HYPERSOCIAL MUSEUM: ADDRESSING THE SOCIAL INTERACTION CHALLENGE WITH MUSEUM SCENARIOS AND ATTENTION-BASED APPROACHES

Matei Mancas[1], *Donald Glowinski* [2], *Paul Brunet* [3], *Francesca Cavallero* [2], *Caroline Machy* [4], *Pieter-Jan Maes* [5], *Stella Paschalidou* [6], *Manoj Kumar Rajagopal* [7], *Stefania Schibeci* [2], *Laura Vincze* [8], *Gualtiero Volpe* [2]

[1] Laboratoire de Théorie des Circuits et Traitement du Signal (TCTS), Faculté Polytechnique de Mons (FPMs), Belgique
[2] InfoMus Lab, DIST - University of Genova, Italy
[3] Queen's University of Belfast, UK, [4] Multitel, Belgium,
[5] Institute of Psychoacoustics and Electronic Music (IPEM), University of Ghent, Belgium
[6] University of Crete, Greece, [7] Telecom ParisTech, France, [8] University of Pisa, Italy

ABSTRACT

This work intended to show that the use of museum scenarios and computational attention models are a good way to achieve social interaction studies. The use of attention-based algorithm allows the possibility to dynamically analyze visitors' behavior in museums in long-term context where some typical behavior models are learnt. Moreover, by providing feedback to visitors, it is possible to interact with them and also to foster interaction between visitors: the system here becomes a sort of mediator in human to human interaction.

KEYWORDS

Social signal, social interaction, expressive gesture, non-verbal communication, context, computational attention, saliency

1. INTRODUCTION

1.1. Social Signal and Non-verbal Cues

Recent ambient intelligence applications based on user-centric concepts have created a common motivation across different research disciplines to understand and model non-verbal affective and social behaviors. Affective and social interactions are an extremely relevant component for analysis of human behavior, since they provide significant information on its affective and social context.

Social intelligence or social competencies, understood as the ability to deal effectively in interpersonal contexts, is a paradigmatic human ability, widely studied in psychology and more recently in neurophysiology. Social intelligence is also receiving a growing interest from the ICT communities, e.g., in the framework of social networks, networked media, Future Internet, user-centric media.

Obtaining automatically behaviors patterns that inform on affective and social activities of the visitor can not only guide an organization in making changes in its physical and information technology environment but also in evaluating the effect of these changes.

Moreover, providing people with feedback on their own or other people behavior is likely to influence their current reactions and foster human to human cooperation.

In this work we investigated two approaches of social intelligence: the social signal analysis and the social interaction influence in museum-based scenarios.

1.2. Museum Scenarios

Museum scenarios offer real-life ecological conditions and enable the analysis of individual as well as collective behaviors: people in museum rooms can be found alone, with friends or of course, with unknown people. Analysis can thus be carried out in most of the social situations and interactions.

Moreover people in museums are likely to have some time to interact either with other people either with computers. This point is very interesting for fostering social behavior through different interactions with computers. Ethnographic studies demonstrated since the 80ies the feasibility of the modeling of museum visitors, for both profiling of museographic projects and for the design of more effective interactive museum setups.

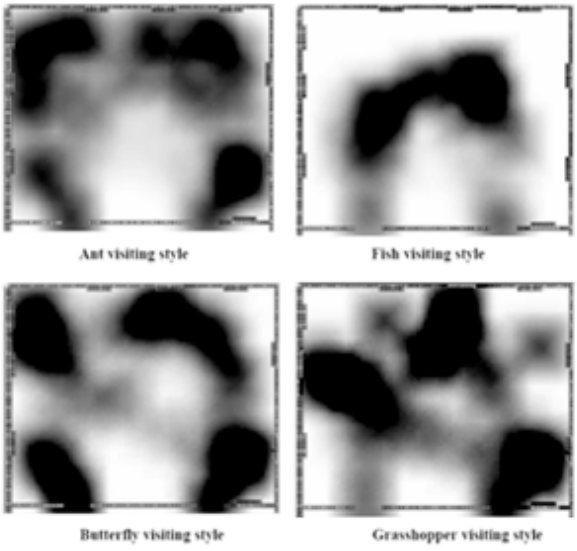

Figure 1: Veron and Levasseur museographic typology

For example, the typology of 4 museum visiting styles has been proposed by Veron and Levasseur [10]. This typology is able to make some restrictions on the motion of visitors while their behavior remains free and ecological enough. It is based on 4 museum visiting styles as shown in figure 1: the "ant" visitor who tends to follow a linear path and spends a lot of time observing almost all the exhibits, the "fish" visitor who moves mainly in the

centre of the room, the “butterfly” visitor who often changes his visiting direction and stops frequently and finally, the "grasshopper” visitor who seems to have a preference for some preselected exhibits and spends a lot of time observing them while ignoring the others.

2. SYSTEM ARCHITECTURE

2.1. Social Signal and Attention

Social interaction is based both on verbal and non-verbal cues. We focus here on non-verbal affective and social behaviors which are of a great importance in social communication. Non-verbal behaviors include the expressive gestures which are, according to Kurtenbach and Hulteen “a movement of the body that contains information” [5]. Thus, gestures can be named expressive since the information they carry is an expressive content, i.e., content related to the emotional and social communication sphere. Expressive gestures are used to communicate voluntary or involuntary information to other people and their interpretation can more or less modify the emotional state of the other people. Expressive gestures can be used either to initiate an interaction (before using verbal interaction) or to provide additional information to the verbal one (which can be the same, different, or in contradiction with the verbal communication). They can also be used both in small and large group interaction. A multilayered framework for automatic expressive gesture analysis was proposed by Camurri et al. [3]. In this framework, expressive gestures are described with a set of motion features that specify how the expressive content is encoded. Different attempts can be found in literature to map a set of expressive gesture features with one of the emotional dimensions that are considered to describe the entire space of conscious emotional experience [11] which are valence and activation. For example activation dimension has been mapped to expressive features such as the amount of energy of a person [2] . However, the main shortcoming of expressive gesture analysis is the scarce consideration of the context in which expressive gestures take place. The context we focus on has to be considered both related to the temporal dynamics of a motion feature and to the spatial context of this feature if the behavior analysis of more than one user is performed.

First studies related to the context-aware analysis of expressivity which established a relationship between the arousal level of an emotion and the uncertainty of a visual stimulus can be found in [1]. Mehrabian and Russell formulated the information rate-arousal hypothesis and confirmed a linear correlation between information rates of a real environment and emotion arousal [9]. These studies put in evidence that the saliency of an event can be related to the novelty of an expressive content.

2.2. Computational Attention or Automatic Modeling of Human Attention

The aim of computational attention is to automatically predict human attention on multimodal data such as sounds, images, video sequences, smell or taste, etc... The term attention refers to the whole attentional process that allows one to focus on some stimuli at the expense of others.

A very important common objective concept between social signal and attention is the context. Social interaction is impossible out of a context and attention is also a notion which cannot be computed out of any context. Moreover, the purpose of attention is to detect surprising behavior which should contain a high amount of information. In social communication, the detection of novel events is susceptible to bring new information and thus to modify the emotional context of the interaction.

Human attention mainly consists of two processes: a bottom-up and a top-down one. Bottom-up attention uses low-level signal features to find the most salient or outstanding objects. Top-down attention uses a priori knowledge about the scene or task-oriented knowledge in order to modify (inhibit or enhance) the bottom-up saliency. While numerous models were provided for attention on still images, time-evolving two-dimensional signals such as videos have been less investigated. Nevertheless, some of the authors providing static attention approaches generalized their models to videos, but very few to audio or time-evolving feature signals (for a detailed review, see [6]). Most of these methods provide bottom-up attention approaches. To our knowledge, a majority of these computational models focuses on low-level motion features (e.g., displacement of people). We suggest in this paper that computational models would gain considering higher-level motion features related to full-body movements to better capture the expressive gestures that characterize the communication of an emotion. Our approach is able to easily adapt to different spatial, short and long temporal contexts.

As we already stated in [7, 6] 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. In the same way, motion can be as interesting as the lack of motion depending on the context. The main cue which involves bottom-up attention is the rarity and contrast of a feature in a given context. A low-computational-cost quantification of rarity was achieved referring to the notion of self-information [8]. This approach uses context at three levels (spatial context, short-term and long-term contexts). Concerning social signal, the most relevant context are the spatial and the long-term ones.

2.2.1. *Instantaneous Level*

Let us consider a collective context, e.g., a group with interacting people. Motion features (e.g., speed, direction) characterizing each moving person are compared at each instant. Salient motion behavior (e.g., one person speed very different from the others) immediately pops-out and attracts attention. This refers to pre-attentive human processes, usually faster than 200 milliseconds. In our approach, motion saliency detection at instantaneous level is computed over time intervals of 200ms to 1s.

2.2.2. *Long-term Attention Modulation*

The long-term memory (LTM) [10] component of the model processes the saliency index in a time interval from several seconds to much longer periods (related to the application time scale). The output is a modification of the instantaneous attention indexes in such interval according to their considered recurrence. The attention amplitude map in the different locations of the observed scene along time is progressively built. This leads to the definition of areas, which capture attention more than others: e.g., a street accumulates more attention than a grassy area. The scene can thus be segmented into several areas of attention accumulation and the motion in these areas can be summarized by only one motion vector per area. If a moving object passes through one of these areas and it has a motion vector similar to the one summarizing this area, its attention is inhibited (usual motion). If this object is outside those

segmented attention areas or its motion vector is different from the one summarizing the area where it passes through, the moving object will be assigned with high attention (novel motion).

3. PROFILING AND INTERACTING IN MUSEUMS

3.1. Visitors Profiling

Visitors profiling is an important issue for the optimization of the visiting paths and the artworks location in museums or other public places. It is able to provide the designer with valuable information about how interested are the visitors in the different artworks, how well the visitors' flow is optimized, etc...

The long-term attention model can be used here to recognize pre-defined paths. The mean behavior but also the individual blob behavior can be analyzed by comparing it to previously learnt models. Models (taking into account people speed and movement direction for example) can be built and than new visitor's motion can be compared to those models. The difference between the long-term models and the visitors' current path represents the "surprise" or the attention which should be granted to the visitors if their path is different from the one of the previously learnt model. By comparing the mean amount of attention generated by the visitors related to the models, which correspond to the visiting styles, it is possible to see which one of these models is closer to the visitors current behavior (the model which generates the lower attention in respect to the visitor's movement).

This technique enables a dynamic profiling of the current behavior of the visitors in a long term context (previously learnt models). Pre-recorded models (as Veron and Levasseur) can be built based on a general ethnographical study. But models can also be built simply from previous groups of visitors.

If the use of pre-defined models seems to be better from an academic point of view, as those typologies properties have been extensively studied, it seems to be more difficult to have reliable models for complex areas (more complex than simple square rooms with one entrance and one exit). The second approach which is to use models learnt from previous visitors seems more realistic in terms of model learning feasibility in complex museographic environments.

3.2. Adapt Visitors Feedback to their Behavior

More than just an analysis of the expressive and social behavior of visitors, some systems could provide some feedback to visitors and adapt it depending on the visitors' social behavior analysis.

Features extracted from the trajectory (fast/slow, empty space/full space...) could be used to adapt the artworks explanations (if the visitor is fast, the explanations should be short, if the visitor uses too much the empty space in the middle of the room, he should be attracted to the artworks by animations, etc...).

The artworks "personal space" (space close to them) can be used to directly attract people to one painting or to project explanations on the most visited paintings which pushes people to go to the less visited ones in the same room by making some references to them for example.

Finally, some interactions with people could help in visitors' flow management by detect people going in the wrong direction (contrary to the main stream) and provide them with negative feedback or no reward in that case. This could be interesting to avoid potential collisions and social negative aspects of a too crowdie room.

3.3. Foster Social Interaction

Finally, the feedback from the computer could become a mediator in fostering human to human interaction and thus foster the social interaction.

One way to do that is to create an expressive interaction within a group of people in the same place and at the same time which should push to cooperation. This system uses attention in a spatial context. The idea is to have two behaviors inside the group (order or synchronization vs. chaos,

outliers vs. not outliers, occupancy of a given space in majority vs. in minority, etc...). Then one of the two contradictory behaviors should be rewarding, the other not. The reward policy could change dynamically, thus people have to work together to find the interaction law.

A second way to foster social interaction would use again attention in a long-term temporal context to initiate and interaction with a group of people in the same place but not necessarily in the same time. Project models of previous groups could be projected on the ground for example. People should easily visualize their position compared to the previous paths. Colors could indicate directions and intensity could indicate speed for example. Than several possibilities of interaction could rise:

- current people follow the path of the previous visitors
- current people direction is different
- current speed is different
- current people both speed and direction are different

Depending on the application, a feedback which is rewarding or not is then provided to the user. This creates interaction based on the previous visitors behavior but also with other people interacting with the system in the same time and which also provide their behavior as a model for current and future visitors.

4. TEST CASES

4.1. Visitors Profiling

The purpose of this application was the real time dynamic analysis of the visitors' style. As shown in figure 2, the protocol was the visit of a room containing five images displayed in a gallery style. The visit is achieved twice: first by a visitor alone and second by a visitor in presence of other unknown visitors.

The five paintings are chosen after a test using the EyesWeb [4] mobile interface (figure 3): 10 images from the same artist were pre-rated on attractiveness by 10 subjects. RM-ANOVA revealed no significant differences in attractiveness and the

5 most similar paintings were chosen. This is done in order to be sure that there is no outstanding painting which will attract attention more than another. In such a case the visitor behavior may be biased.

By using the long-term models of the four visiting styles from Veron and Levasseur typology, the mean visitor behavior is dynamically analyzed in the case when the visitor is alone (figure 4) or in presence of other visitors (figure 5). On the right side of the image a graph with four colors corresponding to the four visiting styles evolves in time. The mean visiting style is the lower curve. While most of the people performed ant-like styles (green curve) when performing alone, the mean visiting style in a social context (with other visitors) seem to be more a grasshopper (blue curve).

This is due to the behavior of the two others visitors but also to the initial visitor. An interesting finding is that there is not a big

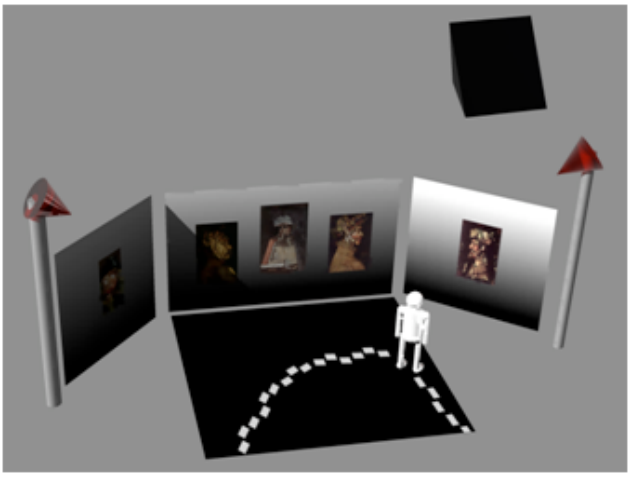

Figure 2: The visitor is in a five paintings museum room

Figure 3: EyesWeb mobile interface used to select the five paintings in the room

Figure 4: A visitor performs alone

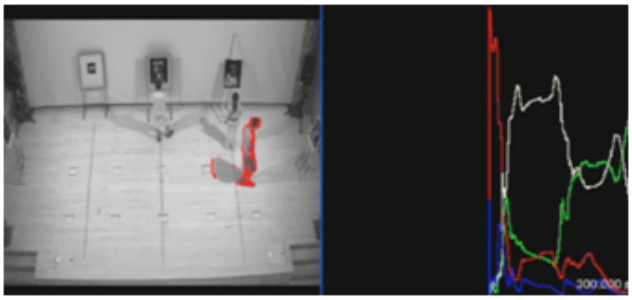

Figure 5: The same visitor as in figure 4 performs with two other people

difference between the main visiting style and the second closest one. This means that the real visiting style may be more a mix of several canonical styles.

Personality tests were applied to each one of the visitors and the big five inventory (ocean) and the Cheek and Buss shyness and sociability scales were computed. The result of this personality test was also very interesting and two categories of people were highlighted: those who change their style when other visitors are present and those who don't. There is a high correlation between the personality tests and people behavior observation: people with little change in visiting style when they are in a social context score significantly higher on sociability (i.e. a personality trait related to approach tendencies) as they do not fear other people. On the contrary, the visitors who drastically change their visiting style to avoid the other people have smaller scores.

4.2. Adapt Visitors Feedback to their Behavior

One of the applications of attention-based interaction in a social context is to manage visitor flow and position. A feedback which aims in positioning the visitor in a precise location should not have a direct focus on positioning the visitor or this one may not want to obey to computer orders. An experiment was achieved to propose "soft feedbacks" which will indirectly push a visitor to find a precise location. The idea is that during the exploration of the stage, a visual feedback (an image) helps the user to locate himself in the target position. If the user is far from the target position, the image is of bad quality while if the visitor is in the correct position, the image is displayed at full resolution (figure 6).

Two features of image degradation were used: the size and blur of the image. The position of the visitor was tracked and two situations were considered: the X position only and the X and Y position. The mean position of the visitors was plotted and this experiment demonstrated that for both X or X and Y measurements, using blur in addition to simple size change help people to locate more precisely. This can be seen in figure 7 where the target position is more focused if the blur is present (columns 2 and 4). While the size change indicates to the visitor the direction that he should follow to reach the target position, the blur provides cues on the exact position of the target. The use of both features seems to be very efficient: all the 14 participants found their location in less than one minute.

Psychological profiles were again applied to the visitors and an interesting result was found: people who perform low on the sociability scale, are more interested in interactions with a computer. This may be due to the fact that more social visitors prefer interactions with real people and not with computers. The result shows that a computer can be used as a mediator to interact with people who should than interact together. These kinds of interactions are likely to attract more introvert people and thus initiate social communication with other visitors through the same system.

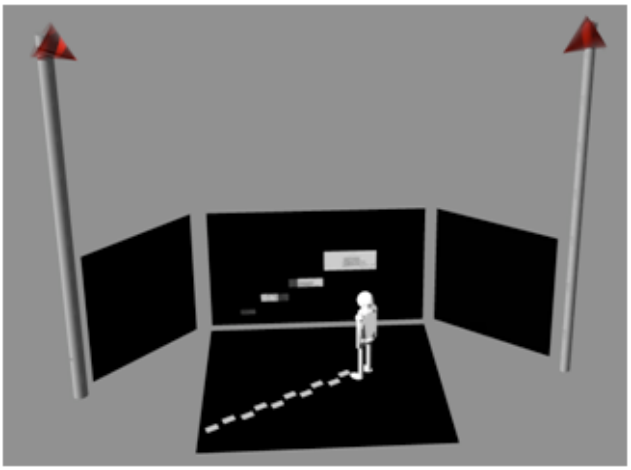

Figure 6: The visitor located in the target position is able to watch a projection to the right interaction

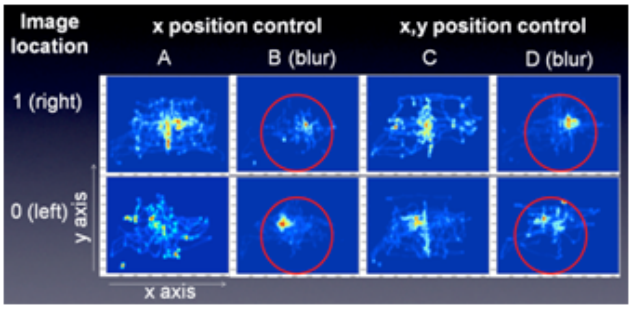

Figure 7: Mean path of the visitors (14 people). First line: target location on the right of the scene. Second line: target location on the left of the scene. First row: only resizing on the X axis used, Second row: resizing and blur on the X axis were used, Third row: only resizing used on both X and Y axis, Forth row: resizing and blur were used on both X and Y axis

4.3. Foster Social Interaction

A last pilot application was achieved in order to be sure that the visitors watched most of the paintings in the museum room. The "personal space" of the painting is used to compute the time that people spent close to the paintings. If the contrast between the paintings' personal spaces occupation is very high (higher than a threshold), the system attention is attracted on the painting which has the most contrasted personal space compared to the others. If this painting is visited a lot, the system will provide the other paintings with animations which should attract the new visitors. If, on the contrary the most contrasted painting is the less visited one, this means that an animation should be directed to this painting (figure 8).

The animations should equilibrate the visit and thus decrease the system overall attention.

Figure 8: A view of the interaction: new visitors entering in the room were encouraged to watch the less visited paintings

5. CONCLUSION

The purpose of this work was to investigate the possibilities of the use of computational attention in museum scenarios in order to analyze behaviors and interact in a social context. Three pilot experiments and a multi-disciplinary approach (engineering and psychology) were used to analyze social interactions and to modify those interactions.

The use of computational attention model in several contexts (spatial and long-term) provided solutions towards a dynamical profiling of visitors and social communication fostering through visitor interactions with the system.

First promising results show that after refinement, those approaches could become applications which could really interest museums and public places in order to optimize their content and to enhance the interest of visitors.

6. ACKNOWLEDGEMENTS

This work has been supported by the numediart research project, funded by Région Wallonne, Belgium (grant N°716631).

This work was achieved during the eNTERFACE'09 Summer Workshop held at the University of Genova, Italy.

7. REFERENCES

7.1. Scientific references

[1] D.E. Berlyne. *Studies in the new experimental aesthetics*. Taylor & Francis, 1974. P.: 92.

[2] A. Camurri, I. Lagerlöf, and G. Volpe. "Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques". In: *International Journal of Human-Computer Studies* 59 (2003). Pp. 213–225. P.: 92.

[3] A. Camurri et al. "Communicating Expressiveness and Affect in Multimodal Interactive Systems". In: *IEEE Multimedia* 12.1 (2005). Pp. 43–53. P.: 92.

[5] G. Kurtenbach and E.A. Hulteen. “The Art of Human-Computer Interface Design”. In: 1992. Chap. Gestures in Human-Computer Communication, pp. 309–317. P.: 92.

[6] M. Mancas. “Attention in Cognitive Systems”. In: vol. 5395. Lecture Notes in Computer Science. Springer, 2009. Chap. Relative influence of bottom-up and top-down attention, pp. 212–226. P.: 92.

[7] M. Mancas. *Computational attention: Towards attentive computers*. Ed. by SIMILAR Network of Excellence. Presses universitaires de Louvain, 2007. ISBN: 9782874630996. P.: 92.

[8] M. Mancas et al. “Real-time motion attention and expressive gesture interfaces”. In: *Journal On Multimodal User Interfaces (JMUI)* (2009). P.: 92.

[9] A. Mehrabian and J.A. Russell. *An approach to environmental psychology*. The MIT Press, 1974. P.: 92.

[10] E. Veron and M. Levasseur. *Ethnographie de l’exposition*. Bibliothèque Publique d’Information, Centre Georges Pompidou, 1983. Pp.: 91, 92.

[11] D. Watson, L.A. Clark, and A. Tellegen. “Development and validation of brief measures of positive and negative affect: The PANAS scales”. In: *Journal of Personality and Social Psychology* 54.6 (1988). Pp. 1063–1070. P.: 92.