speak some predefined sentences such as greeting the user thus limiting the capability of natural conversations.

The EU project, STAR [5] aims to enhance service and training in real factory environments using virtual life-like humans. Cavazza et al. [6] places a live video avatar of a real person into a mixed reality setting and interact with a digital storytelling system with body gestures and language commands. These two systems rely on predefined animation/video such that it is served merely as a storytelling system.

Barakonyi et al. [7] presented the AR Lego system where an animated agent demonstrates assembly of toy blocks in a real environment. The system provides step-by-step assembly instructions as to which block to mount next and how to verify whether the user is at the correct stage in the construction. The agent autonomously shows the proper demonstration by animated behaviors over real toy blocks. The user is restricted to interact with the system through the use of a tracked PDA which acts as a multi-purpose interaction device.

Barakonyi et al. [8] later proposed an autonomous agent system for augmented reality games: monkey bridge. This application describes an AR game in which two users play against each other by building a bridge consisting of real and virtual elements for their monster-like avatars across a virtual ocean. The shape and spatial distribution of the bridge elements influence the decisions of the AR Puppet-based behavior engine that autonomously generates actions for character motion and affective behavior without explicit scripting or user instructions.

The aforementioned systems introduced interesting forms of human machine interactions through the virtual agent. However, these systems suffered from the same shortcomings; they lack physical awareness of the real world environment. Agents that resides on the fiducial marker did not have geography knowledge of the physical environment hence they are unable to move around the environment naturally owing to their incapability to handle physical constraints.

The previous agent systems also have limited their communication channels between the user and the agent. For instance, the Monkey Bridge and AR Lego system only allowed user to interact with the system by manipulating physical objects. Other systems such as Welbo restricted the interaction to only speech.

However, humans communicate through a gestalt of actions which involve much more than just the speech or gestures alone. Non-verbal expressions such as facial expressions, head, body and hand movements all potentially contribute information to the communicative act. They allow the speakers to add further information to their messages including psychological state and attitude.

The complexity of this communicative act should be taken into consideration for agent design; aiming at modeling and improving such interaction by developing user friendly applications which should simplify and reduce user's cognitive load to use agent systems. Psycholinguistic studies have confirmed the complementary nature of verbal and non-verbal aspects in human expressions demonstrating how visual information processing integrates and supports message comprehension [9-11].

Several research works on mutual contribution of speech and gesture in communication and on their characteristics have being carried out. Some models include characteristics of emotional expressions [12, 13] as well as speech and gesture modeling, multimodal recognition and synthesis of facial expressions, head, hands movements and body postures for virtual agents [14-18].

4. AGENT COMPONENTS

In this section we describe the basic elements of our multi-modal AR agent (Fig. 1). Namely: static and dynamic visual appearance, physical awareness and user action recognition.

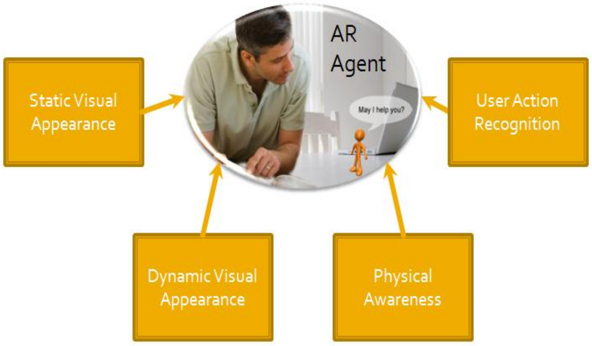


Figure 1. Fundamental components that make up our AR agent

4.1 Agent Static and Dynamic Visual Appearance

Our perception of the agent consists of two distinctive components: static and dynamic visual appearance. Static visual appearance refers to the inanimate figure with visual properties. These properties constitute the static qualities of the agent's visual appearance. When these qualities are brought to life in the forms of gestures, speech and gaze, we encounter the dynamic qualities of the agent's appearance. Together the static and dynamic qualities add up to the overall visual experience.

4.1.1 Static Visual Appearance

According to [19], designing an agent's static appearance is a complex task that spans over several dimensions. These dimensions may include humaness, physical property and graphical style. For the humanness dimension, the agent can be modeled as humans, animals, nonliving objects or some bination of these entities. In the physical property dimension we look at the basic attributes of the agent such as the shape, size color and type of skin, hair, clothes. The graphic style dimension is the most complex of the three where the agent appearance can be decided upon a wide range of different styles or expressions. For example, the agent can be really naturalistic such as the characters from *the sims*, or possess a rather abstract appearance commonly seen in cartoon animations. Similarly, the agent can have a detailed appearance that is photo-like or reduced to semi-detailed contour drawings.

If a virtual agent in its visual appearance naturalistically resembles a human being, we will also expect the agent to behave as a human being and in particular as intelligent as a human being. A conflict between the abilities of an agent and the users' perception of agent abilities can inflict frustration. The more visually realistic the agent is, the higher the expectation of the user in relation to the suitability and 'intelligence' of agent's utterances and actions. Past research have proved a higher tolerance of the limitations of a 'character' that is more sketchily represented [20]. Similarly V. Mulken et al [21] reviewed studies suggesting that agents with a more naturalistic visual appearance can be confusing owing to their misleading behaviors.

However higher degree of naturalism in virtual agents emphasizes the fact that in human-human interaction we are exploit information from human gestures, facial expressions, eye movements naturally. This is known to considerably increase the smoothness of interaction [21, 22]. Therefore, by employing a computer-based agent with gestures and facial expressions that are as human-like as possible, human–computer interaction will also become efficient [22].

Another important aspect of the agent static appearance is its physical size. For instance, considering a desktop AR environment where the user is sitting in front of a desk, asking an agent to find and show him/her the location of an object. In this case it would be more preferable to employ a small size agent over a real-human sized agent since the small agents would have the capability to move between objects on the desk. The agent's body size should be tailored according to the size of the working environment and the type of tasks it may perform.

4.1.2 Dynamic Visual Appearance

Expectation clash is not only a crucial aspect in static appearance design, but should also be considered carefully when designing the agent animation since mismatches between a high-quality stylized static appearance and awkward animations or awkward language and behavior can cause irritation. Similarly, when the agent performs an action which mismatches its static appearance, it may appear as abrupt and unnatural to the user. For example, considering an agent which has a creature appearance of a caterpillar, it may be perceived as surprising and strange if the agent performed a flying action. The dynamic visual appearance should depend on the chosen static visual appearance of the agent.

The planning and modeling of avatar animation can be drawn from a wide range of computer animation literatures. Kiss. S [23] presented a review of computer animation literature, concentrating on articulated characters, their interactivity and real time simulation. Modeling agent actions according to the environment is a complex task. According to [24, 25], it is desirable to introduce autonomy to AR agents so they do not need constant user guidance or carefully scripted behavior prepared for all possible situations. Instead the agents can proactively make decisions based on information from the sensors in the real world environment. As the result, only high-level goals are needed to be set, while the agent's reasoning engine takes care of low-level details to achieve the goals as quickly as possible.

The previous agent systems have been focusing on the realization of AR agents. As discussed in section two, they provide interaction in means of limited channels of communication. One of the novelties of this research is to further improve the quality of interaction by incorporating eye behavior into AR agent systems which has been overlooked in the past.

Eye behavior serves many functions in human interaction [26]. Eye signals may be used as content acts and as dialogue management acts. A speaker may communicate beliefs, intentions or affection by means of eye behavior. For example, express uncertainty or emotional state. A hearer may indicate degree of attentiveness by means of his or her eye behavior. This is often used in human-human communication to show clear acknowledgement of the other person's initiative and that he or she has been heard. As for managing the flow of interaction, eye behavior is efficient in structuring the dialogue such as turn taking [27]. For example, looking away is a means for a speaker to keep the floor. Resuming eye contact with a dialogue partner is a way to pass the turn to and ask for feedback from the hearer.

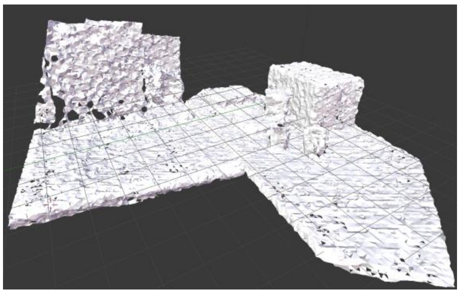
Eye gaze is often used for spatial referencing in human-human communication. When a speaker talks about an object to a hearer, he or she may show the physical location of that object by means of eye gaze direction and subsequently the hearer may lay his or her eye on the object. This is also known as mutual gaze. It regulates not only the flow of the conversation but also reassuring the speaker that the hearer is aware the subject of his utterance. We advocate that this kind of eye behavior is desirable in terms of human-agent communication.

4.2 Physical Awareness

AR agents and physical robot agents exist and assist user task in the real world environment. They are often classified together as "real world agents". Physical robots possess actual bodies that are fit to handle physical constraints in the real world; preventing them from going through other physical objects. In comparison, AR agents have virtual bodies that appear in the real world environment which can be modified more easily than actual robot bodies. However their flexible embodiment does not provide them with the ability to handle physical constraints like physical robots.

Therefore, in order to allow agents to move freely according to physical constraints of the environment, there is a need to provide the agent with the representation of the environment.

Recent advancement in computer vision has allowed dense surface reconstruction in real time [28]. R. Newcombe et al proposed KinectFusion [29] which generates dense reconstruction with a handheld kinect depth camera. In our implementation, we merged several point clouds captured by a single Kinect camera for scene reconstruction. This is further processed to generate a walk mesh representing the “walkable” region of the AR agent. Fig. 3 shows the sample output of surface reconstructions and its walk mesh.



(a) Reconstruction of our physical environment

(b) Walk mesh.

Figure 3. Scene reconstruction of our workspace (a) and the relative walkable region represented as a walk mesh (b).

4.3 User Action Recognition

The most natural ways of human communication is through a combination of speech and non-verbal expressions such as facial expressions, head, body and hand movements. In particular, speech and hand gesture have been known to be the most communicatively effect and natural ways to convey message [30]. Thus it is useful to equip the agent with the ability to decrypt the underlying message from these gestalts.

In an AR desktop environment, the agent may help the user with tasks in the real world environment where location indication may be useful. In this case, deictic gestures (pointing at an object while talking about it) are typically employed to convey location information. They can be nicely map onto deictic expressions like “this” and “that”, “these” and “those”, “here” and “there”. Thus, they can substitute particularly well for certain linguistic expressions, especially spatial expressions. At the same time, deictic gestures form indices to individual things. They can move the hearer’s attention from the speaker to the approximate region of the referent [31]. If speaker and hearer both know that their attention is focused on a similar region, then this facilitates reference resolution.

5. AGENT APPLICATION

In our research we are considering two potential applications for our AR agent in a desktop environment: A desktop secretary agent and an urban design agent.

5.1 Secretary agent

The primary purpose of secretary agents is to relieve user’s workload in a desktop environment allowing the user to work more efficiently. One simple example is helping user to locate objects in the real world environment. The agent would first identify the object, generate a virtual path and then guide the user to the object. To indicate the object in the real world, the agent may employ a combination of spatial dialogue speech and pointing gestures.

Another example function of the secretary agent is to assist the user task by providing the associated support information. This can be integrated with our previous work: ARDictionary [32] (Fig. 4). Considering a typical scenario where a user wishes to look up the definition of a word. He points at the word then inquires the AR agent for the word meaning. The agent moves towards the word, showing acknowledgement of the inquiry, search in the internet and finally inform the user of its findings.

Since the secretary agent can aid the user task in the real world environment, it must be capable of moving to arbitrary location in the environment. However this requires the agent to be physically aware of the environment layout so that it can generate and move along an appropriate path without violating physical constraints.

In our system (Fig. 5), we use a kinect camera and HMD overlooking the work space. The physical awareness is acquired by combining multiple mesh data captured from a single Microsoft Kinect camera. This mesh information is registered with the real world environment. The system is then able to calculate walk paths in real time (Fig. 6).

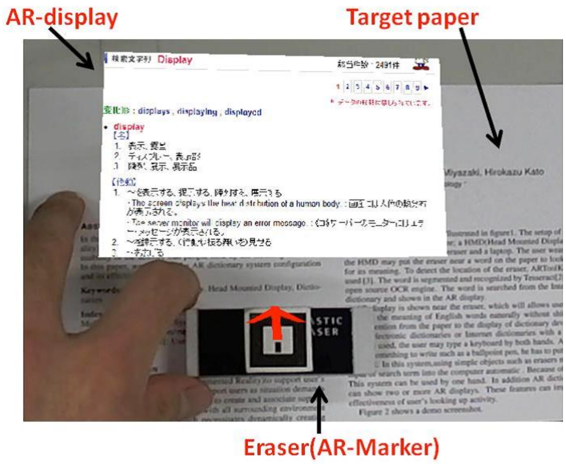


Figure 4. AR Dictionary [32]

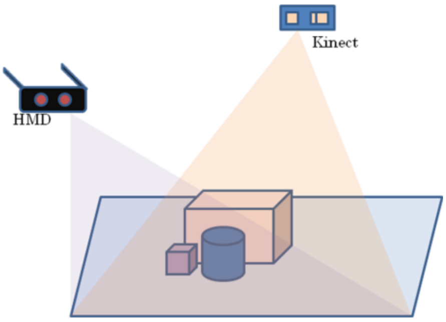


Figure 5. physical setup of the system.

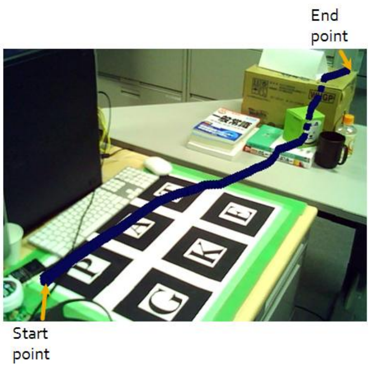


Figure 6. path generated from start to end point.

5.2 Urban Design Agent

Urban Design involves the design and coordination of all that makes up cities and towns. This is usually comprised of buildings, public space, streets, transport and landscape.

In the urban design process, designers have to materialize their designs and present the ideas using realistic representations. This is often achieved by creating a physical miniature model of the city. The design is then refined through a loop of presenting urban layouts, gathering new ideas and making alterations until it is finalized.

However it is tedious to remake the miniature model whenever new ideas need to be applied. The magic cup [33] offers a solution by allowing the user to add, remove and manipulate virtual objects in an AR setting using a cup-shaped tangible interface.

The virtual objects may be overlaid on top of the existing miniature model making it easier to visualize designer's idea. In the magic cup system, whenever the user wishes to change the function of the cup, he/she has to move the cup away from the scene to another marker set area. This draws the user's attention away from the main scene. This problem can be solved by replacing the magic cup interface with an AR agent. The user can simply point at an object or location and ask the AR agent to perform the operation. Another advantage of using AR agent in this application is that the agent can also act as a presenter, giving guided tours around the urban design and may answer the user's questions regarding the design.

6. CONCLUSION

In this paper, we have discussed the shortcomings from the past AR agent systems. A basic framework for AR agent design is proposed based on the urge to improve the naturalness of agent appearance and smoothness of human-agent communication. Physical awareness is a particularly desirable sense which provide the agent with mobility to move in the physical environment while retain physical constraints. We also advocate the employment of eye behavior to enhance the quality of interaction between user and AR agent. Finally, two agent applications have proposed for this research.

REFERENCE

[1] Author ed.^eds., "Augmented reality agents for user interface adaptation," *Computer animation and virtual worlds*, 2008, p.^pp. Pages.

[2] S. D. Barakonyi István, "Augmented Reality Agents in the Development Pipeline of Computer Entertainment," in 4th International Conference on Entertainment Computing, 2005.

[3] T. D. Pattie Maes, Bruce Blumberg and Alex Pentland., "The ALIVE system: wireless, Full-body interaction with autonomous agents," in ACM Multimedia Systems, 1997.

[4] H. K. Mahoro Anabuki, Hiroyuki Yamamoto, and Hideyuki Tamura, "Welbo: An embodied conversational agent living in mixed reality space," in Conference on Human Factors in Computing systems, 2000.

[5] V. L. L. Vacchetti, G. Papagiannakis, M. Ponder, P. Fu, "Stable real-time interaction between virtual humans and real scenes," in 3DIM, 2003.

[6] O. M. Marc Cavazza, Fred Charles, Steven J. Mead, and Xavier Marichal, "Interacting with virtual agents in mixed reality interactive storytelling," in Intelligent Virtual agents,

Kloster Irsee, 2003.

[7] T. P. I. Barakonyi, and D. Schmalstieg, “Agents that talk and hit back: animated agents in augmented reality,” in ISMAR, 2004.

[8] M. W. Istvan Barakonyi, Thomas Psik, Dieter Schmalstieg, “MonkeyBridge: Autonomous Agents in Augmented Reality Games,” in ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, 2005.

[9] M. M. Esposito A., "What Pauses Can Tell Us About Speech and Gesture Partnership," *Fundamentals of Verbal and Nonverbal Communication and the Biometrical Issue*, 2007.

[10] B. N. G. Esposito A., “The Role of Timing in Speech Perception and Speech Production Processes and its Effects on Language Impaired Individuals.,” in 6th International IEEE Symposium BioInformatics and BioEngineering, 2006.

[11] N. R. Ochs M., Pelachaud C., Sadek D., “Intelligent Expressions of Emotions,” in 1st International Conference ACII05, 2005.

[12] A. Esposito, “The Amount of Information on Emotional States Conveyed by the Verbal and Nonverbal Channels: Some Perceptual Data.,” in Progress in Nonlinear Speech Processing, 2007.

[13] G. M. Heylen D., Nijholt A., op den Akker R., “Facial Signs of Affect During Tutoring Sessions,” in LNCS, 2005.

[14] M. M. Hartmann B., Pelachaud C., “Implementing Expressive Gesture Synthesis for Embodied Conversational Agents.,” in Gesture Workshop, LNAI, 2005.

[15] W. I. Kopp S., “Synthesizing Multimodal Utterances for Conversational Agents,” *Computer Animation and Virtual Worlds*, 2004.

[16] M. J.-C. Pelachaud C., Andre E., Chollet G., Karpouzis K., Pele D, “Intelligent Virtual Agents,” *LNAI 4722, Springer*, 2007.

[17] E.-H. A. Petrovska-Delacretaz D., Chollet G., "Automatic Speaker Verification: State of the Art and Current Issues."

[18] T. D., “Talking to Virtual Humans: Dialogue Models and Methodologies for Embodied Conversational Agents.,” *Modelling Communication with Robots and Virtual Humans*, 2008.

[19] M. H. A.Gulz, “Visual Design of Virtual Pedagogical Agents: Naturalism,” in Design and Engagability, 2006.

[20] C. Dowling, “Intelligent agents: some ethical issues and dilemmas,” in AIC, 2000.

[21] D. Dehn, and van Mulken, S, “The impact of animated interface agents: a review of empirical research,” *Human computer studies*, 2000.

[22] S. M. J. Gratch., A. Egges., A. Eliëns., K. Isbister., A. Paiva., T. Rist., and P. Hagen, “Design criteria, techniques and case studies for creating and evaluating interactive experiences for virtual humans,” in *Dagstuhl seminar 04121 on Evaluating ECAs*, 2004.

[23] S. Kiss, *Computer animation for articulated 3D characters*, 2002.

[24] M. W. István Barakonyi, Thomas Psik, Dieter Schmalstieg, "MonkeyBridge: Autonomous Agents in Augmented Reality Games."

[25] S. D. Barakonyi István, "Augmented Reality Agents in the Development Pipeline of Computer Entertainment."

[26] I. Poggi, Pelachaud, C., and Rosis, F. d., “Eye Communication in a Conversational 3D Synthetic Agent,” in Special Issue of Artificial Intelligence Communications, 2000.

[27] R. Vertegaal, Shell, J., Chen, D., Mamuji, A., “Designing for augmented attention: Towards a framework for attentive user interfaces,” *Computers in Human Behaviour*, 2006.

[28] R. A. N. a. A. J. Davison, "Live Dense Reconstruction with a Single Moving Camera."

[29] D. K. Shahram Izadi, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, Andrew Fitzgibbon, “KinectFusion: Real-time 3D reconstruction and Interaction Using a Moving Depth Camera,” *ACM SIGGRAPH*, 2011.

[30] Author ed.^eds., “Gestures and Phases: The Dynamics of Speech-Hand communication,” *Ecological Psychology*, 2008, p.^pp. Pages.

[31] M. M. a. B. Louwerse, A, “Focusing Attention with Deictic Gestures and Linguistic Expressions,” in Annual Conference of the Cognitive Science Society, 2005.

[32] C. C. T. Yuichiro Fujimoto, Makoto Fujisawa, Toshiyuki Amano, Jun Miyazaki and Hirokazu Kato, “AR Dictionary: Dynamic Annotation for English Words in Printed Documents,” in 20th International Conference on Artificial Reality and Telexistence, 2010.

[33] K. T. H Kato, M Tanabe, T Nakajima, Y Fukuda, “MagicCup: a tangible interface for virtual objects manipulation in table-top augmented reality,” in Augmented Reality Toolkit Workshop, 2003.