal motion selection: the role of the depth map,” in *Proceedings of the 8th International Conference on Computer Vision Systems (IVCS2011)*, 4 2011.

[5] M. Mancas, *Computational attention: Towards attentive computers*, similar edition ed. CIACO University Distributors, 2007.