

Virtual Human Simulation on Memory Acquisition and Walking with the Memory

N. Mukai^a, Y. Hayashi^a, Y. Chang^a

^aGraduate School of Engineering, Tokyo City University, Stags, Tokyo, Japan, 158-8557
Email: mukai@cs.tcu.ac.jp

Abstract: In virtual simulations, one of the most difficult tasks is how to model human behavior and how to validate the model by performing the simulations. Actually, many researchers have tried to model virtual humans for the behavior. Some researchers focused on how to decide the path for each person to move in a crowded situation. Others considered that human behavior cannot be generated with just one model, and the model should be categorized to some patterns such as individual, groups and crowd. Then, human actions are generated by applying each model according to the situation. In crowded situation, people might collide with each other if they move their own way without considering other person's path. Then, collision avoidance model was also investigated. In addition, there are some simulations considering real situations such as crowded train station or epidemics of infectious disease.

On the other hand, vision and memory are necessary for people to take actions so that some researchers proposed methods how to obtain the vision and construct the memory. In the real world, we memorize some landmark buildings at each intersection and build up our own map. With the help of the map, we can walk the same way several times and also can return to the original position. Furthermore, we sometimes intentionally change the way to find a new path. In this case, people decide the path by integrating the vision they are obtaining with their memory acquired until then.

Therefore, we have performed a simulation for virtual humans on how to acquire their memory and decide the direction by integrating the vision with the memory in order to investigate real human behavior. People memorize the information they have acquired by walking as a form of map, which is divided into two types as the following.

Route map: 1 dimensional information they have acquired along the walking path.

Survey map: 2 dimensional information they have reconstructed from the route map.

We have modeled how to construct the route map and transform the route map information into the survey map. The route map and the survey map include landmark images they have encountered at each intersection. Then, we have performed two kinds of simulations as follows.

Same route walk simulation: simulation to check if people can walk the same path based on the survey map information they have constructed. People can recognize intersections with pattern matching for the scene they are looking and the landmark images stored in the survey map. If the recognition is correct, people can walk the same way, while people take a wrong way if the recognition is incorrect.

Different walk simulation: simulation to check if people can take the correct path even if they have started from another position different from the point they used when the route map was constructed. People can take the correct path if they correctly recognize the landmark they are looking, or they take a different direction if the landmark image does not correspond to the one they have in the survey map.

In the result of the simulations, a virtual human took the same path when he started to move at the same point. On the other hand, in the second simulation, there have been two kinds of results. One is that the virtual human took the correct path by the correct recognition of landmark images, and the other is that he did not take the correct path due to the incorrect recognition. This result reflects human behavior that people sometimes mistake the recognition of a street and move to a wrong direction.

Keywords: *Computer graphics, virtual human, memory acquisition, walking simulation with memory*

1 INTRODUCTION AND RELATED WORKS

One of the most challenging issues in computer graphics is to generate human models and to perform the simulations in a virtual space. In order to generate virtual human models, many technologies should be integrated such as motion capture, face and body modeling, motion control, groups and crowd simulation, interaction with 3D objects, and so on (Magenat-Thalmann and Thalmann (2004)). In addition, human behavior model should be divided into some categories such as individual, groups and crowd (Thalmann (2004)).

Reynolds (1999) decomposed the motion of human behavior into 3 levels: action selection, steering, and locomotion. Li et al. (2001) considered that a crowd is composed of a leader and many followers, and that the path decided by a leader of a group becomes the motion constraint of the path to be decided by other leaders. The generated motions are collision-free for multiple virtual crowds. In order to simulate human behavior precisely, it takes much time to simulate the action process and render the human models. Then, Guy et al. (2010, 2012) simulated crowd behavior at interactive rates by introducing biomechanical measurements and PLE (Principle of Least Effort) to define the trajectory of large-scale crowds.

On the other hand, Musse and Thalmann (2001) divided virtual human behavior models into 3 categories: individual, group and crowd, and generated each behavior by applying scripted, reactive and guided behavior, respectively. The crowd motion generated by Li et al. (2001) is collision-free; however, in general, people might collide with each other in a crowded situation so that Koh and Zhou (2014) proposed a collision avoidance model to generate some kinds of virtual human behavior, which model can be applied to a path planning in a dynamic environment.

In order to generate autonomous human behavior, virtual humans should have their vision and decide their action according to the judgment based on the vision. Then, Rymill and Dodgson (2005) introduced psychologically-based vision; however, in the method, an attention manager knows the whole situation in the crowd and can pass on the necessary information to each person. On the other hand, Mukai et al. (2015) introduced a human model that has his/her own eyes in a virtual space and performed crowd simulation just by applying the individual human model without decomposing human behavior.

There are many applications related to the virtual human model and its simulation. Shao and Terzopoulos (2005) presented autonomous pedestrians in urban environments by applying the individual model. Kim et al. (2012) simulated dynamic patterns of crowd behavior by using stress modeling based on General Adaptation Syndrome Theory, which gives general responses to some stresses such as cold, fatigue, fear and so on. Liam Delaney and Williams (2013) applied virtual human behavior to risky situations such as epidemics of an infectious disease, which method is based on PMT (Protection Motivation Theory) that is a psychological framework for people to change their behavior to protect their health.

These researches are the methods how to generate human behavior in the current situation; however, in the environment where the situation changes, people have to decide the path to move by integrating the current situation with their memory that has the previous situation. James J. Kuffner and Latombe (1999) presented a method of modeling synthetic vision, memory, and learning to generate autonomous animated humans. The information virtual humans obtain from their vision is saved in a spatial memory as a set of object IDs, and the set is updated by deleting old observations and by adding new ones. With the method, they performed a simulation that a virtual human explored an unknown maze environment. On the other hand, Yang et al. (2010) designed virtual human memory by decomposing it to three layers. The first layer saves newly obtained or old but unforgettable knowledge, and the second layer saves the knowledge that is no longer new but easily recalled. The last layer includes a database that records all obtained knowledge, and the memory is maintained by following Ebbinghaus forgetting law.

In the real world, we can walk on the same path several times or return to the original position by memorizing landmark buildings at each intersection and building up our own map. In addition, we can modify the map by adding new images that are encountered on a different path. Therefore, in this paper, we report a method that models the process generating a map and a simulation to check if a virtual human can walk on the same path with the constructed map. In addition, we have performed another simulation to check if virtual humans can take the correct path by utilizing the map even if they start to move from a different position. This work can be utilized for the investigation of real human behavior.

2 VIRTUAL HUMAN MODEL

2.1 Vision

We model virtual humans who can take actions autonomously by setting their eyes. In the real world, humans can obtain various information through their eyes, and the information can be divided into two kinds: color and distance information. In our simulation, virtual humans are generated with modeling techniques of computer graphics as one object in object oriented programming. In the rendering process, the color and distance information can be obtained through frame and Z buffers, respectively. Virtual humans can recognize the environment such as buildings, road, and so on, with the color information, while they can recognize the distance from the position they are standing to the target place, with the distance information. Fig. 1 shows the examples of frame and Z buffers. We can recognize the color and shape of houses, walls and roads from the frame buffer, while from the Z buffer, we can understand the distance from the position the virtual human is standing to each object such as houses or walls. The Z buffer shows that the whiter the pixel is, the shorter the distance is.



Figure 1. Frame and Z buffers

2.2 Corner Detection

In order to walk autonomously in the virtual environment, virtual humans have to recognize if there is a road ahead or if there is a corner in the right or the left. This recognition can be achieved as follows with Z buffer.

Front Recognition: Scan the center part of Z buffer. If the distance is longer than a threshold, there is nothing in front of the person and he/she can walk ahead.

Corner Recognition: Scan the horizontal line in the middle part of Z buffer in the vertical direction. If there are two pixels which values are different larger than a threshold (two pixels are apart), there is a corner on the way. Fig. 2 shows the frame buffer, and the two pixels, p1 and p2, are on the horizontal line in the middle part of the image in the vertical direction. Fig. 2 (a) shows there is no corner because Z buffer values of two pixels are similar since both pixels are on the same wall, while (b) shows there is a corner in the left because there is distance between two pixels since they are on different walls.

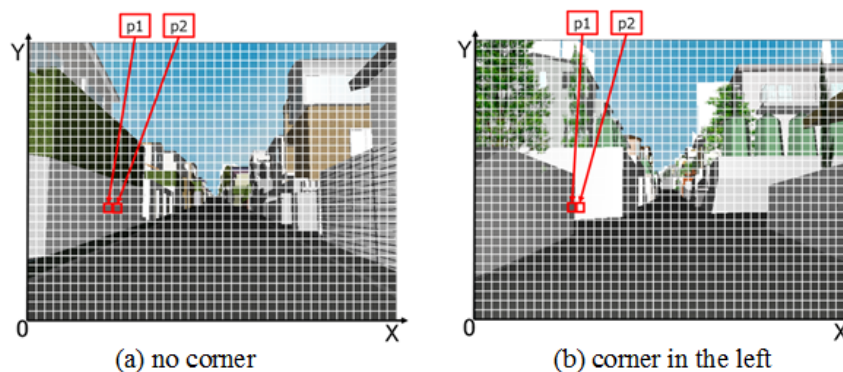


Figure 2. Horizontal line scan for corner detection

3 MAP GENERATION

In order to walk autonomously, virtual humans have to memorize the environment and store the information as a map. The map is known as cognitive map, which is divided into 2 types: route and survey maps (Yokozawa et al. (2004)). The construction method of each map is described as follows.

3.1 Route Map

Route map is one of cognitive maps, which stores information that people obtain on their way according to the sequential order in which they encounter some landmarks. In order to construct route map, space information for construction of route map is defined as the following (Mori and Iwai (1994)).

<Space information constructing route map>

Paths: space people can move: road (by walk), railroad (by train), air road (by airplane), and so on.

Nodes: point connecting paths: intersection, station, airport, and so on.

Landmarks: mark with which people can recognize the path: building, park, signboard, and so on.

Edges: border line over which people cannot move: building wall, seacoast, stream bank, and so on.

In this research, the following information is stored in route map.

<Element stored in route map>

Nodes: intersections are used.

Landmarks: house or wall images are used.

Distance: path length between nodes: long or short is used in the simulation.

Direction: walking direction at each node: straight, right or left is used in the simulation.

Here, 4 house or wall images at each intersection are stored as Landmarks in the route map. Fig. 3 shows the process of extracting 4 images as the landmarks. Fig. 3 (a) shows the frame buffer image, and there is an intersection on the way, which corners are marked as red circles. Fig. 3 (b) shows the Z buffer image, with which we divide the image into far and near places by checking values at red circles. Fig. 3 (c) and (d) show the far and near images, respectively. Then, 4 landmark images can be extracted by cutting each image with a vertical line that passes each corner. Fig. 3 (e)-(h) show the extracted 4 images that are stored in the route map.

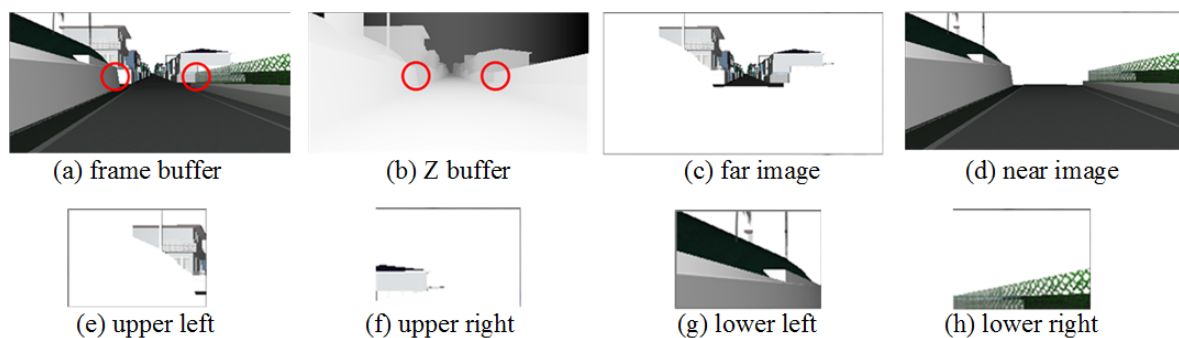


Figure 3. Landmark image extraction process

3.2 Survey Map

Survey map, which is 2 dimensional information, can be constructed with a set of route maps, which are 1 dimensional sequential information. Fig. 4 shows an example of survey map that has 2 information of node (1) and (2). When virtual people get to the intersection (1), they understand the about distance until the intersection from the previous point; however, real people do not know the exact length so that “long” or “short” is used as the distance. In the real world, people memorize the houses or buildings at the intersection. Then, virtual people also store the 4 images at the intersection. They can also know if there are paths on the way by using 2.2 method. In the simulation, it is assumed that intersections are crossroads. Then, the direction is straight, right or left.

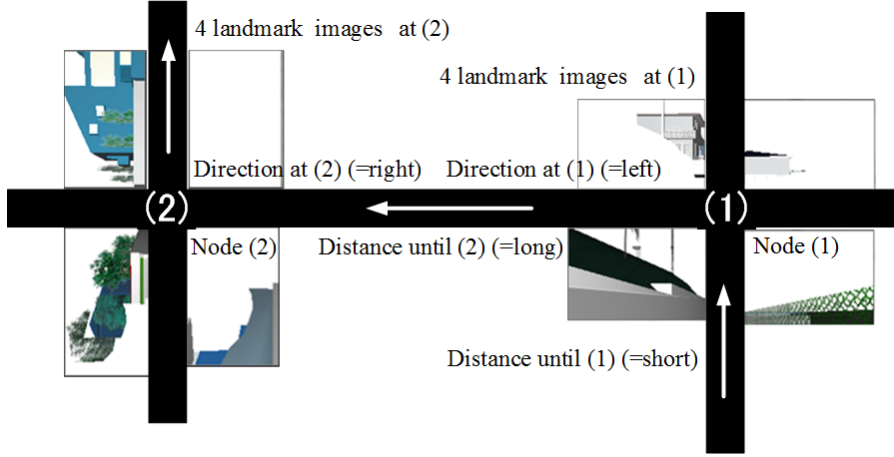


Figure 4. An example of survey map

4 SIMULATION

4.1 Landmark Recognition

Virtual humans take actions, which are the selection of the path and walking to the direction according to the recognition of the landmark images. Landmarks are stored as 4 images at each intersection. Template matching method is used for the recognition of landmarks. Landmark images have color information; however, the images are converted into gray scale and used for the template matching. Normalized Cross-Correlation (NCC) is used as the matching evaluation as the following.

$$R_{NCC} = \frac{\sum_j \sum_i \{T(i, j) \cdot I(i, j)\}}{\sqrt{\sum_j \sum_i T(i, j)^2} \sqrt{\sum_j \sum_i I(i, j)^2}} \quad (1)$$

Where, i and j are for the horizontal and vertical indices of the image, respectively. $T(i, j)$ and $I(i, j)$ are the gray scaled template image and the gray scaled current image that virtual humans are looking at the intersection. R_{NCC} has the range of $[0.0, 1.0]$ and 1.0 means that the images are completely coincident. If R_{NCC} is greater than a threshold, virtual humans understand that the intersection is the same place as the one they encountered before.

4.2 Walk Simulation based on the Memory

Based on the memory, we have performed the following two kinds of simulations.

Simulation 1 (Same route walk simulation): simulation to check if people can walk on the same path according to the survey-map information.

Simulation 2 (Different walk simulation): simulation to check if people can take the correct path even if they started from another position different from the previous point.

Fig. 5 shows a survey map, which has been constructed by a virtual human who started to walk at the point of (S1). (S1) and (S2) are the start points for simulation 1 and 2, respectively. On the simulation 1, he was able to take the same route as when he constructed the survey map.

On the other hand, on the simulation 2, he understood that the intersection (1) is the same as the map, and took the correct direction although he started at a different point (S2). At this time, “straight” is the correct direction although “right” is the correct on the map. At intersection (2), he easily recognized that the intersection was the same as the map because he took the same way in the same direction. At intersection (3), he had taken the “straight” direction before, however, in the simulation 2, he took the different direction and turned to the “right” although he took the same way in the same direction.

As the result of 5 time simulations, 3 times were the same route as before, however, 2 times were different from before. The reason is due to the template matching method that calculates the coincident rate for the whole image. If the position from which landmark images are taken is different, a different image is used for the matching, and the coincident rate becomes lower. On the other hand, these results can be considered to reflect that human memory is ambiguous and people sometimes take a wrong decision.

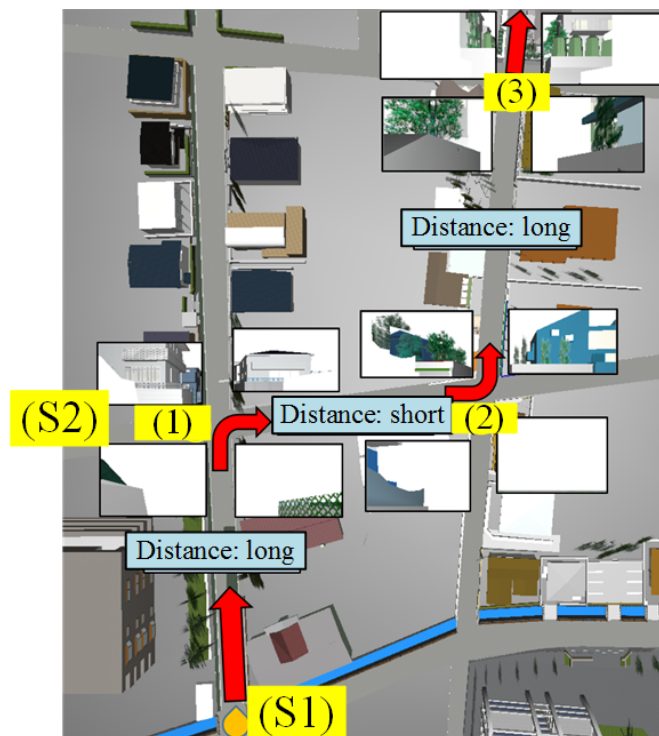


Figure 5. Survey map for simulation

5 CONCLUSIONS AND FUTURE WORKS

We have proposed a memory acquisition method in a virtual space, which memory is stored as 2 dimensional survey map through 1 dimensional route map. We have also performed two kinds of simulations to check if virtual humans can take the correct directions according to the acquired memory in order to investigate real human behavior. For the simulation where people start at the same position as when they acquired the memory, virtual humans were able to take the same directions. On the other hand, for the simulation where people start at a different position, we have found that virtual humans sometimes take wrong directions just as we do in the real world.

This result reflects ambiguity of human memory; however, also shows the template matching method is not robust. One reason is that gray scaled images are used for the template matching, and the color information was not used. The second reason is that the template matching consistent rate is calculated for the whole image. The template image should move over the target image, and search for the best matching position. In addition, some template images viewed from different positions and angles should be stored in the memory for a robust template matching. Then, we plan to improve our method by considering these issues.

REFERENCES

- Guy, J., J. Chhugai, S. Curtis, P. Dubey, M. Lin, and D. Manocha (2010). PLEdestrians: A least-effort approach to crowd simulation. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographic Symposium on Computer Animation*, pp. 119–128.
- Guy, S. J., S. Curtis, M. C. Lin, and D. Manocha (2012). Least-effort trajectories lead to emergent crowd behaviors. *Physical Review E* 85(016110).
- James J. Kuffner, J. and J.-C. Latombe (1999). Fast synthetic vision, memory, and learning models for virtual humans. In *Proceedings of Computer Animation*, pp. 118–127.
- Kim, S., S. J. Guy, D. Manocha, and M. C. Lin (2012). Interactive simulation of dynamic crowd behaviors using general adaptation syndrome theory. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pp. 55–62.
- Koh, W. L. and S. Zhou (2014). An extensible collision avoidance model for realistic self-driven autonomous agents. In *Proceedings of the 11th IEEE International Symposium on Distributed Simulation and Real-Time Applications*, pp. 7–14.
- Li, T.-Y., Y.-J. Jeng, and S.-I. Chang (2001). Simulating virtual human crowds with a leader-follower model. In *Proceedings of the 2001 Computer Animation Conference*, pp. 93–102.
- Liam Delaney, Adam Kleczkowski, S. M. S. R. and L. Williams (2013). Reflections on a virtual experiment addressing human behavior during epidemics. In *Proceedings of the 2013 Summer Computer Simulation Conference*, Number 33.
- Magenat-Thalmann, N. and D. Thalmann (2004). *HANDBOOK OF Virtual Humans*. England: John Wiley & Sons., Ltd.
- Mori, T. and A. Iwai (1994). cognitive map. *IEICE Journal* 77(8), 836–839.
- Mukai, N., K. Tanaka, and Y. Chang (2015). Crowd simulation by applying individual human model with vision. In *Proceedings of the 2015 International Conference on Cyberworlds*, pp. 210–215.
- Musse, S. R. and D. Thalmann (2001). Hierarchical model for real time simulation of virtual human crowds. *IEEE Trans. on Visualization and Computer Graphics* 7(2), 152–164.
- Reynolds, C. W. (1999). Steering behaviors for autonomous characters. In *Proceedings of Game Developers Conference*, pp. 763–783.
- Rymill, S. J. and N. A. Dodgson (2005). Psychologically-based vision and attention for the simulation of human behaviour. In *Proceedings of the 3rd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*, pp. 229–236.
- Shao, W. and D. Terzopoulos (2005). Autonomous pedestrians. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 19–28.
- Thalmann, D. (2004). Crowd and group animation. In *Proceedings of the ACM SIGGRAPH 2004 Course Notes*, Number 34.
- Yang, Z., Y. Fan, B. Zhang, and X. Cheng (2010). A computation memory model with human memory features for autonomous virtual humans. In *Proceedings of the 2010 International Conference on Computer Application and System Modeling*, pp. V3–246–V3–250.
- Yokozawa, K., E. Wada, and H. Mitsumatsu (2004). The coding and transformation of cognitive maps in a virtual environment. *IEICE Trans. J87-A(1)*, 13–19.