ABSTRACT

We create a framework for physical interaction with virtual human. The virtual human reacts to the input of the user in realtime. The visual human has an ocular system model, which mimics human ocular system, to generate natural and human-like gaze motions and to create inputs for the cognitive model of the virtual human.

The visual input for virtual human is generated from the dynamics simulator for the virtual world. The gaze target is chosen by vying of bottom-up and top-down attentions. The motion controller for the eyes and head is based on observations and models of human gaze motion.

We create a virtual boxing game as an application of proposed framework. Proposed framework succeeds to create natural gaze reaction. In addition, by changing the balance of the bottom-up and top-down attention, looks of concentration of the virtual human can be tuned.

Keywords: Gaze, ocular system, attention


1 INTRODUCTION

Recently, virtual humans are often used in games and movie productions. The reality of graphics and motions is remarkably improved. The appearances of virtual humans come to be real with GPUs and motion capture techniques. Particularly, highly interactive contents such as arcade games require virtual humans to react to inputs of users in real time. Conventional computer games create realistic reactions of virtual humans by choosing and connecting offline-prepared motion data which are appropriate for the user’s button inputs. This technique is suitable for limited digital input.

However, due to advances in human interfaces of interactive applications, the variations of user inputs increase and more and more variations of reactions are required. More direct interaction through analog pads or haptic interfaces requires new paradigm of reaction generation.

[8] proposed to create reactive motions of virtual humans based on dynamics simulation and motion control model of humans. In this paper, we propose an ocular system model of virtual human and generate motions of eyes and heads and appropriate behaviors of virtual humans extending [8]. Because humans obtain informations of outside world via their eyes, their gaze shows their intention and what they will do next. Visual attention of human is driven by the saliency of visual input(bottom-up attention) and can be biased consciously(top-down attention). In our method, the target of gaze for virtual human is decided by vying of bottom-up attention and top-down attention. Therefore gaze of virtual human is important not only for verbal communications but also physical and/or direct interactions and cooperative works between human and virtual humans. In addition, we describe the framework for interaction with the virtual human through an avatar controlled by the user via haptic interface.

2 RELATED WORKS

Many virtual humans are employed in computer games such as RPGs, sports games and action games. Fighting games such as Virtua Fighter [21] are pioneer users of virtual humans. In these games, reactions of virtual humans are created by connecting motions in a motion database such as Motion Graph [13]. Therefore, variations of the reactions are limited and it takes large costs to create a motion database. Lee and Lee [14] created realistic motions of boxers with a database approach. However, their method does not create dynamic reaction for contacts and reactive motions of eyes and heads.

The advances of the human interface and virtual reality technologies for entertainment computing bring wide variations of user’s inputs. Therefore, virtual humans are required to perform many variations of reactions. Seungzoo et al. applied a database approach to create a reactive virtual human for haptic interaction [11]. The interaction in their system was a catch ball. It was not a direct contact interaction because of the restriction of the variation of motions.

Space time constraint methods [20] generate optimal trajectory which minimize some optimization function. However, these methods must solve optimization of the trajectory and it is difficult to run them in realtime. Dynamic control and simulation methods use controllers to compute joint torques based on current states and desired actions. These methods create dynamically correct motions from specified motions [26, 24], state machines [9] and environmental physical input [18]. These methods generate appropriate motions for variations of the input. State machines are used to create motions and actions of virtual characters corresponding to the environments [4] and user’s instructions [1]. However, these researches did not include direct interactions.

Some models for gaze generation were proposed. [10] propose a computational model of visual attention from knowledges of neuropsychology. [19] [16] propose a bottom-up (saliency based) eye gaze model and implementation to virtual humans, which generates eyes and heads motions based on fields of view of virtual humans. Both models focus on automatic generation of gaze and do not provide users with interactions. Their model does not create reactive behaviors of virtual humans.

The role of gaze on communications are often discussed and gaze model of virtual humans for communication agents are proposed [5, 23]. [15] proposes a method to generate saccadic eye movement during oral dialogs using a statistical model. [17] pro-
poses to generate heads and eyes motions of virtual actors corresponding to conversation in realtime. These gaze models focus on oral dialogs and does not regard motions of the user.

3 PROPOSED FRAMEWORK

Figure 1 shows overview of the proposed framework. The proposed framework consists of a haptic interface and a visual display, dynamics simulators for the virtual world and internal models of the virtual human, human body models for user’s avatar and the virtual human and a cognitive model of virtual human with ocular system model, behavior planning model and motion controllers.

The user controls the avatar through a haptic interface. The human body model for the avatar traces the motion of the user, while forces which act on body model for the avatar are feedbacked to the user. The ocular system model of the virtual human corrects the internal virtual world model from the information in the field of view of the virtual human. The behavior planning model predicts the virtual world, selects the next action of the virtual human and gives an action instruction to the motion controller for the body of the virtual human. The target of gaze is decided by the bottom-up attention direct from the ocular system model and top-down attention from the action instructions of the behavior planning. The motion controller decides the joint torques, which act on the human body model of the virtual human, from the instructions.

3.1 Dynamics Simulator

We employ a dynamics simulator named Springhead [7], which is suitable for haptic interaction. The dynamics simulator, Springhead, employs Featherstone’s method for joints and penalty method for contacts and requires less computation time for each iteration.

3.2 Human Body Model

In this paper, we model the upper part of the body for virtual boxing application. We use this model for virtual human, and for the avatar, we use the same model except the eyes. We set the dimensions, weights and inertias of the body parts and the limits of the joint angles referring databases of human characteristics [12, 22]. In addition, we give a default angle and a spring-damper model for each joint to give a default posture of the human body model.

3.3 Ocular System Model of the Virtual Human

The ocular system of virtual human get information from the dynamics simulator of the virtual world. The information is limited by the field of view of the virtual human. Then, a value of attractiveness for each object is given to create the bottom-up attention. The behavior planning also gives a value of attractiveness for the target object as a top-down attention. Then, the object with the largest value of attractiveness gets the attention of the ocular system and is chosen as the gaze target. The eyes and the head cooperatively moves to gaze the target based on the saccade and smooth pursuit of human.

3.3.1 Visual Input for the Virtual Human

The dynamics simulator for the virtual world has whole information of the virtual world. The virtual human perceives some part of this information via their ocular system. The virtual world consists of some solids here. The virtual human and the avatar consist of some solid objects as described in section 3.2. The virtual human perceives the information on solids inside the field of view.

3.3.2 Bottom-up attention

Humans give attention to and glance at one of the objects in their field of view unconsciously. This is called bottom-up attention. The choice of the target for the glance depends on the attractiveness of each object. The attractiveness of each object can depend on the interest of the human and properties of the object such as shape, color, texture and motion.

In our framework, we focus on interaction with user’s motions. Therefore, our system calculates the value of the attractiveness for the virtual human from the motion of the object. Humans perceive three dimensional motion of objects from two dimensional input images. They perceive parallel translation to gaze direction as enlargement-reduction of object images and orthogonal translation as translation of object images and rotation around the gaze direction as rotation of object images. It is known that there are some neurons which correspond to these motions in human brain. Therefore, we pick velocities of objects corresponding to these motions, and create an evaluation function for the attractiveness of objects.

The value of attractiveness of an object \( A_i \) is defined by:

\[
A_i = \| v_i - v_j \cdot e \cdot \frac{e}{|e|} \| + \left( v_i \cdot e \cdot \frac{e}{|e|} \right) + \left( L_i \cdot e \cdot \frac{e}{|e|} \right)
\]

where \( e \) is gaze direction, \( v_i \) is velocity of object \( i \), \( L_i \) is angular velocity of object \( i \).

3.3.3 Top-down attention

Humans give attention to targets of behaviors and give gaze to them. This is called top-down attention, because it comes from the will of the human.

The behavior planning of the virtual human (see section 3.4) selects a target object as a target of the selected behavior. In addition, the behavior planning gives the target object some value of attractiveness at the same time. Then, an object with the maximum value of attractiveness is chosen as a gaze target of the virtual human.

3.4 Behavior Planning of the Virtual Human

Humans recognize environments and predict changes of the environments. Then, they decide their next behaviors. This creates natural and smooth motion of humans. The behavior planning of the virtual human mimics this function of the human mind in the proposed framework. The ocular system model has an internal model of the virtual world. The behavior planning copy the model to predict the user’s motion.

In the proposed framework, the internal model is represented by a dynamics simulator, which is a copy of the simulator for the virtual world. The cognitive model predicts the motion of the user and the virtual human itself with a dynamics simulator which runs faster than realtime. Then, the cognitive model analyzes the motions and contacts to decide the next behavior. Many motions of the upper part of the body can be represented by combinations of reaching motions. We represent an action of a behavior as a reaching motion.
of a body part. The part, the duration and the target are notified to the motion controller.

### 3.5 The Motion Controller for the Virtual Human

The motion controller creates motor control signal of the virtual human from the three parameters (the part, the duration and the target).

Flash [3] proposes a control model of human reaching motion from measurements of human motion and simulations. The model consists of a minimum jerk model and a PD control of joints. We use this model for the motion control of the virtual human.

### 3.6 Motion Controller of the Eyes

As described in section 3.3, the ocular system model of the virtual human chooses an object to gaze. The eyes and the head of the virtual human cooperatively move to gaze to the target object. Motions of human eyes are classified into two categories, saccade and smooth pursuit.

#### 3.6.1 Saccade

When the gaze target is changed or moves fast, a saccade arise. [6] models saccade motion of the virtual human by an equation of minimum jerk with speed limit (Eq.(3)), which generates a motion near to the saccade motion of a body part. The part, the duration and the target are notified to the motion controller.

\[
\theta(t) = \begin{cases} 
\theta_{\text{slow}}(0.5) \leq \omega_{\text{max}} \\
\theta(t) = \theta_{\text{slow}}(t) \\
\theta(t) = \theta(t/T) + \omega_{\text{max}}(t-T/2) \quad (T/2 < t \leq t_{\text{end}} - T/2) \\
\theta(t) = \theta(T/2) + \theta_{\text{accel}}(t) \quad (t_{\text{end}} - T/2 < t \leq t_{\text{end}}) 
\end{cases}
\]

where

\[
\begin{align*}
f(t) & = 6(t/T)^5 - 15(t/T)^4 + 10(t/T)^3 & (3) \\
\theta_{\text{slow}}(t) & = \theta f(t) & (4) \\
\theta_{\text{max}}(t) & = \omega_{\text{max}}/f(0.5T) & (5) \\
\theta_{\text{accel}}(t) & = \theta_{\text{max}} f(t) & (6) \\
t_{\text{end}} & = T + (\theta_0 - \theta_{\text{max}})/\omega_{\text{max}} & (7)
\end{align*}
\]

\[
\begin{align*}
\theta_0 & : \text{the angle between the starting and the target direction.} \\
t & : \text{time from starting of the saccade.} \\
\omega_{\text{max}} & : \text{maximum speed of eyeballs = 8.7266 [rad/s]} \\
T & : \text{minimum time span of saccade = 0.05 [s]}
\end{align*}
\]

We refer to [6] for the value of $T$. The eyeball is PD-controlled to the direction of $\theta(t)$.

#### 3.6.2 Smooth Pursuit

Human eyes can track object moving slow. This motion of eyes is called smooth pursuit. We employ a model proposed in [25] to implement smooth pursuit to the virtual human. [25] models binocular motor system of human. The model reproduces different dynamic characteristics of conjugate and vergence eye movements including smooth pursuit and vestibuloocular reflex. The output of the model is target orientations of eyes. We calculate torques to give the eyeballs by PD controllers.

### 3.7 Motion control of the head

When the gaze target direction is outside of the movable area of the eyes, the human head moves to include the direction of gaze in the movable area. If the angle between target direction and front direction of the face is larger than about 30 degrees, head motion arises [2]. Usually, the head motion starts 20ms after the starting of the eyeball motion. However, if the target position can be predicted or the motion comes from active searching, head motion can occur before the eyeball motion [2]. From the above, we define two conditions for startings of the head motion of virtual human.

- The angle between the gaze target and direction front of the face is larger than 30 degrees.
- 20ms have passed after the starting of the eyeball motion or the motion comes from top-down attention section 3.3.3.

In human head motions, target directions of head motions depend on wills. In maximum cases, the front direction of head moves to the target direction. In minimum cases, the head moves to make the angle between the front and the target direction 30 degrees [2].

For the virtual human, the larger attractiveness (see section 3.3.2 and section 3.3.3) means more important gaze and generate larger head motions. The target of the head motion of virtual human is decided as followings:

\[
\theta_{\text{head}} = \theta_{\text{attention}} - \frac{30}{1 + e^{-(\frac{2\theta_0 - \theta_{\text{max}}}{A_2 - A_1})}} [\text{deg}] 
\]

where

- $\theta_{\text{attention}}$: direction of gaze target
- $A_1$: a constant for minimum motion of head = 1.0
- $A_2$: a constant for maximum motion of head = 3.0
- $S$: a constant for sharpness of motion = 7.0

The framework generates smooth head motion of the virtual human based on the minimum jerk model and PD control like section 3.5.

### 4 Virtual Boxing Application

We create a virtual boxing application as an example of the proposed framework. The application creates reactions of the eyes, the head and the upper part of the virtual human’s body such as attacking, blocking and dodging motion.

Following sections explains the implementation of the application.

### 5 Evaluation

We evaluate the proposed framework and the virtual boxing application. We use a PC with a processor of Pentium4 3.2[GHz] for the evaluation.

We asked to four subjects to play the virtual boxing application. Figure 2 shows a scene of the interaction. We tried cutting off the gaze model to see the effect of the gaze model. Subjects reported
that the gaze model makes the virtual human more real and attractive. Subjects also reported that the force feedbacks helped to know the hits of the attacks of both the virtual human and the user.

5.1 The Balance of the Top-down and Bottom-up attention
Most subjects playing the virtual boxing reported that the gaze of the virtual human moved too frequently and looks curious and less concentrated. Therefore, we tune the value of the attractiveness for the top-down attention. When the attractiveness for the top-down attention is small, the virtual human looks around more and looks less concentrated. When the attractiveness for the top-down attention is large, the gaze of the virtual human concentrates to the target object and looks strong. Figure 3 shows the difference on the gaze between different attractiveness values.

6 Conclusion
In this paper, we proposed a framework for realtime interaction with a reactive virtual human with gaze model. We employed human scale projection display and haptic interface for the interaction. The virtual human in the framework has an ocular system model to generate gaze and recognize the virtual world. The ocular system model has both bottom-up and top-down attention models. The virtual human predicts the state of the virtual world from the information recognized via the ocular system and decides the next behavior.

We created a virtual boxing system as an application of the proposed framework. Players report that reactions with ocular system model and gaze are more attractive than reactions without it. In addition, by tuning the attractiveness of the target object, the ocular model generates both concentrated and curious eyes.

References