Simulated Crowd: Towards a Synthetic Culture for Engaging a Learner in Culture-dependent Nonverbal Interaction

Sutasinee Thovuttikul, Divesh Lala, Hiroki Ohashi,
Shogo Okada, Yoshimasa Ohmoto, Toyoaki Nishida
Graduate School of Informatics, Kyoto University
Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan
{thovutti@ii.ist, lala@ii.ist., ohashi@ii.ist., okada_s@, ohmoto@, nishida@}i.kyoto-u.ac.jp

ABSTRACT
Difficulties in living in a different culture are caused by different patterns of thinking, feeling and potential actions. A good way to experience cultural immersion is to walk in a crowd. This paper proposes a simulated crowd as a novel tool for allowing people to practice culture-specific nonverbal communication behaviors. We present a conceptual framework of a simulated crowd using an immersive interactive environment. We discuss technical challenges concerning a simulated crowd, including real-time eye gaze recognition in a dynamic moving situation, sensing of nonverbal behaviors using multiple range sensors, and behavior generation based on novel temporal data mining algorithms.

Author Keywords
Synthetic Culture, Embodied Conversational agents, Multi-modal interaction, learning by mimicking

ACM Classification Keywords
I.2.11 [Artificial intelligence] Distributed Artificial Intelligence –Artificial agents, multi-agent system.

INTRODUCTION
In daily life, humans use nonverbal communication to expand the meaning of verbal feelings. Eye movement plays a critical role. It helps the actor to both effectively and efficiently communicate her/his intention to the partner without speaking.

As pointed out by Knapp and Hall [1], nonverbal behaviors depend on numerous factors, including age, status, gender, role, context, emotion, mood, personality, and cultural background. When one enters another culture, s/he might feel uneasy from time to time until s/he has learned the patterns of thinking, feeling and potential actions shared in that culture [2].

2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction.
February 13, 2011, Palo Alto, California, USA.

Practice is often better than precept. A good way to practice culture is to walk in a crowd in which one need to pay attention to culture-specific nonverbal signals by which people cooperate or negotiate with each other to avoid collision and achieve the respective goals.

This paper proposes a simulated crowd as a novel tool for allowing people to practice culture-specific nonverbal communication behaviors. A simulated crowd can be characterized as one possible instantiation of synthetic culture [3], which specifies an artificial environment habited by synthetic agents behaving according to a parameterized norm. To realize a simulated crowd, we introduce an immersive interactive environment. We discuss real-time eye gaze recognition in a dynamic moving situation, sensing of nonverbal behaviors using multiple range sensors, and behavior generation based on novel temporal data mining algorithms as key technical contributions.

In what follows, we first discuss how nonverbal communication depends on culture. Then we introduce the idea of synthetic culture and simulated crowd as its instantiation. We then present the realization of simulated crowd using an immersive interaction environment. Finally, we discuss technical challenges together with preliminary results.

NONVERBAL COMMUNICATION
Nonverbal communication "involves those nonverbal stimuli in a communication setting that are generated by both the source (speaker) and his or her use of the environment and that have potential message value for the source or receiver (listener)" [4].

In principle, sending and receiving of messages occur in a variety of ways without the use of verbal codes, i.e., touch eye contact (eye gaze), volume, vocal nuance, proximity, gestures, facial expression, pause (silence), intonation, posture and smell. There are two basic categories of non-verbal language, nonverbal messages produced by the body and nonverbal messages produced by the broad setting (time, space, silence).

Eyes not only function to see things but also serve as a stimulus to be seen by others. Eye gaze cues are used to make inferences about others' cognitive activity, including their focus of attention, intention, desire, and knowledge about the current state of affairs. Kleinke [5] summarized how gaze functions (a) provide information, (b) regulate interaction, (c) express intimacy, (d) exercise social control, and (e) facilitate service and task goals. One of the important roles of eye gaze is as a social cue guiding attention. People share information about intentions and future actions using eye gaze.

Communicative functions implying future actions are particularly important in situations where a person is moving with others, such as playing sports, collaborating in physical tasks or walking in a crowd. The eye gaze manages the directions and the timing that people move and provides their intentions as to whether each person accepts the movement or not. The most significant point is that we can communicate that information in a short time by using eye gaze and other information. The function of eye gaze plays an important role in establishing quick and smooth interaction among the people.

Since nonverbal communication is rather polysemic, integrating multiple modalities is mandatory for robust interpretation of nonverbal behaviors. Morency et al [6] constructed a probabilistic model which predicts when to give listener backchannel using not only eye-gaze of the speaker but also the prosody and spoken words. They achieved better prediction of visual backchannel cues than previous studies by enabling the system to automatically select the relevant features. Huang et al [7] also used multimodal cues to realize a quiz agent that is attentive to its users' situation. They used audio and visual information to estimate the atmosphere of the interaction and who among the users is leading the conversation. They changed the agent's behavior according to those barometers. We believe the multi-modal information is very useful for our setting as well because we obviously use not only eye gaze but also some other cues to walk through the real crowd without conflicts. To infer the situation of the leaner, we propose using multi-modal information such as eye gaze, hand gesture, head direction, and torso direction.

Maia et al [8] devised an experiment to study the impact of eye gaze of humanoid avatars in conversation. They set four conditions: video, audio-only, random-gaze avatar and informed-gaze avatar. The four experiments compared the impact between avatar and non-avatar communication conditions. The informed-gaze avatar can relate the agent's eye gaze and conversation better than the random-gaze avatar. They concluded that the best result of the conversation is on agent who can relate the gaze to the conversation.

CULTURAL DIFFERENCES IN NONVERBAL COMMUNICATION

Different cultures have different type of communicative expressions, where verbal and nonverbal behavior is different. For instance, in some cultures standing close shows a familiar relationship between people. On the other hand, other cultures consider standing close as being rude. There are also differences in gestures, touching behaviors and eye gaze patterns [1].

Fehr and Exline [9] suggested that gaze is associated with dominance, power, or aggression. For example, when people are in the crowd or an elevator, they can adjust their personal space if they agree or limit eye contact [10]. Eye gaze has different meaning in each culture. Argyle et al [11] reported that English and Italians use direct eye gaze during conversation, whereas for Japanese and Chinese from Hong Kong seldom use direct eye contact in conversation. Watson [12] classified two categories, “contact culture” and “noncontact culture”. Contact cultures such as Arabs, Latin Americans and Southern Europeans engage in more gaze, touch and close interpersonal distance during conversation than noncontact cultures.

The behavior of a human crowd is diverse. For example, the differences of Thai and Japanese culture in a crowd are walking speed, posture and eye contact. There is evidence that walking speed differs across cultures[13] Preliminary through observation, the posture of walking of Japanese people seems more gentle and graceful than Thai people, and eye contact of Japanese people is focused more toward their destination more than Thai people.

In the physical age, the learner has had to go to a different country for observation, recognition, and imitation of cultural behavior, if they would like to practice nonverbal culture. In the information age, however, we can use a simulated system to represent the virtual world where human communication takes place in a different place, culture and language. The virtual world allows us to design various types of synthetic culture which are introduced in the next section.

Christopher [14] presented the theory of mind which is used as a computing process of agent behavior. This theory controls agent interaction behavior, their eye, head and body directions, locomotion and greeting gestures. The direction detector detects the agent's eye gaze, i.e., direct or averted. The intentionality detector decides if the goal object is the desired one. Theory is the module store of the mental state of the agent. The result of this module is based on the interaction between the agents. In our approach, we use cultural behavior as agent behavior. The selection of suitable agent reaction behavior is for future work. The proper selection of agent behavior is based on cultural.

SYNTHETIC CULTURE

The idea of synthetic culture has been put forward by Hofstede [3] and is described as “role profiles for enacting dimensions of national culture”. Hofstede [2] defines five dimension of national culture based on then aspect of each culture when measure relative with other cultures.

Dimension 1 Power distance. High power distance cultures believe the power in institution is distribute unequally.

Dimension 2 Individualism versus collectivism. This represents the difference between people. The collectivism culture feels about in-group or out-group.

Dimension 3 Masculinity versus femininity or achievement oriented versus cooperation oriented. This dimension describes how gender influences roles. In high femininity cultures both genders are assume to be cooperation oriented.

Dimension 4 Uncertainty avoidance. Weak uncertainty avoidance cultures believe "what is different is curious", but strong uncertainty avoidance cultures tend to think "what is different thing is curious is dangerous".

Dimension 5 Long-term versus short-term orientation. Long-term orientation cultures are driven by future-orientation (perseverance and thrift), while short-term orientation refers to cultures that are driven by past and present-orientation (respect for tradition).

Synthetic culture "does not exist in reality", but only exists in a gaming or training context. Hofstede himself gave ten such profiles of synthetic cultures, which could be of use in simulation and games.
The ten Synthetic-Culture Profiles [3] extend from the five dimension of national culture.
Dimension 1 Power Distance: *Hipow* is high power distance culture. *Lopow* is low power distance.
Dimension 2: *Indiv* is highly individualistic. *Collec* is extremely collectivity.
Dimension 3: *Achievor* is highly achievement orientation. *Caror* is highly cooperation orientation.
Dimension 4: *Uncavo* is strong uncertainty avoidance. *Unctol* is weak uncertainty avoidance.
Dimension 5: *Lotor* is absolutely long-term orientation and *Shotor* is very short-term orientation.

The variety characters are considered in the ten Synthetic-Culture Profiles and eye gaze is an important expression of many intentions. For instance, the expression of interest in *Hipow* is positive and animated, with no eye contact but the expression of interesting *Lopow* is animated, with eye contact and interjections. It is clear that the way to express the same intention is different in a contrary synthetic culture.

The concept of synthetic cultures is useful for the analysis of different cultures and intercultural awareness. In particular, our goal is to create agents which exhibit cultural attributes. Although infusing agents with these characteristics is difficult, synthetic cultures give us a method to achieve this by not having to recreate complex social norms, experiences or behavior that has been built throughout many generations. Using eye gaze is just one of the initial steps which we can take to achieve agents that possess cultural attributes. For example, we may find that avoidance behavior can be expressed through the lack of eye contact. In this case, a synthetic culture can be created which uses eye gaze in a way that can be recognized by the learner. By the same token, agents with a synthetic culture may recognize eye gaze from the learner and react to it accordingly.

SIMULATED CROWD AS AN INSTANTIATION OF SYNTHIC CULTURE

The simulated crowd is an instantiation of synthetic culture in which the learner can gain virtual experiences with a culture through interaction with synthetic characters that behave according to the parametric model of the culture. It allows the learner to interact with characters that embody various kinds of the behavior in the given culture. Figure 1 shows a person who is interested in nonverbal behavior in a different culture in the synthetic crowd. From the information above, we can create scenarios and role plays to drive learners to understand the differences of people from other cultures. In figure 1, the learner walks through the crowd among many people with varying cultures. Due to learner observation of nonverbal behavior, s/he can interact with people in the synthetic crowd.

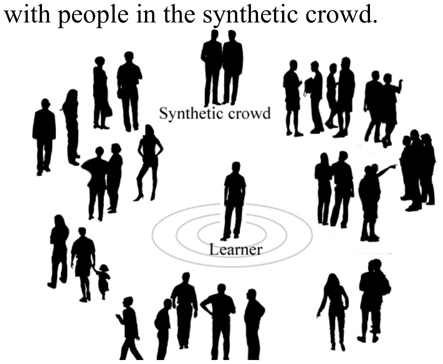


Figure 1. Synthetic Crowd and learner.

Klaus Dorfmueller-Ulhaas [15] developed an immersive 3D game consist of a 3D projection screen with shutter-glasses, a 3D sound system and an optical tracking system. Their system created a crowd simulation for a 3D game. The player controls his upper body and avoids the characters in crowd. The 3D technique is an ideal interaction for a player. s/he feel not only width and height but also depth on the screen. From this research we know this type of environment is important for a learning system.

We plan to set the environment to enclose the learner like s/he is walking among people in the crowd. We use the learner's head direction as the future walking direction of learner. When the learner walks, s/he feels like they are walking through real crowd. We set various actions for agent behavior. The learner can observe the agent behavior for learning and recognizing the cultural behavior in a crowd or interacting with the agent such as walking close to, looking at, waving their hand to the agent.

Figure 2 shows our implementation of a simulated crowd using an immersive display system. The learner is among the virtual crowd. The player walks through the crowd for play the game. This research is a good example of walking in crowd. For our approach we discuss a including nonverbal behavior from a variety of cultures into an agent. The learner can then practice interaction with various cultural agents.

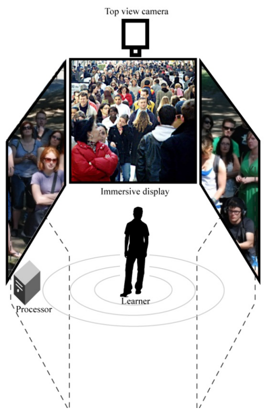
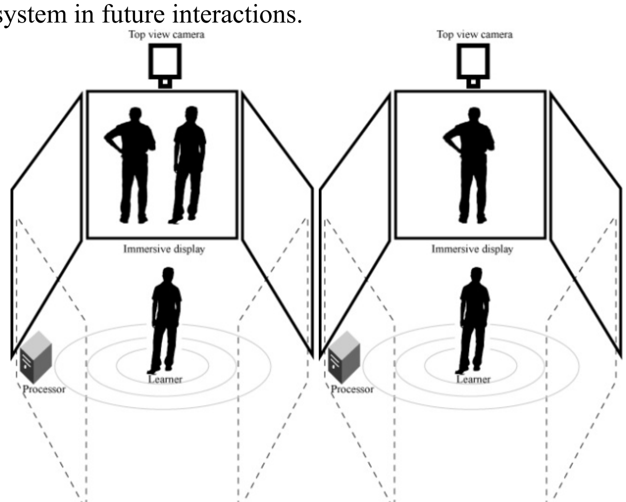


Figure 2. Implementation of a simulated crowd using an immersive display

One may or may not display an avatar of the learner on the screen. In the avatar mode, both the agent and avatar represented learner are displayed on the screen as figure 3(a). In contrast, there is only agent on the screen in non-avatar mode as figure 3(b). From the learner's viewpoint, non-avatar mode represents a directly mutual the gaze between learner and agent (on the screen). Moreover, user distance is directly detected from the length between user and screen. On the other hand, in avatar mode the user sees both the avatar and the agent. The learner interprets the distance between the avatar and the agent on the screen. The learner's gaze need to be mapped as the avatar behavior and the learner is expected to interpret the gaze interaction of the avatar and the agent. In this approach we selected the non-avatar be proper in the initial step. However each approach has its own advantages and disadvantages, we will decide the best solution for the system in future interactions.



(a) (b)

Figure 3. show avatar and the agent on the screen (a) with-avatar (b) without-avatar Mode

We focus on several scenarios that are possible for investigating eye-gaze and the simulation of crowds. The following is one such scenario as the simple act of walking past a person coming from the opposite direction. For illustrate the scenario in figure 2-4, the avatar is a man on the left (in figure 2) and the agent is a woman in the right (in the figure 2). We show both the avatar and the agent on the screen because the interaction between them is mapped in the immersive environment.

Step 1: In figure 4, the agent recognizes the learner walking towards them and can see that they are going to collide if they keep going along their current paths. It reasons that somebody will have to change direction if the goals of both parties are to be achieved.

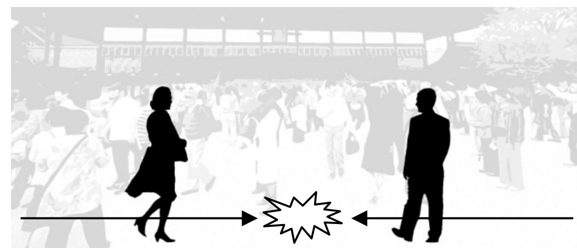


Figure 4. Step 1: recognizing the situation

Step 2: The agent observes the learner and recognizes via eye gaze that the learner cannot see the agent. In figure 5, the learner looks down and cannot see the agent, causing the agent to change his walking direction.

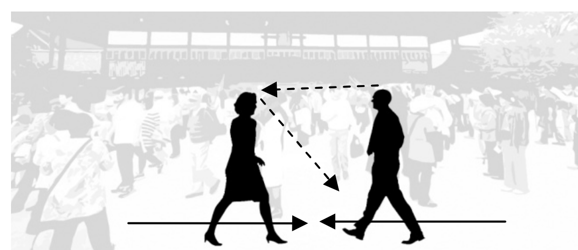


Figure 5. Step 2: Checking each other

Step 3: The agent subtly motions with their body that they are changing direction, while checking the gaze of the learner to see if they recognize their action. On the other hand, the agent may also recognize the subtle movement of the learner and use eye gaze to acknowledge their intention as the figure 6.

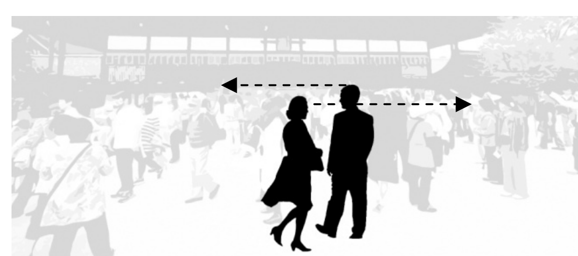


Figure 6. Step 3: Collision avoidance by changing motion

Step 4: Figure 7 is when the agent is sure that collision can be avoided, both parties' goals can be achieved, and the agent continues, reverting to its higher level goal.

Figure 7. Step 4: Confirmation of the safety

The above scenario involves activities that we take for granted in everyday life. Humans instinctively manage to achieve this many times in a day, through subtle non-verbal interactions and eye gaze. These gestures must be able to be identified quickly and an appropriate action performed, in order for agents to react accordingly in those situations.

Eye gaze and behavior differ across cultures using this scenario. We can use the virtual environment to create a number of similar scenarios in which eye-gaze coordinates human behavior.

Simulated crowd in immersive interaction environment

We are building a system that implements a simulated crowd using an immersive interaction environment. It permits the learner to achieve a goal by using eye gaze and multiple actions of nonverbal behavior. Our main focus is to create agents that can exhibit cultural behavior by characterizing synthetic cultural attributes.

The immersive interaction environment consists of 7 immersive displays, 4 range sensors and 3 top-down cameras. The immersive environment is shown in Figure 7. The immersive display screens the simulated crowd around the learner. There are 7 immersive displays for represent in 315 degrees. We use a regular octagon less one display for use as the immersive environment entrance. When we use the scenario for training, the learner has a 315 degree view of the scenario of the immersive crowd (Figure 8).

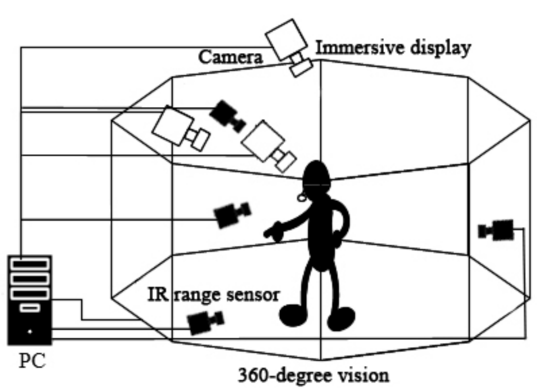


Figure 8. The immersive environment

The range sensors and top-down cameras are used for capturing learner behavior e.g. the direction of the head (as the eye gaze direction), walking speed and position of the learner in the immersive environment. We apply this behavior as an input of the nonverbal communication behavior learning system. The learner can interact with the system in 315-degree as well. Range sensors are the optimal sensing devices for our environment because they give us distance information in real time without any complicated calculation. This enables us to easily distinguish the leaner who stands in the cylindrical display from other people who are possibly shown on the display. Contact-type sensors are powerful but inadequate for our environment since they possibly prevent the leaner from acting freely. Among various contactless sensors, optical motion sensors will allow us to easily capture the motion of people. However, they cannot properly get data when the markers are hidden by something. In order to deal with this problem, we must place many cameras and try to observe the person from various points. We, however, cannot use optical motion sensors in this study because we try to capture the motion of a leaner, who is in a cylindrical display and the places for putting cameras are limited in such an environment.

Figure 9 represents the system architecture that we plan to implement for simulated crowd. The system is designed to recognize head gesture, torso, eye movement, and hand gesture and generate the behaviors of multiple synthetic characters based on the game script. A machine learning system (temporal data mining) is used to extract patterns from the log of the user-agent interactions to generate an action model of the synthetic characters. The simulated crowd has not been implemented but we have started to develop some techniques for supporting the simulated crowd learning system. Ohmoto et al implemented head gesture recognition and torso recognition [16]. This approach focused on eye movement recognition for interact ion with an agent and part of a learning system.

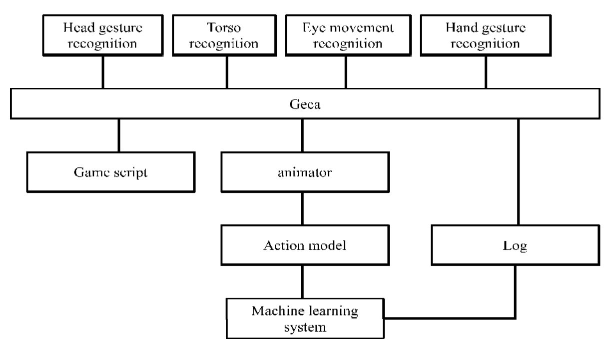


Figure 9. The system diagram

We now discuss three technical problems that we have to address in order to realize the idea mentioned above: real-time eye gaze recognition in a dynamic moving situation, sensing of nonverbal behaviors using multiple range sensors, and behavior generation based on novel temporal data mining algorithms.

REAL-TIME GAZE ESTIMATION IN MOVING SITUATION SUCH AS WALKING IN A CROWD.

When humans walk, they usually look towards the front. However, humans are not always gazing at the direction because peripheral vision is often used to recognize their surroundings, especially when they are moving in a crowd. While humans are walking in a crowd, they have to communicate with each other in following manner; they recognize whether they have eye contact, they imply their future walking directions (future action) using their eye gaze, and they give their acceptance if necessary. Mutual eye gaze occurs when human are looking at each other. Normally, mutual eye gaze helps human organize interaction [17]. In a crowd, sometimes when mutual eye gaze occurs, humans cannot predict the future walking direction of each other because they are seeing each other and cannot recognize the other eye direction which cues the future walking direction. We often avoid collisions by quickly looking in at the future direction. This point is one challenge of our research for developing the simulated crowd.

When learner walk in the simulated crowd, the system has to capture learner behaviors such as eye gaze, head gesture, hand gesture and torso movement. While the learners are walking in the simulated crowd, the main challenge is to develop a method recognizing the object which the learner is actually watching in real-time.

There are many eye gaze recognition techniques [18-21]. Some of this research uses the front of the face as the input to gaze recognition. There are many process to recognize gaze e.g. face detection, eye detection, and eye gaze analysis.

The users of many gaze direction measuring systems have to wear some devices or fix their head in a small region for measurement. These conditions prevent natural communication. On the other hand, the precise measuring of gaze direction by using image processing is a time-consuming process.

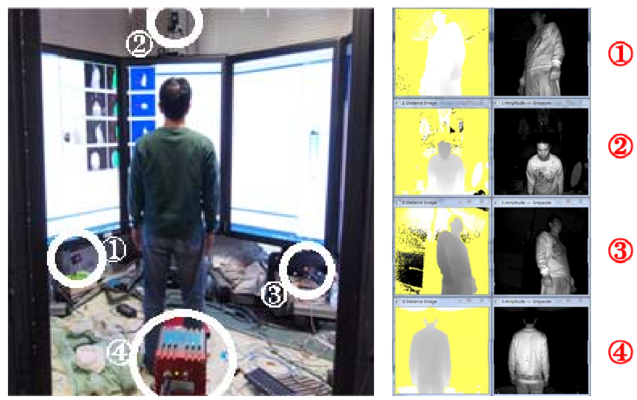
The recognition of precise gaze direction is less important than recognizing whether the agent is watching the learner or whether the learner is watching the agent in a simulated crowd. Therefore, we roughly approximate gaze direction by head direction by taking the time to integrate information such as the learner's eye movements, the agent's gaze direction and the timing of their movements.

Our prototype [16] measuring system can detect the learner's head direction and eye movements in 20 frames per second at least. The prototype system spends most of the time detecting head direction. The body motion measuring system can already detect head direction. The implementation of eye movement detection is left for future research

REAL-TIME ESTIMATION OF THE LEARNER'S 3D MOVEMENT

The estimation of leaner behavior will be calculated from the learner's 3D movement. Learner movements are captured from multi range sensors. Movement data is interpreted for the input of the learning system.

There are two main difficulties in achieving this. One is that since the learner can exhibit various behaviors in the situation, some parts of the learner's body can often be hidden by other parts, and this prevents us from accurate tracking. The other is that the tracking process usually requires a good deal of computational time, and it is difficult to develop a tracking algorithm which works in a reasonable computational time.



(a) (b)

Figure 10. (a) The configuration of the 4 range sensors. (b) Distance image (left) and amplitude image(right) measured by each range sensor in Figure 9. (a)

We need to realize accurate and robust sensing of the learner's motion by using multiple range sensors. We can deal with the problem of occlusion by setting multiple range sensors at mutually complementary position and integrating them. Our configuration of the 4 range sensors is shown in Figure 10 (a), and distance and amplitude images measured by each range sensor are shown in Figure 10 (b). In order to track the learner's motion, we use the 3D human body model which consists of 7 body parts; torso, head, upper arms, lower arms, and lower abdomens (see Figure 11). These parts are approximated by hemisphere cylinders or elliptic cylinders. We update the rotation matrices and translation vectors of each body parts to fit the learner's posture at each frame.

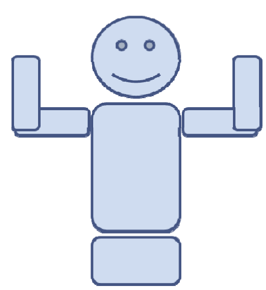


Figure 11. The 3D human body model we use.

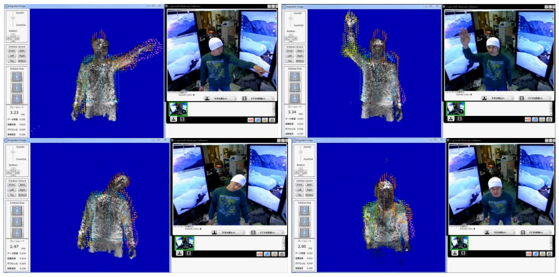


Figure 12. Some examples of recognized body parts

BEHAVIOR GENERATION BASED ON MACHINE LEARNING APPROACH

Agents have to generate appropriate actions (behavior) corresponding to the user's behavior. To generate appropriate actions, we have to model the correlation between the user's multimodal nonverbal pattern and the behavior of the agent. We call this the correlation interaction rule. The numbers of nonverbal patterns which each user uses are unknown because the behavior of users changes depending on the situation and task. Thus it is difficult to define the model prior to execution.

Mohammad and Nishida proposed a novel unsupervised learning technique [22, 23] for discovering the interaction rule from interaction data which is obtained from sensors. It allows for modeling of the user's gesture command, with the robot's action corresponding to the gesture and their association [22]. A constrained motif discovery algorithm elicits gestures and action patterns from continuous time-series data.

Okada et al proposed an incremental clustering approach: Hidden Markov Model based Self-Organizing Incremental Neural Network (HB-SOINN) [23]. HB-SOINN is a hybrid approach which integrates a self-organizing incremental neural network (SOINN), a tool for incremental clustering, and HMM which is used for modeling of time-series data. HB-SOINN has markedly improved the clustering performance over that of traditional clustering methods. However, HB-SOINN cannot be applied for continuous time-series data.

We plan to realize an incremental motif discovery algorithm by integrating CMD and HB-SOINN for discovering the interaction rule from multimodal interaction data which is obtained from sensors. Figure 13 shows the procedure of the machine learning approach.

The main idea of our approach for behavior generation is to learn the interaction model in three main steps which in we plan to realize (observing, learning and acting) the same manner as the approach which is proposed in [22]. The inputs to our system are the learner's behavior (nonverbal patterns) and the agent's action sequence; both are continuous multi-modal time series data. Each modal time-series data is multidimensional. The learning is accomplished in three main phases. First, during the discovery phase the input streams are segmented into meaningful primitive patterns and converted into a discrete integer sequence. Second during the association phase, the discrete integer sequence is analyzed to find associations and correlations between learner and agent behaviors and this information is used to build the behavior generation model. The discovery phase is done using a novel constrained motif discovery algorithm based on HB-SOINN [23].

The association phase is done using simple Bayesian network induction. The controller generation phase is done by modeling actions using Hidden Markov Models. We take into account the correlation between learner and agent behaviors. To implement the incremental multimodal motif discovery, we need a method for matching between multimodal patterns and a clustering approach for multimodal patterns. We plan to implement a new kernel for calculating the similarity between multimodal patterns.

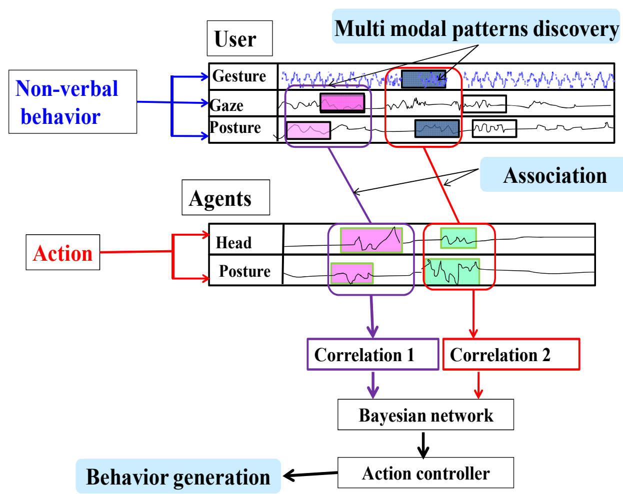


Figure 13. The procedure of proposed machine learning approach

CONCLUSION

This paper presents our research concept for developing nonverbal communication in a cultural learning system. When the learner walks through the simulated crowd, s/he will observe the agents in different culture behavior and interact with agent in real-time. The scenarios are created for helping the learner understand the culture. Eye gaze is important nonverbal behavior because humans usually use eye gaze for explanation or emotion. For eye gaze estimation we plan to analyze head direction that we have implemented as our prototype but for this approach real-time eye estimation is required. This issue is our challenge. We will use real-time estimation of the learner's movement to recognize the torso movement. Currently we have developed human body detection. The challenge of this process each identifying part of the learner body. The last technique described in this paper is behavior generation based on a machine learning approach; this technique generates an appropriate agent behavior corresponding to the learner's behavior. The learning system requires more technique required but for the initial step we propose only the technique that we have been implementing.

REFERENCES

1. Knapp, M.L., and J.A. Hall. Nonverbal Communication in Human Interaction. Fort Worth, TX: Harcourt Brace Jovanovich College Publisher, 1992
2. Greert, H., Gert, J. H., and Michael, M., Cultures and Organizations 3rd edition, Mc Graw Hill, 2010.
3. Hofstede, G., and Pedersen, P. Synthetic Cultures: Intercultural Learning Through Simulation Games, 1999. *Simulation Gaming,* 30:415. Sage Publishing
4. LARRY, A. S., Communication Between Cultures 7th edition, Wadsworth, 2010.
5. Kleinke, C.L., Gaze and Eye Contact: A Research Review, 1986.
6. Morency, L.P., de Kok, I., and Gratch, J. Predicting listener backchannels: A probabilistic multimodal approach. Intelligent virtual agents: 8th international conference, IVA 2008 5208 (September 2008), 176–190
7. Huang, H.H., Furukawa, T., Ohashi, H., Cerekovic, A., Pandzic, I.S., Nakano, Y., and Nishida, T., How Multiple Concurrent Users React to a Quiz Agent Attentive to the Dynamics of Their Game Participation. In van der Hoek, Kaminka, Lesperance, Luck, Sen, eds.: Proceedings of 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010), Toronto, Canada (May10-14, 2010), 1281–1288
8. Maia, G., Mel, S., Simon, B., and Martina, A. S., The Impact of Eye Gaze on Communication using Humanoid Avatarsz. In Proceeding CHI '01 Proceedings of the SIGCHI conference on Human factors in computing systems. (2001), 309-316.
9. Fehr, B. J., and Exline, R. V., Social visual interactions: A conceptual and literature review. In A. W. Siegman & S. Feldstein (Eds.), *Nonverbal behavior and communication* (Vol. 2nd). Hillsdale, NJ: Lawrence Erlbaum. (1987), 225-326
10. Scheflen, A.E., Body language and the social order. Englewood Cliffs, NJ: Prentice-Hall, 1972.
11. Argyle, M., Henderson, M., Bond, M., Iizuka, Y., and Contarello, A. Cross-Cultural Variations in Relationship Rules. International Journal of Psychology 21, 1 (1986), 287-315.
12. Watson, O.M., Proxemic behavior: A cross-cultural study. Mouton De Gruyter, 1970.
13. Geert, H., Culture's consequences: comparing values, behaviors, institutions, and organizations across nations 2nd revised edition, SAGE, 2001
14. Christopher, P., A Perceptually-Based Theory of Mind for Agent Interaction Initiation. International Journal of Humanoid Robotics © World Scientific Publishing Company, (June 8, 2006).
15. Klaus, D.-U., Dennis, E., Oliver, G., Nicolas, S., Volker, W., and Elisabeth, A., An Immersive Game - Augsburg Cityrun, In Perception and Interactive Technologies (PIT), ser. LNAI 4021, E. Andrè et. al., Eds., Springer-Verlag Berlin Heidelberg (2006), 201-204.
16. Ohmoto Y., Takahashi A., Ohashi H., and Nishida T., Capture and Express Behavior Environment (CEBE) for realizing enculturating human-agent interaction. Culture and Computing, Lecture Notes in Computer Science, Volume 6259/2010,(2010), 41-54
17. Argyle, M. Bodily Communication 2nd edition. London: Methuen, 1988.
18. Helton, M.P., Anna, M.G.G., and Adriao, D.D.N., Image Processing for Eye Detection and Classification of the Gaze Direction. Neural Networks, 2009. IJCNN 2009, page 2475 – 2480, 14-19 June 2009.
19. Ba, L. N., Eye Gaze Tracking. Computing and Communication Technologies, 2009. RIVF '09, page 1 – 4, 13-17 July 2009.
20. Cadavid, S., Mahoor, M.H., Messinger, D.S., and Cohn, J.F., Automated classification of gaze direction using spectral regression and support vector machine, Affective Computing and Intelligent Interaction and Workshops, 2009., page 1-6, 10-12 Sept. 2009.
21. Kristiina, J., Nishida, M., and Seiichi, Y., On Eye-gaze and Turn-taking. IUI 2010 Workshop: Eye Gaze in Intelligent Human Machine Interaction, February 7, 2010.
22. Mohammad, Y, Nishida T., and Okada, S., Unsupervised Simultaneous Learning of Gestures, Actions and their Associations for Human-Robot Interaction, IEEE International Conference on Intelligent RObots and Systems (IROS2009) 2009
23. Okada, S., and Nishida, T. Incremental clustering of gesture patterns based on Self-Organizing Incremental Neural Network. IEEE International Joint Conference on Neural Networks (IJCNN2009), (2009)
24. Lala, D., and Nishida, T., A Layered Construct-Based Model to Analyze Trust in Culture Contexts. (submitted)