Personal Space Modeling for Human-Computer Interaction

Toshitaka Amaoka[1,3], Hamid Laga[2], Suguru Saito[3], and Masayuki Nakajima[3]

[1] Department of Information Science, Meisei University
amaoka@is.meisei-u.ac.jp,
[2] Global Edge Institute, Tokyo Inst. of Technology,
hamid@img.cs.titech.ac.jp,
[3] Graduate School of Information Science and Engineering, Tokyo Inst. of Technology,
{suguru,nakajima}@img.cs.titech.ac.jp

Abstract. In this paper we focus on the Personal Space (PS) as a non-verbal communication concept to build a new Human Computer Interaction. The analysis of people positions with respect to their PS gives an idea on the nature of their relationship. We propose to analyze and model the PS using Computer Vision (CV), and visualize it using Computer Graphics. For this purpose, we define the PS based on four parameters: distance between people, their face orientations, age, and gender. We automatically estimate the first two parameters from image sequences using CV technology, while the two other parameters are set manually. Finally, we calculate the two-dimensional relationship of multiple persons and visualize it as 3D contours in real-time. Our method can sense and visualize invisible and unconscious PS distributions and convey the spatial relationship of users by an intuitive visual representation. The results of this paper can be used for Human Computer Interaction in public spaces.

1 Introduction

Recent advances in Human Computer Interfaces (HCI) brought a new range of commercial products that aim at connecting the physical and the virtual worlds by allowing the user to communicate with the computer in a natural way.

The goal of this paper is to provide a human-computer interaction tool in which the computer detects and tracks the user's states and his relation with other users, then initiates actions based on this knowledge rather than simply responding to user commands. In this work we focus on the concept of Personal Space (PS) [6] which is a non-verbal and a non-contact communication channel. Everyone holds, preserves, updates this space, and reacts when it is violated by another person. In public spaces, for example, people implicitly interact with each other using the space around them. Psychologists also studied the existence of PS in virtual and cyber worlds such as Second Life [16, 3]. They found that many users keep a PS around their avatar and behave in accordance with it in the virtual world. PS is also one factor that makes a virtual agent behave naturally and human like inside the virtual world [4]. It is also an important factor when simulating and analyzing the behavior of people in a crowd.

Based on this concept, we provide a new human-computer interaction framework. In a first step, we propose a mathematical model of the PS based on four parameters: the distance between people, their face orientations, age, and gender. The proposed model can be used for simulating the behavior of virtual agents and also for non-contact Human Computer Interaction. In both cases, the first two parameters can be estimated automatically while the two others are set manually since they are attributes of the personality. In the case of HCI application, we automatically estimate the distances between persons and their face orientations from image sequences using Computer Vision (CV) technology. Based on this model we calculate the two-dimensional relationship of multiple persons and visualize it as 3D contours in real-time.

The remaining parts of the paper are organized as follows; Section 1.1 reviews the related work. Section 1.2 outlines the main contributions of the paper. Section 2 details the mathematical model we propose for modeling the PS. Section 3 describes the system and algorithms we propose for estimating the parameters of the PS. Results are presented and discussed in section 4. We conclude in section 5.

1.1 Related work

Existing Human Computer Interaction (HCI) technologies are mostly based on explicit user inputs acquired through devices such as keyboard, mouse, and joystick. There have been also many attempts to emulate these devices using cameras and speech so that the user will be able to use his hands for other purposes. Examples of such input devices include the camera mouse [1] and voice commands in home environments [13].

There is however no HCI technology that interprets the meaning of distances between people, neither the use of communication through space and distance. This concept is reflected in the notion of Personal Space (PS) which is a non-verbal communication and behavior.

The concept of Personal Space, since its introduction by Edward T.Hall [6, 5] and the discussion by Robert Sommer [12], is studied and applied in many fields; In robotics, the PS is considered as a factor for selecting a communication method between a robot and a human [14]. It can be used for example to model of the intimacy of a robot to other users.

To simulate realistic behavior of virtual agents in virtual worlds, researchers try to apply the rule of real human behavior to the agents. There are many studies for simulating natural conversation and behavior of virtual agents. Rehm et al. [10] focused on the conversation and cultural difference to make natural behavior in virtual world. They treated the "first meeting scenario" to apply their model. Mancini et al. [7] proposed an Embodied Conversational Agents (ECAs) called Greta that are virtual embodied representations of humans that communicate multimodally with the user or other agents through voice, facial expression, gaze, gesture, and body movement. They studied what happens in real life, how people differ in their behavior depending on the personality and situation and build up their ECAs. Furthermore, many researchers study to make natural conversation and behavior to virtual agents as macro-level like group or crowds. McDonald et al. [8] proposed human behavior models (HBMs) that are able to control a single simulated entity (or a single group of simulated entities). However, HBMs developed by different groups are unable to interact with each other. This work mainly treats crowd control.

Quera et al. [9] models the space around an agent as a 2D function that reflects in each spatial location the degree of non-satisfaction of the agent if he moves to that location. The optimal configuration of a crowd of agents emerges from the local interactions. This is the closest model to our work. The non-satisfaction function they propose depends on inter-agent distances only. However, the PS concept is based on many complex rules. Its shape and size are affected by several factors such as gender, age, and social position in addition to distances and face orientations. In this paper, we propose a new model that takes into account all these factors.

1.2 Overview and contributions

We propose a HCI framework based on the concept of PS to create interactive art. The idea is to interpret distances between people and make interaction with it. The proposed method estimates the shape of the personal space by measuring first the user's position and face orientation. Then we can simulate the way people communicate between each other through the interaction of their personal spaces. It generates and varies the contents according to the relationship between users.

The contributions of this paper are three-fold; first we propose a mathematical model of the PS. It is controlled by four parameters: the user's age, gender, position in the 2D space, and the face orientation. Second, we propose to estimate these parameters. User's position and face orientation are estimated automatically using a set of cameras mounted around the operation area. At the current stage the gender and age are set manually by the user. Finally, we visualize the PS in 3D using computer graphics.

The proposed method in this paper enables the detection of the user's mobile territory and his relationship with other. Our targeted application is interactive and collaborative digital art, but results of this work can be applied to modeling the behavior of virtual agents, as well as analyzing people behavior in a crowd.

2 Modeling the Personal Space

Edward T.Hall in his study of human behaviors in public spaces [6] found that every person holds unconsciously a mobile territory surrounding him like bubbles. The violation of this personal space by a tierce person results in an effective reaction depending on the relation of the two persons. This suggests that the

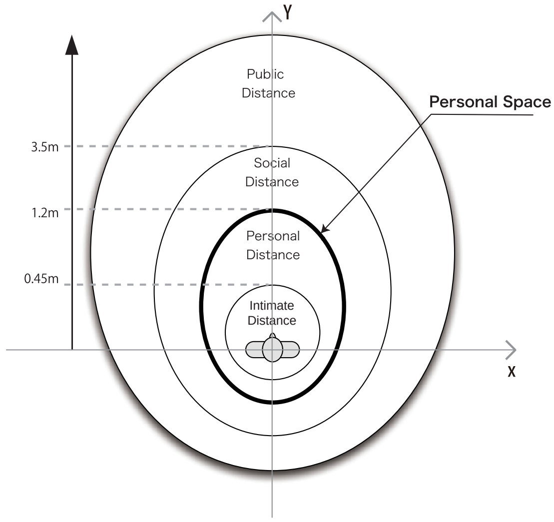

Fig. 1. Definition of the Personal Space. The space around a person is divided into four regions: the intimate distance, the personal distance, the social distance, and the public distance. The figure shows typical sizes of each zone.

Table 1. Standard personal space distances.

	Distance	**Type of relationship**
Intimate distance	$0 - 45$cm	Very intimate relationship
Personal distance	$45 - 120$cm	Friends
Social distance	$1.2 - 3.5$m	Strangers
Public distance	> 3.5m	Public speaking

concept of PS is a non-verbal communication between two or more persons. The personal space as defined by Edward T.Hall and shown in Fig. 1 is composed of four areas: the intimate space, the personal space, the social space and the public space.

The shape of the PS is affected by several parameters. In this paper we consider four of them: gender, age, distance, and face orientation. The relationship between gender and PS is well studied by sociologists [6]. They suggested also that the shape of the PS varies with the face orientation. For example, the PS is twice wider in the front area of a person than in the back and side areas.

2.1 The model of the PS

Given a person P located at coordinates $p(x, y)$ we define a local coordinate system centered at p, with X axis along the face and Y axis along the sight direction as shown in Fig. 1. The personal space around the person P can then be defined as a function Φ_p which has its maximum at p and decreases as we get far from p. This can be represented by a two-dimensional Gaussian function Φ_p of covariance matrix Σ, and centered at p:

$$\Phi_p(q) = e^{-\frac{1}{2}(q-p)^t \Sigma^{-1}(q-p)}. \tag{1}$$

where Σ is a diagonal matrix:

$$\Sigma = \begin{pmatrix} \sigma_{xx}^2 & 0 \\ 0 & \sigma_{yy}^2 \end{pmatrix}. \tag{2}$$

The parameters σ_{xx} and σ_{yy} define the shape of the PS. Considering the fact that the PS is twice wider in front along the sight line than the side (left and right) areas, we define $\sigma_{yy} = 2\sigma_{xx}$.

This model assumes that the shape of the front and back areas of the PS are similar. However, previous studies pointed out that people are more strict regarding their frontal space. Shibuya [11] defines the PS in the front of people as twice larger as the back, left and right ares. We use this definition in our implementation. We build this model by blending two Gaussian functions as follows:

$$\Phi_p(q) = \delta(y_q)\Phi_p^1(q) + (1 - \delta(y_q))\Phi_p^2(q). \tag{3}$$

where $q = (x_q, y_q)^t$, $\delta(y) = 1$ if $y \geq 0$, and 0 otherwise. Φ_p^1 models the frontal area of the person and is defined as a $2D$ Gaussian function of covariance:

$$\Sigma_1 = \begin{pmatrix} \sigma_{xx}^2 & 0 \\ 0 & 4\sigma_{xx}^2 \end{pmatrix}. \tag{4}$$

Φ_p^2 models the back area of the person and is defined as a $2D$ Gaussian function of covariance

$$\Sigma_2 = \begin{pmatrix} \sigma_{xx}^2 & 0 \\ 0 & \sigma_{xx}^2 \end{pmatrix}. \tag{5}$$

Notice that the standard deviation of Φ_p^1 along the Y axis is twice the standard deviation of Φ_p^2 along the same axis. The function δ blends the two functions and therefore it allows to take into account the face orientation. This concept is illustrated in Fig 2. In our implementation, we model the space where people (or agents) are interacting as a 2D plane parallel to the floor plane. We define a unique world coordinate system and encode a the floor plane as a 2D matrix. Every person P_i holds such a matrix (herein after referred as M_i). Each element of the matrix encodes the importance of the corresponding location to the person. The matrix is dynamic and is updated every time step δt. Fig. 1(b) shows the variation of this matrix according to location and face orientation.

In our implementation we define σ_{xx} as the threshold to which a specific zone of the space is violated. For example, to model the intimate space of a standard person we set $\sigma_{xx} = \sigma_0 = 0.45/2 = 0.255m$ as shown in Fig. 1. The figure gives also the standard values of σ_{xx} for different zones of the PS.

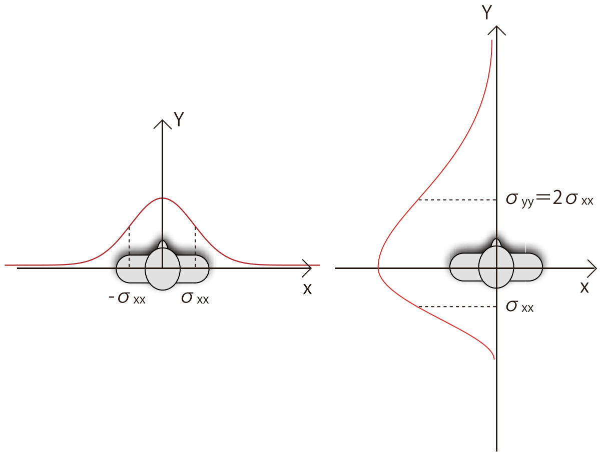

Fig. 2. The Personal Space model based on the face orientation. The PS in the front area of a person is wider that the back and side areas.

2.2 Parameterization of the PS

The personal space does not depend only on the position and face orientation but also on factors related to the person such as age, gender, social position, and character. These personal factors can be included in a function f that affects the value of the standard deviation σ_{xx}. In our implementation we consider only age and gender, hence:

$$\sigma_{xx} = f(\sigma_0, age, gender). \tag{6}$$

In the simplest case, f can be a linear function that scales σ_0 with a factor α reflecting how much a person is kin to protect his intimate space. This model is however not realistic. In our implementation we encode the age and gender dependency as a lookup table where each entry corresponds to the value of σ_{xx} given the age and gender. Table 2 shows an example of such table as defined in [2]. The distances are given in centimeters, and are for Anglo ethnic group. The table also shows that the personal space varies depending on the situation such as being in indoor or outdoor environment. In summary the Personal Space model is parameterized by two types of factors:

- The inter-personal factors, which include distances and face orientations, are embedded inside the parameter σ_{xx}. These two parameters are estimated automatically as will be explained in Section 3.
- The personal parameters such as age, gender, and social position embedded in the function f. These parameters are input manually by the user and are encoded as a lookup table.

In the following we describe the computer vision platform we developed for estimating the inter-personal factors used for simulating the Personal Space.

Table 2. Variation of the frontal area of the personal space with respect to age and gender. The values vary with ethnicity. We considered only Anglo ethnic group in indoor and outdoor locations [2]. The distances are in *cm* and are equivalent to $2\sigma_{xx}$.

Sex combination	Indoor			Outdoor		
	Adult	Teenage	Child	Adult	Teenage	Child
M-M	83	83	63	83	75	62
M-F	71	63	58	79	68	58
F-F	75	62	59	75	73	67

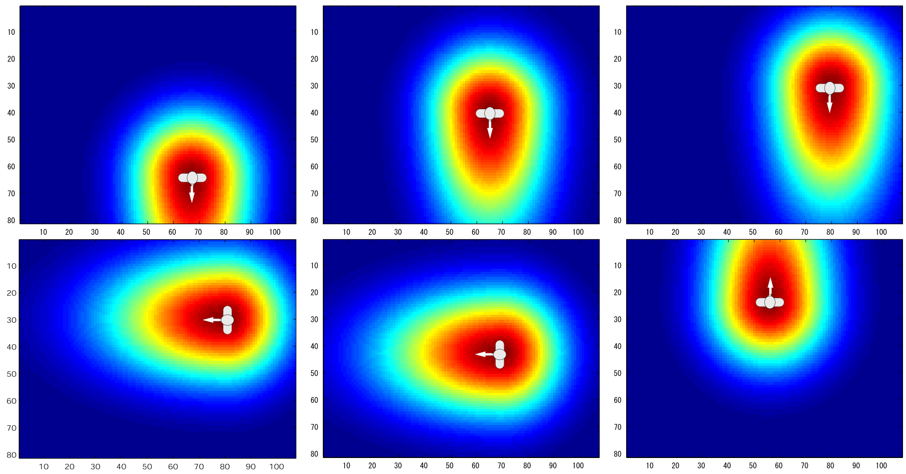

Fig. 3. Illustration of the representation of the PS function as a 2D matrix where each cell encodes the importance of the corresponding location to the agent / person. The arrow indicates the direction of the face.

2.3 Implementation details

In our model, the personal space is computed with respect to the floor (XY) plane. We represent the space as a 2D matrix where each cell (i, j) represents a floor location. Each person keeps and updates every time step Δt his own matrix, herein referred as M_a. It encodes in each cell the importance of the corresponding location which is computed using Equations 3 and 6.

Figure 3 shows how the matrix is updated with respect to the agent (or person) location and face orientation.

3 System

Using the model of Eq. 3 for interaction requires the estimation of the user's position and face orientation. We build a computer vision platform to track the

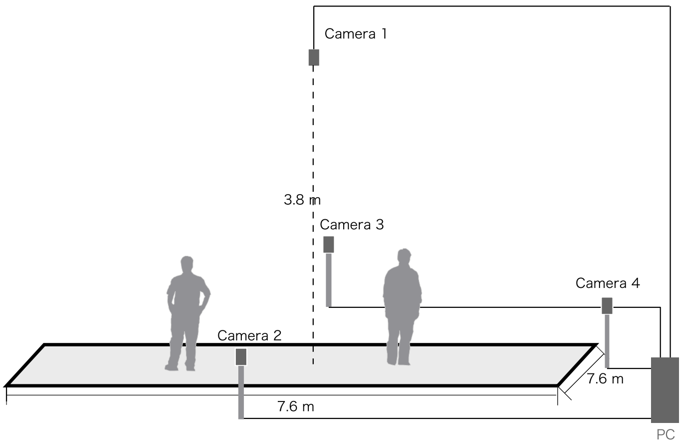

Fig. 4. Overview of the system setup. The top camera (camera 1) is used for people detection. The side cameras (camera 2-4) are used for detecting faces and estimating their orientation.

user's behavior in front of the screen. Using a 3D sensor for efficient localization and tracking is possible and will provide more accurate data. However, 3D sensors such stereo cameras and laser range scanners are expensive. The system, as shown in Figure 4, is composed of four IEEE1394 cameras: one overhead camera for people detection and tracking, and three frontal cameras for detecting faces and estimating their orientation.

3.1 Multiple user tracking

We detect people by detecting and tracking moving blobs in the scene. For efficient detection and tracking, we setup the system in a studio under controlled lighting conditions. The system then operates as follows:

- First, using the overhead camera, we take shots of the empty scene and use them to build a background model by learning a Mixture of Gaussian classifier. Given a new image I, the classifier classifies every pixel into a background or foreground.
- The persons are detected as connected components of the foreground pixels.
- Second, during the operation, we track people over frames using the Mean Shift algorithm.

Estimating distances between people requires recovering the 3D location of the detected persons. To do so, we setup a world coordinate system in such a way

Fig. 5. Example of person and face detection and tracking using four cameras.

that the XY plane corresponds to the floor plane, and the Z plane oriented to the upper direction. The cameras are calibrated and therefore, the exact locations of the persons. We use the face location as the location of the persons. We explain how faces are detected in the next subsection.

3.2 Face detection and orientation estimation

To estimate the face orientation in the 3D space we

- Detect faces from each of the three frontal cameras using the Viola and Jones face detector [15] freely available with the Intel's OpenCV library.
- Assign the detected facial images to each of the persons detected with the overhead camera.

Since the entire camera system is calibrated, i.e., the transformations matrices between the frontal cameras and between each frontal camera and the top one are computed in advance, the faces are assigned to the persons by assigning each face to its closest blob in the overhead camera.

In our current implementation, we are considering only four orientations of the faces corresponding to the angles $0, 90, 180, 270$ degrees. When the a person is facing camera 2 (see Fig. 4), only camera 2 detects the frontal face. Therefore, we assign the angle 0. We apply the same procedure to the other cameras. Camera 4 corresponds to angle 90 and camera 3 corresponds to angle 180. We assume the orientation 270 when no face has been detected by the three frontal cameras.

Figure 5 shows an example where people positions and faces are detected using our system.

4 Results

To visualize the Personal Space of a person we consider the three levels as shown on Fig. 1. We define a PS function for each zone using Equation 3. Position and face orientation are estimated automatically.

The PS function as defined in Equation 3 can be also interpreted as the degree of the user's response to the violation of a zone in his PS. Depending on his relation with the other person, the user activates one of the three functions.

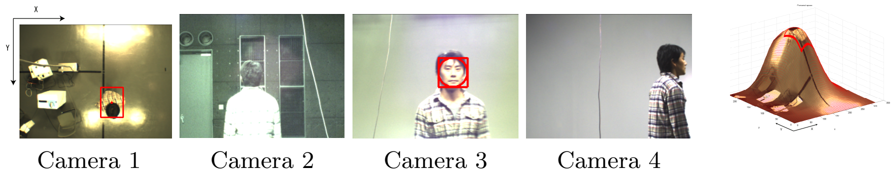

Fig. 6. Visualization of the personal area of the personal space. The user is facing Camera 3 (X axis).

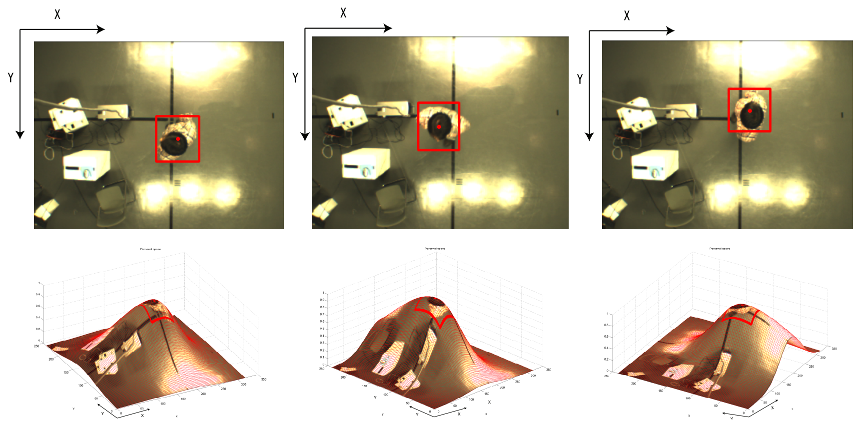

Fig. 7. Visualization of the PS when the face is rotating from camera 2 to camera 4.

To visualize this, we map the image space (taken by the top camera) onto the XY (the floor) plane of the world coordinate system. Since we used only the frontal face detector, we detect three orientations of the face $0, 90$ and 180 degrees. Figure 6 shows an example where the user is facing Camera 3. Notice that camera 3 has detected the face and therefore the area of the PS in front of the user is larger than back and side areas.

Figure 7 shows an example where the user is standing and rotating only his face from one camera to another. Notice how the personal space evolves with the face orientation. This particularly proves the efficiency of our model.

Finally, Figure 8 shows two persons in the same scene and their associated personal space matrices. Each pixel in the matrix encodes how important is the corresponding location to the person. This figure also shows how the PS shape varies according to the face orientation. In our current implementation we considered only the four principal orientations $(0, 90, 180, 270)$. Other orientation can be considered by improving the face detection and orientation estimation part.

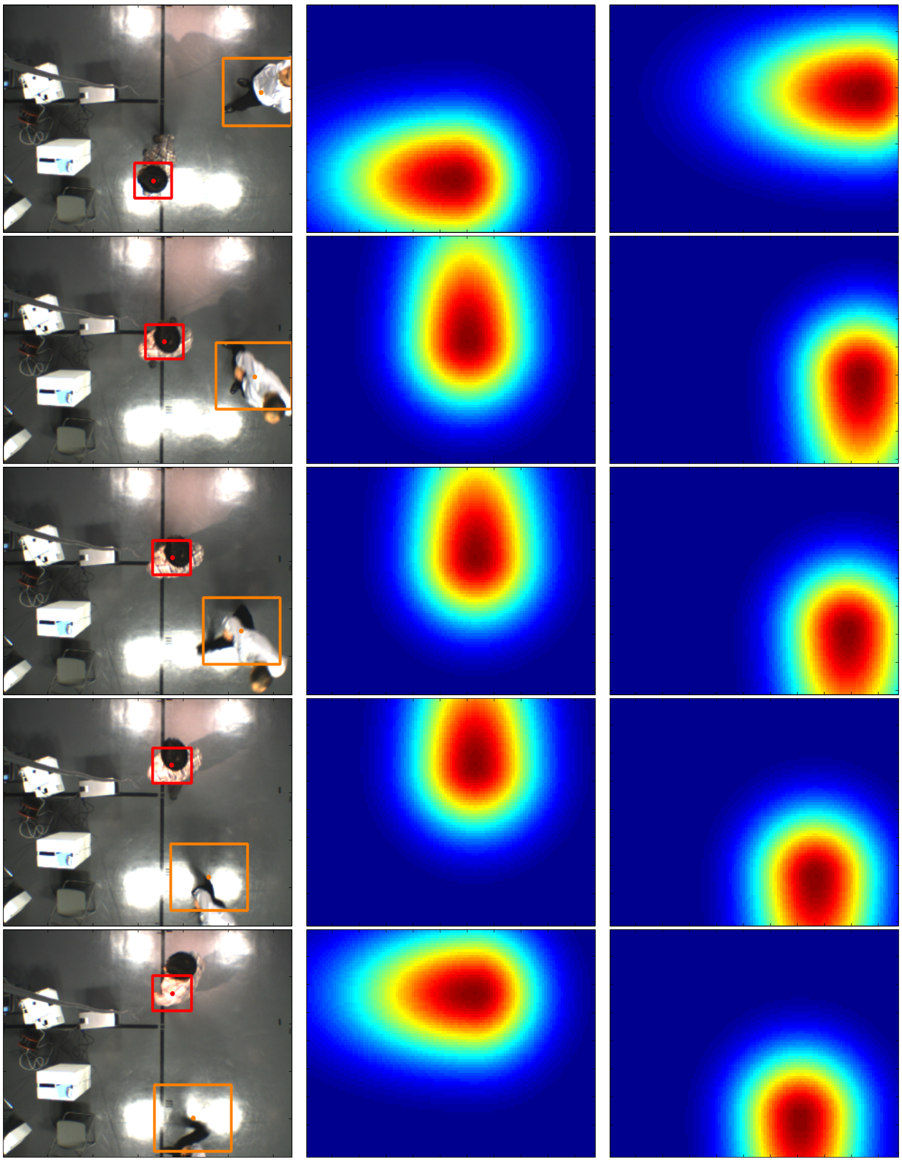

Fig. 8. Variation of the PS with respect to distances and face orientations when two users are interacting. The middle column corresponds to the PS of the person in read.

5 Conclusion

In this paper we have proposed a mathematical model of the personal space. We particularly implemented a method for automatically estimating two parameters of the PS: the people position and their face orientation. As extension we plan in future to automate the estimation of the other parameters such as age and gender. Possible extensions include also the improvement of face detection algorithms for detecting faces at different orientations. We plan also in the future to use this personal space model for HCI and also modeling virtual agents behavior.

Acknowledgement: Hamid Laga is supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology program Promotion of Environmental Improvement for Independence of Young Researchers under the Special Coordination Funds for Promoting Science and Technology.

References

1. Camera mouse. *http://www.cameramouse.org/ (Accessed on November 15, 2008).*
2. J. C. Baxter. Interpersonal spacing in natural settings. *Sociometry*, 33(4):444–456, 1970.
3. D. Friedman, A. Steed, and M. Slater. Spatial social behavior in second life. *Intelligent virtual agents 2007*, pages 285–295, 2007.
4. M. Gillies and M. Slater. Non-verbal communication for correlational characters. *Proceedings of the 8th Annual International Workshop on Presence, London*, September 2005.
5. E. T. Hall. Proxemics. *Current Anthropology*, 9(2-3):83–108, 1968.
6. E. T. Hall. *The Hidden Dimension: Man's use of Space in Public and Private.* Anchor Books, Re-issue, 1990.
7. M. Mancini, B. Hartmann, and C. Pelachaud. Non-verbal behaviors expressivity and their representation. 2004.
8. D. McDonald, R. Lazarus, A. Leung, T. Hussain, G. Bharathy, R. J. Eidelson, N. Pelechano, E. Sandhaus, and B. G. Silverman. Interoperable human behavior models for simulations. 2006.
9. V. QUERA, F. Beltran, A.Solanas, L.Salafranca, and S.Herrando. A dynamic model for inter-agent distances. *In Meyer J A, Berthoz A, Floreano D, Roitblat H L and Wilson S W (Eds.) From animals to animats, The International Society for Adaptive Behavior*, 6:304–313, 2000.
10. M. Rehm, Y. Nakano, E. Andres, and T. Nishida. Culture-specific first meeting encounters between virtual agents. pages 223–236, 2008.
11. S. Shozo. *Comfortable distance between people: Personal Space.* Japan Broadcast Publishing Co., Ltd, 1990.
12. R. Sommer. *Personal Space: The Behavioral Basis of Design.* Prentice Hall Trade, 1969.
13. H. Soronen, M. Turunen, and J. Hakulinen. Voice commands in home environment - a consumer survey - evaluation. September 2008.
14. T. Tasaki, S. Matsumoto, H. Ohba, S. Yamamoto, M. Toda, K. Komatani, T. Ogata, and H. G. Okuno. Dynamic communication of humanoid robot with multiple people based on interaction distance. *Transactions of the Japanese Society of Artificial Intelligence*, 20:209–219, 2005.
15. P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 55(2):137–154, May 2004.
16. N. Yee, J. N. Bailenson, M. Urbanek, F. Chang, and D. Merget. The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *Journal of CyberPsychology and Behavior*, 10:115–121, 2007.