Enhancing Motion Data with Head and Eye Motion

Heegun Lee

School of Computer Science

Carnegie Mellon University

A thesis submitted in fulfillment of the requirements for the Undergraduate Research Program

Abstract

Motion Capture is a widely used tool for synthesizing character animations that exhibit natural body motion. Human data for a variety of subjects, behaviors, and activities is becoming increasingly available. However, the vast majority of human body motion data collected only includes data for the larger body parts, such as the torso, arms, and legs. Finer details such as facial expression and eye movements are almost always missing and often the entire head and neck motion is reduced to that of a single rigid body. These simplifications sacrifice many subtle but important motion details that are present in an actor's face, particularly the gaze direction of the eyes. Omitting these details in an animation results in motion that looks "dead" and unrealistic as the characters perform the activities with a fixed stare.

We present a simple model that can be used to augment existing motion data with gazing behavior to produce enhanced motions that include realistic head and eye motion. Our gaze model can be easily used to create more lifelike characters that are responsive to objects in their environments. We show results of human motion data augmented with head and eye motions that yield significantly improved animations when compared with the original motion data.

Contents

1 Introduction **1**

2 Gaze Model **3**

2.1 Eye Structure 4

2.2 Head-Eye Coordination 6

2.3 Visual Attention 7

3 Implementation **8**

3.1 Procedure 9

3.2 Details 10

3.3 Results 10

4 Conclusions **16**

References **18**

List of Figures

2.1 Structure of the Human Eye 5
2.2 Structure of the Model Eye 5

3.1 calculating the modified head and eye motions 10
3.2 A comparison of the original and modified motions 12
3.3 The eyes track the red ball 12
3.4 The character follows the flight of the bee. 13
3.5 Multiple characters reacting to dynamic obstacles. 14
3.6 Multiple characters in a busy street environment. 15

Chapter 1

Introduction

Natural human motion involves the coordinated movement of numerous body parts. Manually modelling human motion is a difficult process and prohibitively time consuming due to the large number of degrees of freedom that must be specified. Motion capture systems provide a way to digitally record the motion of real human actors in the form of skeletal motions. These skeletal motions can readily be used to generate realistic character animations.

Realistic character animations that exhibit natural human motion are becoming desirable in a widening variety of applications. Recent films have used virtual characters to achieve special effects or in situations where it is dangerous or costly to employ real human actors. Video games have traditionally made use of virtual characters to populate immersive environments. Virtual avatars are being used to provide user-friendly interfaces to computerized systems such as automated help desk attendants.

Exhibiting natural motion greatly enhances the visual realism of these character animations. Motion capture is increasingly being chosen as a effective and efficient method of obtaining motion data for this use. However, there are limitations to the motion capture systems that are currently in wide use.

Conventional motion capture systems focus on capturing the gross motion of the large body parts. Finer details such as facial expression and gaze direction are generally ignored due to the limitations in capturing such small motions on the same scale with the larger motion of the gross body parts. These missing details however, become readily apparent in the produced animations with an unnatural rigidness of the head and face.

Different approaches have been taken to address this problem. Specialized systems have been devised to capture facial expression at close range using a large number of control points. The "Eyes Alive" Lee *et al.* (2002b) system used captured eye data to generate realistic eye motions for virtual avatars based on a statistical model. Deng *et al.* (2005) applied texture synthesis methods in generating eye motion and blinking effects. Neurobiological models of visual attention have been developed to simulate human gazing behavior Itti *et al.* (2003). However, these systems have their own set of limitations and optimal contexts.

Conventional motion capture systems also have difficulty in capturing motion involving environmental interaction. The majority of current motion capture systems are optical-based, where multiple cameras are used to record the position of reflective markers attached to an actor's body. The presence of environmental objects introduces occlusions that are problematic for such systems. In addition, it becomes prohibitively costly and time consuming to capture specific motions for every desired environmental configuration.

In this thesis we present a model that focuses on enhancing deficient motion data with head and eye motions that are consistent with human gazing behavior in dynamic environments. The gaze model that we use is simple and efficient and can easily be used to improve the realism of characters placed in user-defined environments. We show results of our model where the augmented motions substantially improve upon the original motion.

Chapter 2

Gaze Model

Fully simulating the process of visual perception (*synthetic vision*) is the ultimate method for animating a character to react based on what it apparently perceives. The key challenge is to devise a practical scheme for simulated perception that realistically reflects the character's perceptive capabilities and typical behaviors.

The remainder of this chapter introduces our model of gaze. We consider physiological and physcological aspects of human gaze in the context of creating an efficient model that can be used to augment motion capture data with simulated gazing behavior.

2.1 Eye Structure

Figure 2.1 shows a schematic diagram of the human eye. Each of the components that constitute the eye plays a crucial role in human vision. We introduce only the relevant parts that were used in our model.

The retina consists of over 100 million light sensitive cells. The majority of these cells are *rods*, which are sensitive to small intensities of light but unable to distinguish color. The remaining consists of *cones*, which can distinguish fine detail and the colors of the visible spectrum. The cone cells are densely concentrated in a small central area of the retina called the *fovea*. Accurate vision-the ability to distinguish fine detail, is limited to this foveal region, which comes to roughly 2^o about the *visual axis*, an imaginary line drawn through the pupil to the center of the fovea.

A small area accuity forces the eyes to rigourously move in order to process the whole field of view. These eye rotations can be extremely fast with speeds reaching up to $900^o/sec$.

We model the eyes as spheres that are attached to the head at fixed positions. The eyes can rotate about the horizontal and vertical axis, with 45^o being specified as the maximum angle of rotation. We simplify the foveal area to a single point, so that the area of visual accuity is reduced to a to a line. Given that the eyes are fixated at a point in space, this assumption allows us to derive eye orientations using simple 3D vector calculations instead of volumetric intersections as will be shown later.

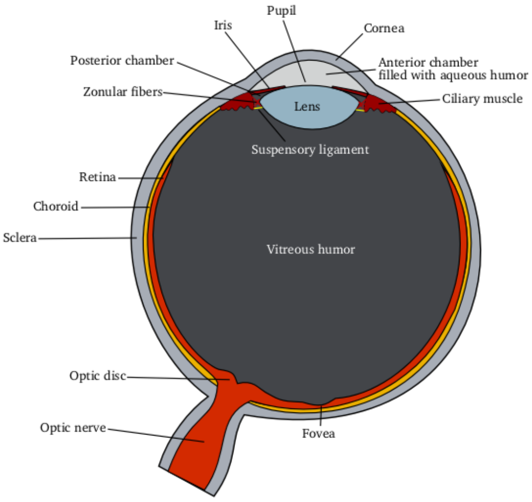

Figure 2.1: Structure of the Human Eye

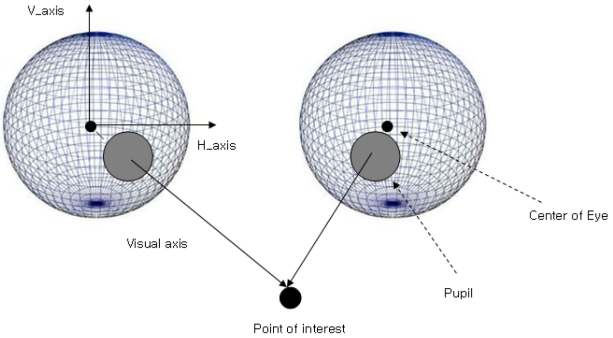

Figure 2.2: Structure of the Model Eye

2.2 Head-Eye Coordination

Human gaze can be simplified to a series of *fixations*. The size of the foveal region restricts the area of high-resolution visibility to a small area, called the point of fixation. A change in point of interest induces a gaze shift from the original point of interest to the new one. This gaze shift is achieved through coordinated rotations of the head and eyes. The eyes normally lead with a fast rotation towards the new point of interest and the head quickly follows behind. When the gaze shift is completed, the eyes are oriented such that their visual axes are aligned with the new point of interest. The exact ratio between the head and eye rotations composing the gaze shift is hard to define generally. We left this is as a user-defined parameter so that the head rotation composes a fraction of the total gaze shift. The eye rotations are calculated after the orientation of the head is determined.

2.3 Visual Attention

Human visual attention is complex and difficult to model. We simplify visual attention to within the context of the user-defined environment, with the aim of producing reactive gazing behavior to moving objects. A character can be in one of two states. The *idle* state is when the enviornment does not contain any interesting ojects within the character's field of view. The *fixated* state is when the eyes are focused at an object in the environment.

When there are multiple objects in the field of view we apply a weighting function to each object to derive the fixation point. This function currently incorporates the size, distance, and velocity of the object relative to the character.

$$Obj_{weight} = k1 * Obj_m + k2 * Obj_d + k3 * Obj_v \tag{2.1}$$

The object with the greatest weight is taken to be the point of fixation.

Chapter 3

Implementation

Our system takes two sources of input. The first is a skeletal motion file which is the original, unmodified motion. We based our implementation on the Acclaim motion data format for its wide use and availability. The Acclaim format uses two seperate files to specify a skeletal motion. The ASF(Acclaim Skeleton File) file specifies the configuration of the skeleton with the hierarchy and properties of the bones composing the skeleton. The AMC(Acclaim Motion Capture) file is a sequence of *postures*, where a posture fully specifies the rotation values of each movable bone in the skeleton for a given frame. The second input to our system is a text file that specifies the environmental information. Environmental information consists of the position and size of all objects in the scene. This information is specified for each frame of the original motion. Dynamic objects can be represented by a change of position over a sequence of frames. The output of our system is a new motion file that incorporates modified head rotations and a seperate text file that contains the eye rotation values for each frame.

3.1 Procedure

The modified head and eye motions are calculated for each individual frame of the original motion using a set procedure. Figure 3.1 shows a sequence of steps in this procedure. Note that the red, green, and blue lines consititute the local coordinate axes of the head.

First, the local coordinate system of the head is derived from the original motion file. This is equivalent to computing the orientation of the head and involves sequentially multiplying the transformation matrices of all the joints lying on the path from the root to the head.

$$R_{head} = R_{upper_{n}eck} * R_{lower_{n}eck} * ... * R_{lower_{b}ack} * R_{root}; \tag{3.1}$$

The visual attention model is then applied to all objects in the character's field of view to derive the fixation point. This is represented by the yellow ball to the right of the character. The direction vector from the head to the fixation point is decomposed into a horizontal and vertical component using the local coordinate system of the head. The angle is then multiplied with a fractional constant and applied to the head. This gives us the modified head motion.

$$head_{mod} = head_{original} + k * \theta_{obj} \tag{3.2}$$

Now, we have the final position of the eyes as well as the modified orientation of the head. We may easily derive the eye rotations by calculating the angle from each eye to the object.

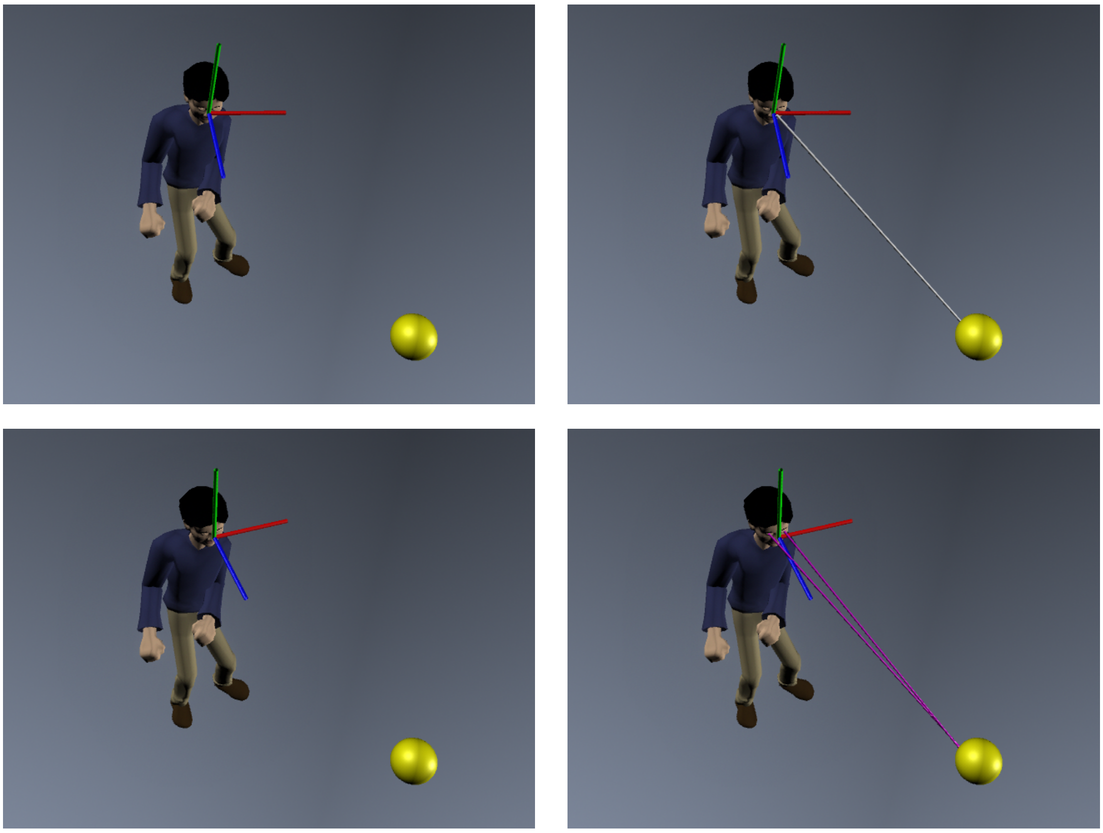

Figure 3.1: calculating the modified head and eye motions

3.2 Details

A custom viewer program for importing motion data and also exporting modified motion data was written in C++ using the OpenGL API for visual representation. FLTK was used to create a user interface window. The component for generating the modified motions was integrated into the viewer program so that the modified motion may be viewed in real-time. The program was run on a 1.8 GHz Pentium laptop with 512MB of ram.

3.3 Results

In order to test the effectiveness of our model we ran it through a series of example inputs that widely varied in original motion and environment scenarios. Our model was able to generate the modified motions in real-time while simultaneously displaying the simplified results. The example figures that follow were rendered offline in MAYA using the modified motions.

Figure 1 and Figure 2 relate to the first example where the original motion consists of a character swinging a stick with his head in a fixed position. By running our model with the ball at the end of the stick set as an interest point, the character realistically tracks the ball with coordinated movements of the head and eyes. Figure 3 shows an example where character's body is set in a fixed posture. By setting the bee in flight as an environmental object, the character accurately follows the flight of the bee with head and eye motions. Figure 4 shows an example of a group of characters placed in a dynamic enviornment composed of rolling spheres. The characters take note of the moving obstacles that pass them by. Figure 5 shows another group example with sixteen characters moving simultaneously in a simulated busy street. The characters pay attention to the car that is crossing the intersection.

The generated head and eye motions for the examples were effective in giving the impression that the characters were alert to changes in their environment. The increase in realism was especially noticeable in close-up shots of the face where the consistent motion of the eyes were noticable. The simplicity of our model enables us to generate the modified motions in real-time on a Pentium 3 machine with 512MB of ram. We believe this model could be used effectively in interactive applications such as video games. However, it can also be used as a postprocessing tool to augment captured motions that are missing data for the eyes and head.

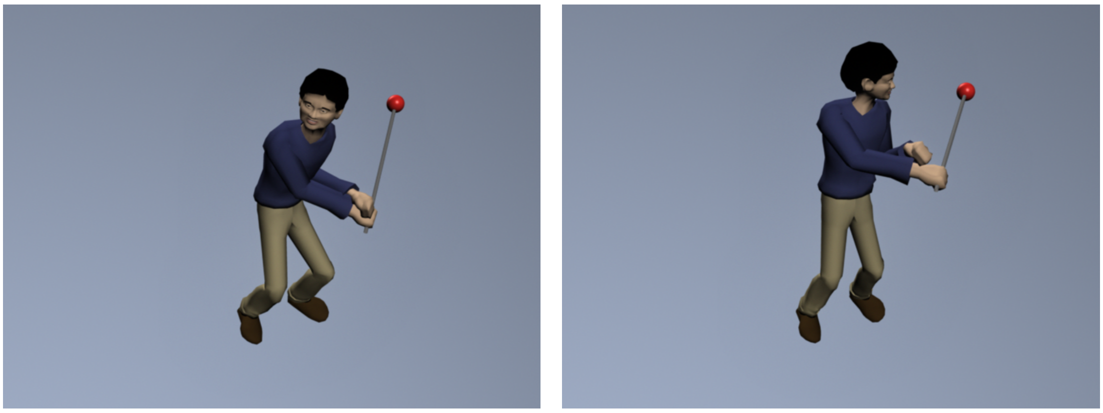

Figure 3.2: A comparison of the original and modified motions

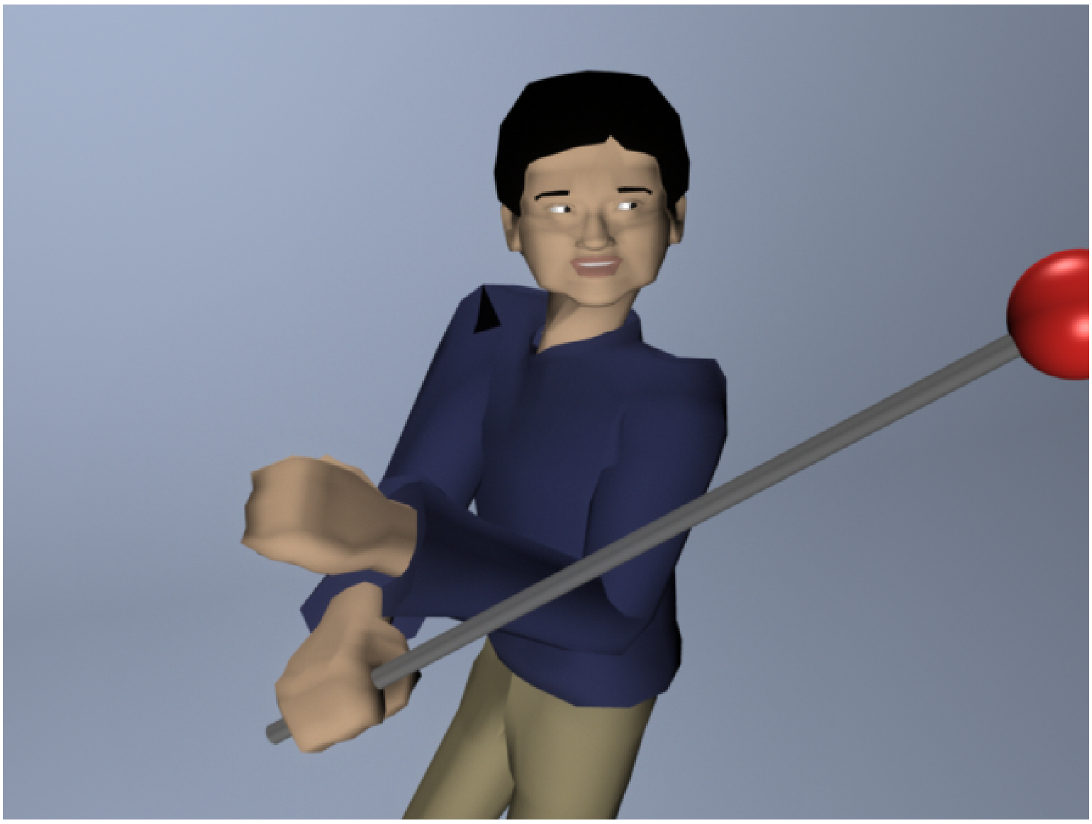

Figure 3.3: The eyes track the red ball

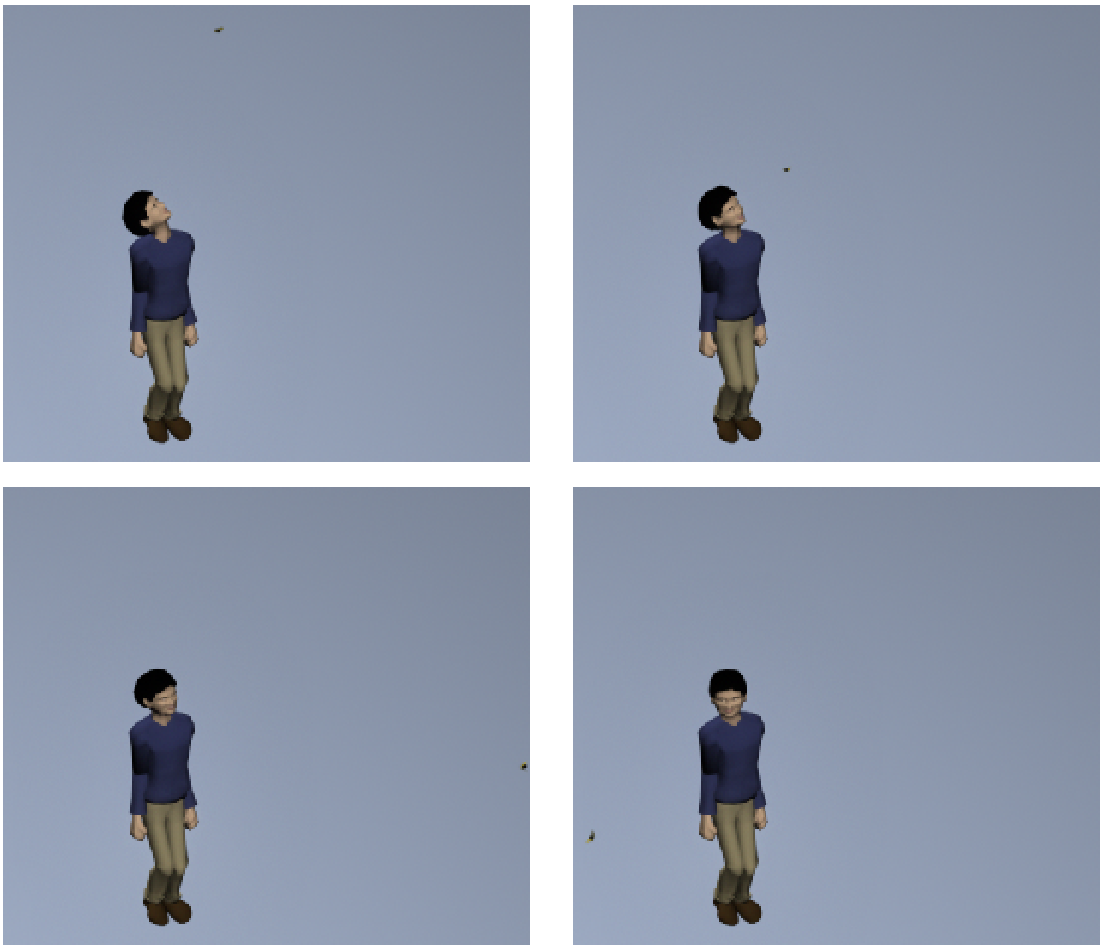

Figure 3.4: The character follows the flight of the bee.

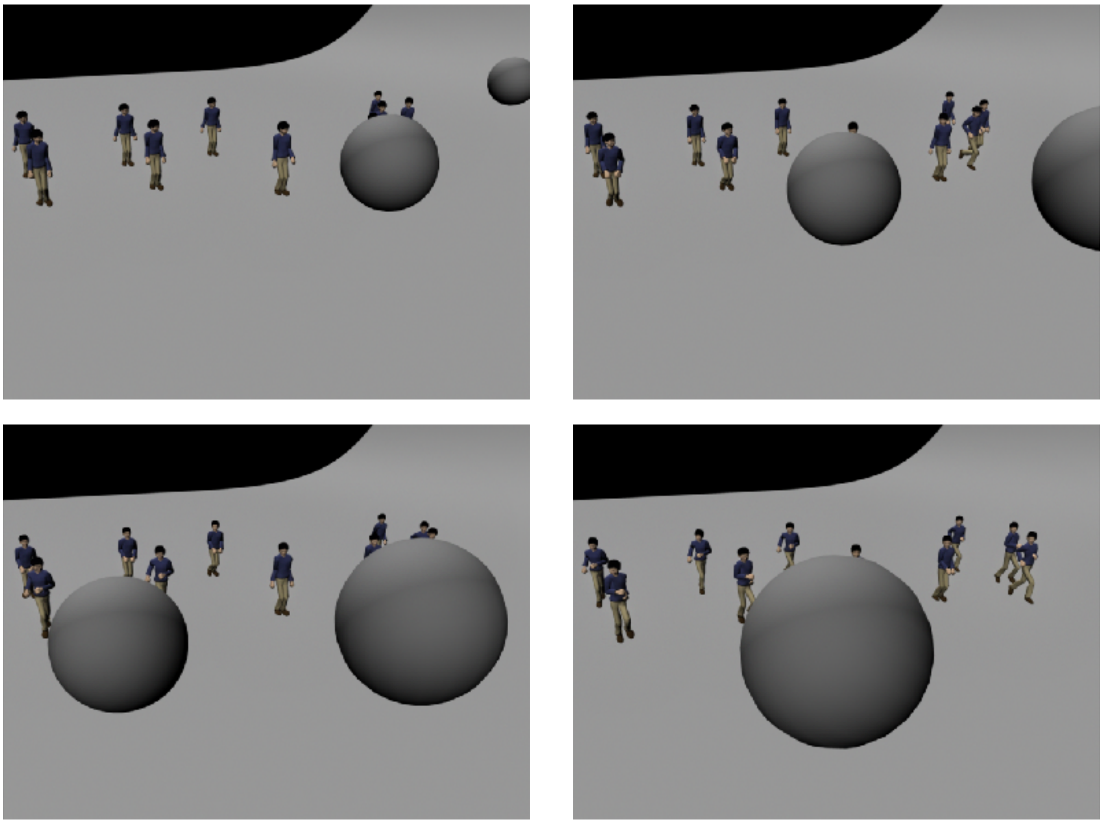

Figure 3.5: Multiple characters reacting to dynamic obstacles.

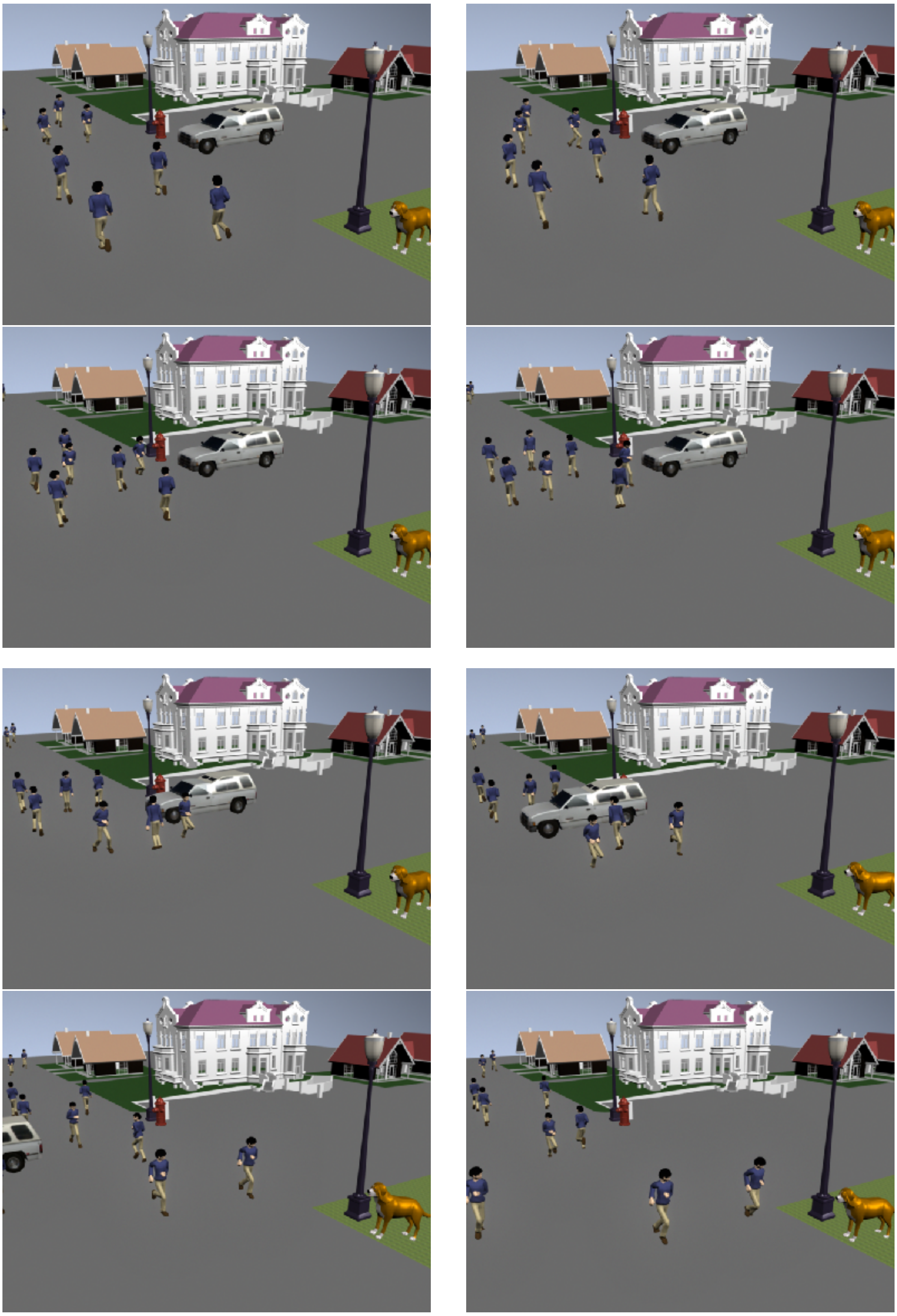

Figure 3.6: Multiple characters in a busy street environment.

Chapter 4

Conclusions

We have presented a simple model that generates coordinated head and eye motions consistent with human gazing behavior in user-defined environments. The addition of gazing behavior was found to significantly improve the realism of character animations. Many improvements to the basic model are possible. Using a more sophisticated visual perception and attention model that includes information filtered for the entire field of view, including proper occlusions and a model of object semantics would likely yield better results. In addition, incorporating existing gaze data and visual attention statistics gathered from real human subjects and biasing object attention to create behavior-specific models would be very interesting as future work.

References

DENG, Z., LEWIS, J. & NEUMANN, U. (2005). Automated eye motion using texture synthesis. *IEEE Computer Graphics and Applications*, **25**, 24–30. 2

GLEICHER, M. (1997). Motion editing with spacetime constraints. In *1997 Symposium on Interactive 3D Graphics*, 139–148.

GLEICHER, M. (1998). Retargeting motion to new characters. In *Proc. ACM SIGGRAPH 98 (Annual Conference Series)*, 33–42.

ITTI, L., DHAVALE, N. & PIGHIN, F. (2003). Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Proc. SPIE Int. Symp. on Optical Science and Technology*, vol. 5200, 64–78. 2

KOGA, Y., KONDO, K., KUFFNER, J.J. & LATOMBE, J.C. (1994). Planning motions with intentions. In *Proceedings of SIGGRAPH 94*, 395–408.

LEE, J. & LEE, K.H. (2004). Precomputing avatar behavior from human motion data. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 79–87, ACM Press.

LEE, J., CHAI, J., REITSMA, P.S.A., HODGINS, J.K. & POLLARD, N.S. (2002a). Interactive control of avatars animated with human motion data. *ACM Transactions on Graphics*, **21**, 491–500.

LEE, S.P., BADLER, J.B. & BADLER, N.I. (2002b). Eyes alive. *ACM Transaction on Graphics*, **21**, 637–644. 2

Palmer, S.E. (1999). *Vision Science: Photons to Phenomenology*. The MIT Press.

Salvucci, D. & Goldber, J. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Eye Tracking Research and Applications Symposium*, 71–78.

xiang Chai, J., Xiao, J. & Hodjins, J. (2003). Vision-based control of 3d facial animation. *Eurographics/SIGGRAPH Symposium on Computer Animation*.

Zhang, L., Snavely, N., Curless, B. & Seitz, S.M. (2004). Shape and motion: Spacetime faces: high resolution capture for modeling and animation. *ACM Transactions on Graphics(TOG)*, **23**.