1. KinAct: social interaction for data processing

Principal investigator: Matei Mancas (University of Mons, Engineering Faculty)

2. Objectives:

There are three main objectives in this project:

1. Analyze the behavior of groups of people in two different scenarios (standing and sitting).

Low-level features like speed and direction can be extracted from complex videos using optical flow algorithms [1]. Moreover, mid-level features related to social signal processing [2] as personal spaces [3],[4],[5] can also be extracted. Figure 1 shows on the left velocity features extraction in a crowd while on the left social signal features are extracted and displayed in 3D: the intimate space in red, the personal space in blue.

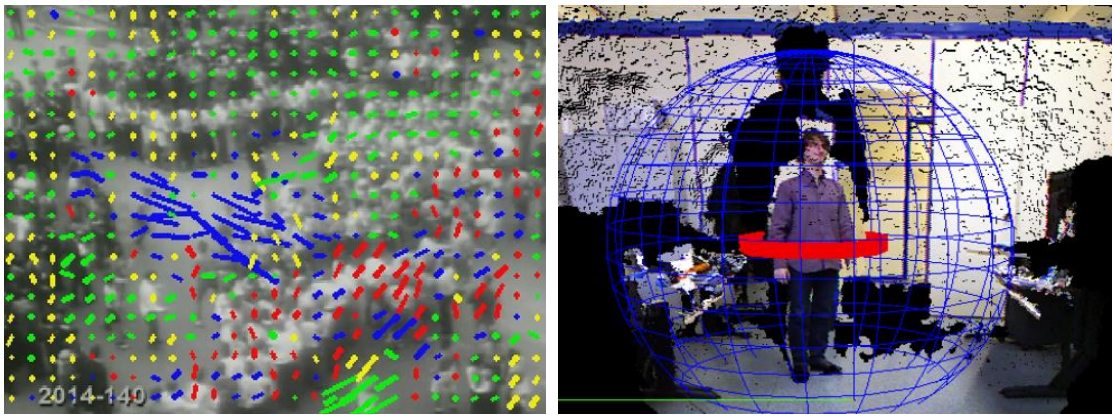

Figure 1: Left image: low-level features extracted from a large group of moving people using optical flow algorithms, Right image: mid-level features: 3D personal spaces around a user acquired using the Kinect 3D camera [10].

2. Use people behavior to interact with media (interactive video and video game)

A first approach here can be to directly use the quantity of motion and distance between the users to interact with video games or other media. A more sophisticated approach uses their relative rarity which highlights the rarest, so the most "interesting" features. This can be done both for still images [6] but also be generalized to videos [7]. This technique can be used for example to highlight unusual social connections…

An idea of the style of game which can be used can be seen in Figure 2. A game similar in concept with the "world of goo" [8] can be used where the distances between people would replace mouse placement to build a structure having physical properties. Also unusual interaction could be shown on the structure by using different colors or audio feedback for example.

A video could also be projected, and depending on the spectators' quantity of motion, the data can be summarized: the most important data obviously will imply more intense reactions of the spectators. Most of the methods of video summarization [9] use features extracted from the video and audio track, but here we will take into consideration spectators' reactions to keep the frames which are the most representatives and discard the others.

Figure 2: Snapshot from the game "world of goo" where blobs can be aggregated into wider structures having physical properties.

3. **Analyzing long-term models of people behavior** through the result of the game or by summarizing the video function on people behavior

Once the results of the social interaction are recorded (lattice with people positions for the interactive game as the one which is under construction on Figure 2, or video summary), it is possible to compare in long-term scenarios different people behavior (people knowing each-other or not, men/women, different cultures, older/younger …). This can lead to interesting findings about people social interactions and also can be used to model those social interactions and find unusual experiments (where different constraints would have been applied like a maximum time to move, to be as far as possible from some persons, etc…).

3. Applications:

Applications of such systems could be in team building and social interaction improvement through games and video summarization. More broadly, those techniques can help in establishing frameworks on the use of humans to help in data processing or to use computers to help in human social relationship.

4. Resources:

We will bring several Kinect [10] cameras which can provide both classical 2D color images and a high-quality depth map. We will also bring a good video-projector, but probably there will be need for one local video projector. The game which will be used will be ready for the workshop.

The project needs a quite large room with/without chairs which has the possibility of video projections on a screen or white walls (typically a class room).

5. Tentative work plan:

Each week is divided in two tracks: “a” and “b”. Track “a” is more technical, while track “b” more high-level. Those two tracks can be hold in parallel.

W1_a: Feature extraction from groups of people (low-level and mid-level features)
W1_b: Scenario 1 preparation: video game with standing people (precise scenario building, video game tests), Scenario 2 preparation: videos with sitting people (precise scenario building, video selection…)

W2_a: Testing of attention algorithm and integration of the results on the game lattice (different colors for the nodes for example – see Figure2)
W2_b: Direct mapping of the extracted features on the game and first tests

W3_a: Integration of attention algorithms in Scenario 2
W3_b: Tests of the full Scenario 1 and first tests on scenario 2

W4_a: Processing of the long-term models (the lattice from the game and the summarization of the videos)
W4_b: Tests on the two scenarios

6. Project benefits:

An output would be a package to be used along with a Kinect camera and a video projector to reproduce at least one of the two scenarios.

7. Team:

- Matei Mancas holds a PhD from the Engineering faculty of the University of Mons in the field of computational attention. He is currently developing attention algorithms for complex videos (crowds, groups of people, etc…)
- Julien Leroy holds a MSc from the Engineering faculty of the University of Mons. He is now a PhD student and he is focusing on social signal processing and mainly on personal spaces detection and analysis.
- Nicolas Riche also holds a MSc from the Engineering faculty of the University of Mons. He is now a PhD student working on the set-up of a multimodal computational attention system.
- Other people needed:
  This is a highly interdisciplinary project. We also need the following profiles:
  - A computer scientist with good programing skills (C/C++) with optional interest to social signal processing and computational attention. He/she would work on system integration and help on social signal processing and computational attention algorithms.
  - A computer scientist interested in video games and/or video montage with sufficient programing skills. He/she would work on the video game adaptation to the scenarios, to video montage for the second scenario and more generally to the module of long-term activity recording from the video game and the summarization of the videos.
  - A psychologist to set-up pertinent scenarios details and the most interesting psychological tests.
  - Of course any interested people are welcome both from engineering and humanities.

8. References:

[1] G. Farnebäck, “Two-Frame Motion Estimation Based on Polynomial Expansion”, Lecture Notes in Computer Science, 2003

[2] A.Vinciarelli, M.Pantic and H.Bourlard, “Social Signal Processing: Survey of an Emerging Domain”, Image and Vision Computing Journal, Vol. 27, No. 12, pp. 1743-1759, 2009

[3] E.T.Hall, ”The silent language” Doubleday, 1959

[4] E.Sundstorm, I.Altman, ”Interpersonal Relationship and Personal Space: Reseach Review and Theoritical Model”, Human Ecology,Vol.4,No.1, 1976

[5] T.Amaoka, H.Laga,S.Saito,M.Nakajima, ”Personal Space Modeling for Human-Computer Interaction”, ICEC, 2010

[6] M. Mancas, “Relative Influence of Bottom-Up and Top-Down Attention”, Attention in Cognitive Systems, Lecture Notes in Computer Science, Volume 5395/2009, 2009

[7] Dynamic saliency map in complex contexts: http://il.youtube.com/watch?v=iKYLpP6GQMY

[8] World of Goo video game: http://www.youtube.com/watch?v=BzQuTsIKd14

[9] A. G. Moneya and Harry Agius, “Video summarisation: A conceptual framework and survey of the state of the art”, Journal of Visual Communication and Image Representation, Volume 19, Issue 2, 2008

[10] PrimeSense website: http://www.primesense.com/