Assessment of Broader Attention-focus Perspective

Eyal Dim
Information Systems Department
The University of Haifa
Mount Carmel, Haifa, 31905, Israel
dimeyal@bezeqint.net

Tsvi Kuflik
Information Systems Department
The University of Haifa
Mount Carmel, Haifa, 31905, Israel
tsvikak@is.haifa.ac.il

ABSTRACT

Mobile personal devices nowadays are equipped with variety of sensors that can be used for context-related measurements such as location, orientation, motion and more. These measurements may allow systems to fuse, reason and abstract context related data into meaningful contextual cues, and use them for personalized services delivery. An example for such a scenario may be two people in front of a product in a shopping center or in front of an exhibit in a museum. They may pay "social attention" to each other, or "object attention" to the exhibit or product. Capturing their attention-focus may enable better adaptation of services to their needs. "Social attention" and "object attention" are broader perspectives in comparison to knowing which specific object attracted the attention, or how exactly a person attracted the momentary attention of the other. This work shows how even a small number of simple sensors combined with a relevant model may enable the assessment of broader attention-focus perspective.

Author Keywords

Ubiquitous computing, context aware computing, social signal processing, user model, group model.

ACM Classification Keywords

H1.2. Models and principles: user/machine systems. H5.2. Information interfaces and presentation: user interfaces.

General Terms

Experimentation, Human Factors, Measurement.

INTRODUCTION

Imagine two people in a shopping center, standing in front of a product. At any given moment, they may be facing the product, or facing each other. In the first case, there is high probability that they are interested in the product. In the second case, there is high probability that they have a social interest in each other. Their general social context is "a group of two people", and their general location context is "being at a shopping center". However, their attention may shift from "socializing" to "interested in a product" or to both. These are broader attention-focus perspectives in comparison to "I listen to you now" or "I see this specific object now". Understanding the focus of attention may enable a system to better adapt its services to these users. For example it may suggest product information when appropriate; avoid interference when people are involved in social attention; or suggest amusement to a person who is not paying attention to the product, enabling the other person to focus on the product. A second example is robotics, where a robot serving a person should adapt to the person's needs, and identify if the person is addressing the robot, or whether it is the right time for the robot to attract the person's attention. A third example is virtual presence, where a real world situation is represented in the virtual world. It would help if the avatars in the virtual world, representing the real world people, would be able to represent cues that reflect the people's attention. (e.g. if a person is facing another person then we, humans, assume that the probability that they are focused on each other increases).

James [1890] presented attention as follows: "Everyone knows what attention is. It is the taking possession by the mind, in a clear and vivid form of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others....". Technologies to measure attention-focus, or cues for attention-focus, vary in their validity, accuracy, precision and maturity. Examples of such technologies are brain computer interfaces using EEG [Hamadicharef et al. 2009] and methods presented for example by Stiefelhagen et al. [2002], such as: eye-gaze detection, eyeball movement detection; face recognition, head pose and head orientation detection, audio cues and speech recognition. Most of the methods require devices, which either require effort for setup and calibration or are inconvenient to carry or non-portable at all. Therefore, a lot of research is done in a stationary environment (e.g. meeting room, standing in front of a product shelf, sitting in front of a computer). The proposed paradigm is to assess a *broader* attention-focus perspective (e.g. a *narrow* attention focus perspective would be knowing which specific object one is looking at, and a *broad* perspective would be understanding that the focus of attention is a product or a group mate). The broader perspective may enable using simpler measurement sensors, available on nowadays devices such as mobile smart phones. Implementation of such broader perspective requires modeling of the user, the group and the context. Low-level measurements as well as the inferred knowledge may be included in a User and Group Model (UM). The UM data is used to adapt to the user/s needs, as surveyed by Kray & Baus [2001]. The UM uses detailed values stored into its properties. These values are measured and inferred, starting with low-level signals, and continuing with the process of fusion and abstraction of these low-level signals. The UM is usually a general model, yet it is based on details. Organization of such details may enable to use them within low-level building blocks that can serve as the bricks of the UM building. This bottom up approach requires investigation of such building blocks as if they were mini-models. This study presents a mini-model of the attention of two people near an object in a museum. Analyzing such a scenario would require to know: (i) Who is in the group (ii) Where the group is (i.e. a museum) (iii) Are the two people in close proximity to each other? (iv) Is each one of them in close proximity to a specific exhibit? (v) Are they facing the exhibit or each other? (vii) Are they talking? The answer to each of the above questions may involve specific sensor data, fusion of data sources and inference. This work proposes a mini-model, of broad attention-focus, which may be combined with a UM while collecting and fusing data, measured by simple sensors.

BACKGROUND AND RELATED WORK

Llinas [2010] described the standard process of data (or information) fusion from several sources. This process prepares the data by common referencing; associates it by generating hypothesis, assessing them, and selecting the preferred hypothesis; estimates the current state and predicts the future states; and finally, exports the fused data for the use of another user or process. Context awareness is one implementation that requires information and data fusion. Dey and Abowd [2000] defined context as: "any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves". They further defined context-aware system: "A system is context aware if it uses context to provide relevant information and / or services to the user, where relevancy depends on the user's task". Dey [2010] presented tools for building context-aware applications. He described two main concepts: (i) Distributed information - the application is responsible for sensor data fusion and usage, while the sensor's data is accessed through a Widget, that prepares the sensor's data for the application. (ii) Centralized repository of data – the "blackboard" approach. The blackboard method may also be used by having several distributed blackboards [Corkill, 1991].

Ubiquitous computing (pervasive computing, ambient intelligence) paradigm [Weiser, 1991] contributes another point of view. It is aimed at utilization of nonintrusive networks of sensors and machine learning techniques for context detection and adaptation [Bettini, 2010] as information is distributed through the networks.

Data gathered following the above-mentioned approaches may be used for user and group modeling. A Group Model represents both the specific model of each individual, as well as the group as a whole. It refers to group dimensions that may be based on sociological theories such as communication, conflicts handling, controversy and more [Pizzutilo et al., 2005]. As a result, a system used by a group can adapt its services to the group using a Group Model [Dim and Kuflik, 2010].

A MINI-MODEL CASE STUDY

Our case study involves two people and an exhibit (object) in a museum environment (applicable also in a shopping center). The people's focus of attention may be used for adaptation and personalization. For example, if a person is interested in socializing with a group-mate it wouldn't make sense offering this person a personal device, which may cause isolation (e.g. earphones). There may be several *Attention-categories*: (i) attention to an object (such as product, label, presentation, commercial or exhibit); (ii) social attention - attention to a group member; (iii) integrated attention - a combination of the first two categories, such as in the case of conversation about an object (museum exhibit) that is being observed; (iv) navigation attention ('watch your step'); and (v) intrinsic attention – attention to internal thoughts, mind wandering, or to external distractions (e.g. noise, color, other people, etc.). The question that rises is how a system can distinguish between these categories of attention by using inference and sensor measurements. The identification of categories (i) through (iii) may be supported by fusion and reasoning about low-level sensor data, such as proximity, orientation, sensor detection sector, and voice detection. Attention-category (iv) may be supported by the change in orientation towards the direction of walk. Attention-category (v) may require additional data from sensors such as brain activity detectors. (iv) and (v) are beyond the scope of this paper. The following discussion focuses on the first three Attention-categories. It starts with a description of low level sensors in our experimental environment; continues with a theoretical analysis of the "two people and an object" scenario; and shows how basic measurements such as proximity and orientation can be integrated to assess an increase or a decrease in focus on the Attention-categories (i) through (iii). Finally, we discuss the ability to measure and reason about users attention in the situation of "two people and an object" scenario. The model may be easily expanded to more than two group members, if it is used to represent subgroups of two people from the group.

The PIL Project Sensor-suite

In the framework of the PIL[1] project [Kuflik et al., 2011] a Radio Frequency (RF) based positioning system was installed at the Hecht Museum[2]. It utilizes a wireless sensor network designed and produced by Tretec[3], composed of small (matchbox size) mobile RF tags called Blinds (Figure

[1] http://www.cri.haifa.ac.il/connections/pil/, accessed Dec 14th, 2011
[2] http://mushecht.haifa.ac.il/info_eng.aspx, accessed Dec 14th, 2011
[3] http://www.3tec.it/, accessed Dec 14th, 2011

1, left) and RF to TCP Gateways (Figure 1 right) that transfer the data reported by the Blinds over a local area network to the PIL server. Each Blind and gateway has a unique identifier. The Blind can be carried by a person (Figure 1 middle), or located near an exhibit (as a stationary beacon).

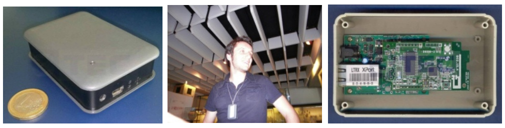

Figure 1. Mobile positioning device and gateway

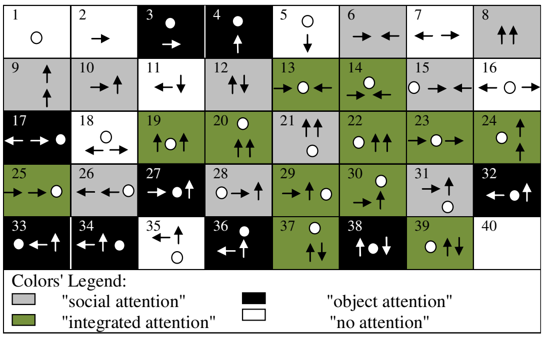

Table 1. Combinations and attention categories of "two people and an object" based on proximity and orientation

The Blind sensor has several important features, including: (i) Measuring proximity among Blinds, which allows to reason about the proximity among visitors. (ii) Detecting voice level and voice activity (due to privacy considerations voice is not recorded), a feature that can be used to assess the level of conversation among visitors as well as their proximity (in this scenario people may have conversation only if they are close to each other). (iii) Detecting orientation of visitors, using embedded magnetometers, enabling assessments such as whether visitors are facing each other, the exhibit, or standing back to back. Finally, (iv) detecting motion by using embedded accelerometers.

Mini-model of the "Two People and an Object" Scenario

The scenario we analyze is a common one: two people coming to a museum or a shopping center. They may be together or apart. They may also be next to an exhibit or a product. The property in focus is the *Attention-categories*: (i) "social attention" that represents high probability that the two people pay attention to each other (one person to the other, or mutual attention); (ii) "object attention" that represents high probability of attention of at least one of the people to the object; (iii) "integrated attention" that represents a combination of (i) and (ii); and finally, (iv) "no attention" where the probability is low that there is attention. These define the mini-model output.

We performed a theoretical analysis of the above scenario, taking into account the proximity and the orientation of the entities (person1, person2 and the object), where orientation resolution was limited to 90 degrees. The results are presented in Table 1. Each person is represented by an arrow symbol pointing at one of the four main orientations (left "←", forward "↑", right "→", and backward "↓"). The object, "○", may be positioned in front, behind, on each side, and between the two people. The analysis yielded 40 unique combinations based on proximity and orientation. Cell 1of table 1 represents an object alone, cell 2 represents a person alone, cells 3 through 5 represent a single person and an object, cells 6 through 12 represent two people (no object), cells 13 through 39 represent two people and an object, and cell 40 represent the case where there is neither a person nor an object. The above theoretical cases are further refined into the Attention-categories, as shown by the color-coded cells in Table 1. White cells represent "no attention" cases, when people are not in close proximity with each other (or with the object) or face away. Gray cells represent "social attention" when the two people are in close proximity and at least one person is facing the other (or standing side by side with the other). Black cells represent "object attention" when at least one person is in close proximity to the object and facing it. Green cells represent "integrated attention", where conditions for both "social attention" and "object attention" exist.

Measuring Attention-categories

The mini-model measurements are collected by three Blind sensors: one located next to the object (exhibit), and the other two are carried by the two people. The mini-model processing starts with the "identification" stage that associates an entity ID (person or object) with a Blind sensor ID. It continues with "proximity assessment" stage that uses proximity between Blinds and associated IDs to assess the entities' proximity. In parallel, the "orientation assessment" stage computes mutual orientation between Blinds. Finally, the "attention refinement" stage uses the entities' proximity and the mutual orientation to generate the Attention-category as the output of the mini-model.

There are some insights in regards to the measurements: (i) Although the orientation sensor within the Blind is accurate (better than 10 degrees), some of the results would remain ambiguous. For example, in combinations 8 and 9 both Blinds will have the same orientation while proximity is detected. Another example relates to 7 and 16 that cannot be differentiated because they have the same orientations, while proximity is not detected in both cases (because of turning away from the other person and / or to the object). (ii) As for proximity, if there is no proximity report, it does not mean that there is no proximity. Occlusions and interferences may cause false negative proximity reports. In this case the reasoning process assumes "no information", until enough messages are gathered. (iii) There are also false positive proximity reports that result from reflections and multi-path of the RF signal. Therefore, a low threshold of 10% of recent proximity messages was set for accepting proximity status. (iv) It is interesting to note that there are cases where the proximity detection limitation becomes an advantage. For example, in combination 7, although in the real world the two people are close to each other, their Blinds do not detect proximity (because the signal is blocked by their bodies, standing back to back), and they are considered having "no attention", which is true in reality because they stand back to back.

Evaluation

We evaluated the above theoretically analyzed model, in order to demonstrate how it can be used in practice. The evaluation was conducted in a natural museum environment with phenomena of RF interferences, occlusion and reflections. Two people were wearing the Blinds on their chest, having the Blind main detection sector in front of them. The object's Blind was positioned on a table oriented at a specific predefined direction. The object's Blind had free 360 degrees transmission and detection; while due to body shielding, detection of the Blinds carried by visitors degraded when facing away from each other or from the object. 36 of the 39 meaningful combinations in Table 1 were tested (in three cases we failed to collect data, and case 40 is meaningless). The tests took 1840 seconds, collecting 4671 Blind messages. Each case lasted for at least 30 seconds, with an average of 51 seconds per case, and an average of 130 Blinds' messages per case. Table 2 presents the results of the Attention-categories detection. Each of the categories (lines in the table) is detected at high recall and precision percentages as presented in table 3. As expected, as a result of the real world environment characteristics, that include shielding, interferences, occlusions and reflections, there are phenomena of false negative (where there are no measurements, even though they are expected) as well as false positive (where there are erroneous measurements), hence the attention assessment becomes probabilistic. Future work may use additional sensors such as voice detection to determine if there is an additional increase in the social attention probability.

Attention	Recall (%)	Precision (%)
No	86	75
Social	86	100
Object	79	83
Integrated	67	67

Table 2. Two people and an object mini-model results

CONCLUSIONS

The study presented the concept of measurement of broader attention-focus perspectives, and demonstrated it with the "two people and an object" scenario. It showed that this mini-model can encapsulate measurements and reasoning process, while exporting only the required output: Attention-category. The mini-model provides a generic framework to bridge the gap between sensors' low-level measurements and the abstract values needed for supporting personalization and adaptation to the momentary need of the small group: socialization, paying attention to the objects in front of them (e.g. a product in a shop or an exhibit in a museum), or both. In general, we showed a general approach where simple measurements from a small number of sensors may enable the inference of abstract UM attributes, given the right process and model.

REFERENCES

1. Bettini, C., Brdiczkab, O., Henricksenc, K., Indulskad, J., Nicklase, D., Ranganathanf, A., Riboni, D. A survey of Context Modeling and Reasoning Techniques. Pervasive and Mobile Computing, 6(2), 161-180 (2010).
2. Corkill, D. D. Blackboard Systems. In AI Expert, Vol.. 6(9), 40-47, September, (1991).
3. Dim, E., and Kuflik, T. Social Signal Processing: Detecting Small Groups Interaction in Leisure Activity. In The Proceeding of the 14th International Conference on Intelligent User Interfaces, IUI'2010, Hong Kong, China, 309-312, (2010).
4. Dey, A. K., and Abowd G. D. Towards a better understanding of context and context-awareness. In CHI'2000 Workshop on the What, Who, Where, When, and How of Context-Awareness, (2000).
5. Dey, A. K. Context-aware Computing. In Krumm, J. (edt.), Ubiquitous Computing Fundamentals, Taylor and Francis Group LLC, (2010).
6. Hamadicharef, B, Zhang , H., Guan, C., Keng, W., Phua, KS., Tee, KP., and Ang, K. Learning EEG-based Spectral-Spatial Patterns for Attention Level Measurement. In: Proceedings of ISCAS2009, Taipei, (2009).
7. James, W. The Principles of Psychology. Harvard UP, Cambridge, MA, (1890/1981).
8. Kray, C., and Baus, Y. A Survey of Mobile Guides. In conjunction with The Fifth International Symposium on Human Computer Interaction with Mobile Devices and Services, (2001).
9. Kuflik T., Lanir J., Dim E., Wecker A. J., Corra' M., Zancanaro M.and Stock O. Indoor Positioning: Challenges and Solutions for Indoor Cultural Heritage Sites. In (IUI 2011), 13-16 February 2011, Palo Alto, CA, USA (2011).
10. Llinas, J. A Survey and Analysis of Frameworks and Framework Issues for Information Fusion Applications. In Hybrid Artificial Intelligence Systems, LNCS, Vol. (6076/2010), 14-23, (2010).
11. Pizzutilo, S., De Carolis, B., Cozzolongo, G., and Ambruoso, F. Group Modeling in a Public Space: Methods Techniques and Experiences. In Proceedings of WSEAS AIC 05, Malta, Spain, 175-180, (2005).
12. Stiefelhagen, R., Yang, J., Waibel, A. Modeling Focus of Attention for Meeting Indexing Based on Multiple Cues. In IEEE Transactions on Neural Networks, Vol. 13, No. 4, (2002).
13. Weiser, M. The computer for the 21st century. In Scientific American, 265(3), 94–104 (1991).