Attentional Model for Perceiving Social Context in Intelligent Environments

Jérôme Maisonnasse, Nicolas Gourier, Oliver Brdiczka, Patrick Reignier
PRIMA, GRAVIR-IMAG
INRIA Rhône-Alpes,
38349 St. Ismier.
France

Abstract. This paper presents a novel approach to detect interaction groups in intelligent environments. To understand human activity, we must identify human actors as well as their interpersonal links. Interaction detection is a good cue to address activity of user groups. An attentional model is derived from gravitational model and cognitive psychology approaches. Whereas determining locally users' focus of attention is a difficult task, this model exploits contextual elements such as position, speed and saliency of objects in the scene to estimate shared attention. The attentional model shows promising results on simulated scenarios where unexpected events occur.

1 Introduction

Human activity recognition is a growing field of research. Recent progress in computer multimodal perception promises new developments in the field of ambient applications and pervasive systems. Such systems aim at offering services by taking into account the current user's activity in a specific situation. In intelligent environments, more and more devices are able of perceiving user activity and proposing appropriate services. Addressing the right user at the right moment is essential. We must detect potential users and their connection while doing an activity. This aspect of human activity is neglected by most studies, in which groups are predefined and unchanging. It appears important to explicit relations between different users and to detect interactions between them. Interaction configuration group is the best detector of activity presence in a scene. Indeed, when a participant is in immediate physical contact with another, they contribute to the same global definition of the situation [5]. Delimiting who is concerned by an activity is a difficult operation. Psychology studies show that human activity is more unexpected than we perceive. Limits between different activities are fuzzy when users share the

Please use the following format when citing this chapter:
Maisonnasse, Jerome, Gourier, Nicolas, Brdiczka, Oliver, Reignier, Patrick, 2006, in IFIP International Federation for Information Processing, Volume 204,Artificial Intelligence Applications and Innovations, eds. Maglogiannis, I., Karpouzis, K., Bramer, M., (Boston: Springer), pp. 171–178

same physical space [7]. Outside laboratory conditions, activity evolves in relation to exte

rnal factors which can not be expected. Identification of the current group configuration of users is necessary to connect activity. The dynamics of group configuration, i.e. the split and merge of interaction groups, allows us to perceive relevant evolution of current activities. Determining user focus of attention is a difficult task. Focus of attention is an intern cognitive task which can not be perceived directly. This perception must be estimated from external observations.

We propose an attentional model to dynamically detect interaction group. An interaction between people occurs when we suppose that they share the same information [11]. A cognitive explanation of how people share information is their ability to product a mutual intelligibility of current situation. Perception of current situation is defined by enabled shared resources in a physical, social and cultural environment, more or less stabilized [10]. In this paper, we model available contextual element in intelligent environment to compute mutual intelligibility. From these results, we analyze focus of attention of users, and detect where interactions take place.

2. Related Work

Based on the idea that social world is organized and understandable in the way action is produced, computer sciences attempt to extract invariant features to describe activity. Most computer vision based research in human activity recognition is focused on data processing issues. Many approaches extract a structured representation of user activity from sensory input data [8]. Visual, acoustic and temporal aspects of activity are concerned. Human activity is cut into a sequence of relevant features from observation in relation to a particular activity. Some approaches are very close to input data. These systems work like a black box learning activity from specific observations. Some systems build higher-level representations of activity. Extracted features are used to match some learned concept representing activity [3]. Relations between detected entities are interpreted as semantic relationships. In the second case, the main issue is to identify entities and concepts describing a specific situation or action. Relation detection then depends on concept and entity recognition. However, interactions between users are an implicit data and almost studies do not consider only one group.

It seems important to explicit interaction between users. A first approach has been applied successfully to speech event detection [2]. It could be completed with other modalities. Psychology offers relevant models to understand how people could interact in relation of contextual element. In order to estimate how attention focus is placed on space, we propose a cognitive model. To compute this model, some relevant and available features are used in intelligent environment. Almost every device could give some information about their internal state and their action in direction of users. For example, when an user is receiving an email or his telephone is ringing, this device will send a message to our system about their actions.

However, understanding what humans do when they execute actions is more difficult. User based approaches are exponentially complex and computationally expensive for this problem. For this reason, we implement an attentional model based on context. This approach requires fewer features from users and objects and is psychologically plausible.

3. Role of Context in production of Mutual Intelligibility

To explain the way in which agents are able to communicate, it is necessary to admit that they share mutual knowledge. Theory of mutual knowledge has a characteristic to produce a regression at infinity. But this theory cannot be integrated into a cognitive explanation of production and comprehension of communicative acts. Sperber and Wilson developed a weaker but empirically more adequate concept, the mutual manifestness. For Sperber and Wilson, "a fact is manifest to an individual at a given time if and only if this individual is able at this time to represent this fact mentally and to accept his representation as being true or probably true [11]. In other words, a fact is manifest when it has the characteristic to be perceptible or deduced by an agent at a given time. A fact can thus be manifest without being known. However, some facts can be more manifest than others. To model this, we associate a degree of salience to each fact. The salience is a function of the perceptual and cognitive capacities of the individual, and of his physical environment. For example, let us suppose that a telephone is ringing in a room where an individual A is sitting at an open window and that at the same time a car is passing in a street. In this case, it will be strongly manifest for A that telephone rang, but less clear that a car passed. Thus, because of the difference of salience between the ringing telephone and the car noise, the fact "telephone is ringing" is more manifest, i.e. has more chance to be perceived or deduced than the fact "a car passed". Sperber and Wilson define the cognitive environment as whole facts which are manifest for a given individual. A shared cognitive environment indicates all the facts which are manifest to several individuals. From the example of telephone by imagining that another individual B is in the same part as A. In this case, by supposing that they have same perceptual capacities, it is manifest for A and B which telephone is ringing. This means simply that they are able to perceive or deduce the same fact, and not that they share a belief, a knowledge, or a representation concerning this fact. The Mutual Cognitive Environment (ECM) indicates a shared cognitive environment in which identity of individuals who have access to this environment is manifest. A and B share a cognitive environment which includes all facts and especially their co-presence. As they share the same environment, they can establish an interaction in relation to their common perception of contextual events. This definition is more precise than Dey's definition of context as "any information that can be used to characterize the situation of entities" [4]. Dey does not precise how user information is selected. We define context as the whole set of objects which are manifest for an individual and capable to modify his interpretation of the situation.

3.1 Focus, Nimbus and Spatial Metaphor

In spatial metaphor, localization contributes to structure interactions between users. By considering their interpersonal distance, in virtual space, users adapt their cooperation situation. Space is inhabited by objects which might represent people, information or other computer artefacts. Cooperation for distant users need to be restored by contextualizing space of work and by giving users a mean to control their interaction. "Objects in space are responsible for controlling interactions on the basis of quantifiable levels of awareness between them. Awareness between objects is manipulated via focus and nimbus, subspaces within which an object chooses to direct either its presence or its attention" [9]. Initially, focus and nimbus are defined as follows:

" - The more an object is within your focus, the more aware you are of it
- The more an object is within your nimbus, the more aware it is of you"[1].

In spatial metaphor, there are objects which manage their awareness by manipulating focus and nimbus subspaces. To define how many objects can interact, we evolve awareness levels from a combination of nimbus and focus configuration. "The level of awareness that A has of object B in medium M is some function of A's focus in M in relation to B's nimbus in M" [1]. Level of awareness defines whether objects may be strongly or weakly aware of each other. This model describes how to quantify level of awareness but not how to compute dynamically focus and nimbus. Indeed, in computer supported co-operative work (CSWC) applications, computing focus and nimbus do not present interest because focus and nimbus are parameters controlled by users as input data. We need a specific model to compute focus direction from contextual elements observation. On the basis of cognitive model, the attractiveness notion of a salient object leads us to another field of research where distance and salience object is useful to understand how objects influence each others.

3.2 Gravitational Model

The first Law of Universal Gravitation has been formulated by Isaac Newton in the 17th Century. Any two objects in the Universe exert gravitational force on each other, with the universal form (1). This force is proportional to the product of their masses and inversely proportional to the square of the separation between the two objects. An example is shown in Figure 1.

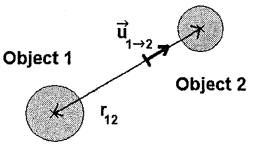

Fig. 1. Object 1 attracts Object 2

$$\vec{F}_{1\to2} = -\vec{F}_{2\to1} = -G\frac{m_1.m_2}{r_{12}^2}\vec{u}_{1\to2} \quad (1)$$

where $\vec{F}_{1\to 2}$ is the gravitational force exerted by object 1 on object 2, $G = 6.67.10^{-11}\ N.m^2.kg^{-2}$ is the universal gravitational constant, m_1 and m_2 are the masses of the two objects, r_{12} is the distance between the two objects and $\vec{u}_{1\to 2}$ is a unitary vector between the two objects. By considering *N* objects in the Universe, the force exerted on each object *i* is equal to the sum of gravitational forces exerted by the *N-1* other objects (2):

$$\vec{F}_{\to i} = \sum_{\substack{j=1 \\ j\neq i}}^{N} \vec{F}_{j\to i} = \sum_{\substack{j=1 \\ j\neq i}}^{N} -G\frac{m_j.m_i}{r_{ij}^2}\vec{u}_{j\to i} \qquad (2)$$

The Fundamental Principle of Dynamics (3) enounces that the derivative of the quantity of movement of an object *i* is equal to the sum of the forces exerted on this object. By supposing that the mass m_i of the object is constant, we can compute the acceleration of this object (4). The acceleration a_i of object *i* stands for the attraction of other objects on object *i*. In particular, it reflects the fact that objects with little masses are more attracted by objects with bigger masses than objects with bigger masses towards objects with little masses.

$$\frac{d(m_i.\vec{v}_i)}{dt} = \vec{F}_{\to i} \quad (3) \qquad \vec{a}_i = \sum_{\substack{j=1 \\ j\neq i}}^{N} -G\frac{m_j}{r_{ij}^2}\vec{u}_{j\to i} \quad (4)$$

In this work, we compute a likelihood interaction as the cue of a shared activity between co-presence users. An attentional model can identify when people share same resources, on the basis of proxemic information and contextual element salience. We have interpreted this cognitive model by transposing some relevant concept from gravitational model.

4. Social awareness as activity detector tool

Focus of a person is defined by attention direction which is the combination of its external and internal factors. External factors of a person are determined by the attraction of the person, objects or artefacts towards its environment. We adapt the gravitational model to simulate persons' attraction towards other persons or objects. Each person or object has a salience *m*. The salience corresponds to the mass in the gravitational model and derivates the concept of nimbus in spatial metaphor. We suppose that salience is invariant, which allows computing the attraction vector of each person towards the people and the objects in the environment using the gravitational model. The salience could be defined on perceptive, social or situation features.

Internal factors of a person are determined by the person's current goal or current activity, regardless of its environment. Cues of internal factor of a person are for example current speed and gaze direction. Internal factors can also be represented by a vector. For the moment, we just take into account the current speed. Both external and internal factors vectors are combined so that the influence of external factors decreases exponentially with the internal awareness, as shown by Figure 2. For our

application, the influence of the attraction becomes negligible when the speed is higher than the top speed v_{max} of human running of 4 m/s. We compute the attention vector as a linear combination of internal and external factors, as in (5).

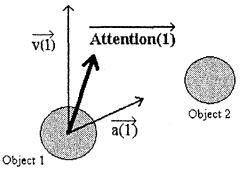

Fig. 2. Attention vector of object 1

$$\overrightarrow{Attention\,(i)} = \lambda \cdot e^{-2\frac{\|\overrightarrow{v(i)}\|}{v_{max}}} \cdot \overrightarrow{a(i)} + \mu \cdot \overrightarrow{v(i)} \quad (5)$$

We consider the interaction area as ellipse constructed as follows; the position of the person is a focus of the ellipse and its addition with the attention vector gives us the center of ellipse. The area of the interaction is called interaction capacity. The interaction capacity is the maximum interaction area which human can act. To calculate the interaction capacity, we consider that two people speaking together at an interpersonal distance of 1.5 meters [6] are in full interaction, and their interaction ellipses recovers fully, as in Figure 6. We define the attention point of a person as the other focus of his interaction ellipse. Beyond a maximal distance of 6 meters, we consider that few social interactions occur, and the great axis reaches its maximum. This prevents us from having too slim ellipse. A fact is salient when it modifies the direction of an interaction ellipse. To determine social interactions and shared activities, we consider ellipses overlaps, as shown in Figures 4 and 5. The use of ellipses reflects the fact that the person stays aware of his surrounding and that the perception field reduces when the attention increases, and that the attention decreases on the opposite direction. In particular, when a person is alone in an empty environment, the attention vector is null, the two foci are equal to his position and his interaction ellipse becomes a circle. This reflect the fact that the person has his attention all around him, as in Figure 3.

Figure 3. Person 0 alone in an empty room

Figure 4. Person 1 close of an salience object 0

Figure 5. Two persons attracted each other.

Figure 6. Two persons in full attraction.

5. Application of the model to social awareness

In this section, we present a simulation which demonstrates the capacities of our attentional model to detect shared activities. Imagine an office environment where three persons (A, B, C) are working at their personal computer (0,1,2). Each person is attracted by his personal task materialized by computer interaction as shown in figure 7. Suddenly, person A starts to speak. The noise produced by his voice is

perceptible for all people. Then B and C, attracted by this perception, move their attention focus from their computer to person A, as in Figure 8. When A stops speaking, a person D enters in the room and begins to speak. All persons move their attention to the newcomer, as in Figure 9. Person D gives his directives and exits the room, when the telephone is ringing. All persons lead their attention to telephone, but persons A and C have more probability to interact in relation to distance which separates them, as shown in Figure 10. The scenario described above illustrate how a system could identify where attention is lead on elements of context. The more attention they share, the stronger is their likelihood to interact.

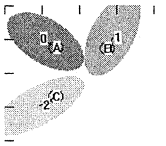

Figure 7. Three persons A, B, C work on their computer 0,1,2

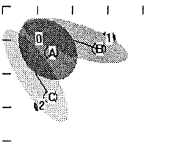

Figure 8. Two persons B, C are attracted by a third person A.

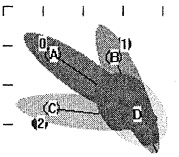

Figure 9. Three persons A, B, C are attracted by a newcomer D who speaks loud

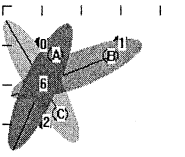

Figure 10. Three persons A, B, C are attracted by the telephone ringing.

Difficulties to use this model come from the choice of parameters for salience for each object. We have developed this model on the basis of a kind of type relation between parameter value and interaction strength. The salience follows an exponential scale. Other parameters could be used for the specification of a particular object salience. For these scenarios, we use parameters indicated in Figure 11.

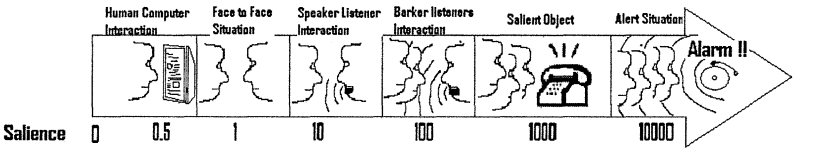

Fig. 3. Salience of right entities on the left entities

6. Conclusion

We propose a new perception tool to recognize human interaction in intelligent environments. On the basis of proxemic information and salience of contextual objects, we attempt to explicit social relationship to determine whether interactions

occur or not. When one interaction is detected, we suppose that users have possibility to define or modify their current activity by articulating their respective actions in relation to new appearing fact. Here, we insist on the importance of context in activity production, and explicit mechanism to improve recognition performance.

This approach is based on context evaluation. Human activity is not represented by a sequence of sensory features by describing an entity at t moment, but integrates the relation between entities and the whole elements present in context which could affect his activity. The main difficulty is to identify objects and evaluate their salience. When the object is electronic, it can give information about its status and an a priori salience can be affected. The task of identifying users' status is more difficult, except if users are equipped of sensors. We need to plug this model to a complete architecture to evaluate real gain for human activity recognition. However, simulated scenarios give some interesting results, and presume to detect activities and theirs unexpected evolutions.

7. References

1. Benford, S.D, and Fahlèn, L. E. : A Spatial Model of Interaction in Large Virtual Environments, Proc. Third European Conference on CSCW (ECSCW'93), Milano, Italy, Kluwer (1993).
2. Brdiczka, O., Maisonnasse J., Reignier P. : Automatic detection of Interaction Groups, Proceedings of ICMI 2005, (2005) 32-36.
3. Crowley, J.L. and Reignier, P. : Dynamic Composition of Process Federations for Context Aware Perception of Human Activity, International Conference on Integration of Knowledge Intensive Multi-Agent Systems, KIMAS'03 (2003) .
4. Dey, A., Abowd, G., and Salber, D. : A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Human-Computer Interaction (2001) 16 (2-4).
5. Goffman, E. : The presentation of self in every day life. Doubleday, New York: Doubleday (1959) (French traduction, 1973, Paris: Éditions Minuit).
6. Hall, E.,T. : The Hidden Dimension. Garden City, N.Y.: Doubleday (1966).
7. Heath, C., & Luff, P. : Collaboration and control: Crisis management and multimedia technology in London Underground line control rooms. CSCW Journal (1992) Volume 1(1), 69-94.
8. McCowan, I., Gatica-Perez, D., Bengio, S., Lathoud, G., Barnard, M., and Zhang, D.: Automatic Analysis of Multimodal Group Actions in Meeting, IEEE Transactions on Pattern Analysis and Machine Intelligence (2004).
9. Rodden, T.: Populating the Application: A Model of Awareness for Cooperative Applications, Computer Supported Cooperative Work '96, Cambridge MA USA, (1996).
10. Salembier, P., Theureau, J., Zouinar, M., & Vermersh, P. : Action/Cognition située et assistance à la coopération. In J. Charlet (Ed.), Ingénierie des connaissances IC2001. Grenoble: PUG (2001).
11. Sperber, D., & Wilson, D.: Relevance. Communication and cognition (2nd edition ed.). Oxford: Basil Blackwell (première édition 1986, Cambridge, MA: Harvard University Press) (2001).