Inferring Gaze Shifts from Captured Body Motion

Daniel Rakita, Tomislav Pejsa, Bilge Mutlu, Michael Gleicher
University of Wisconsin–Madison
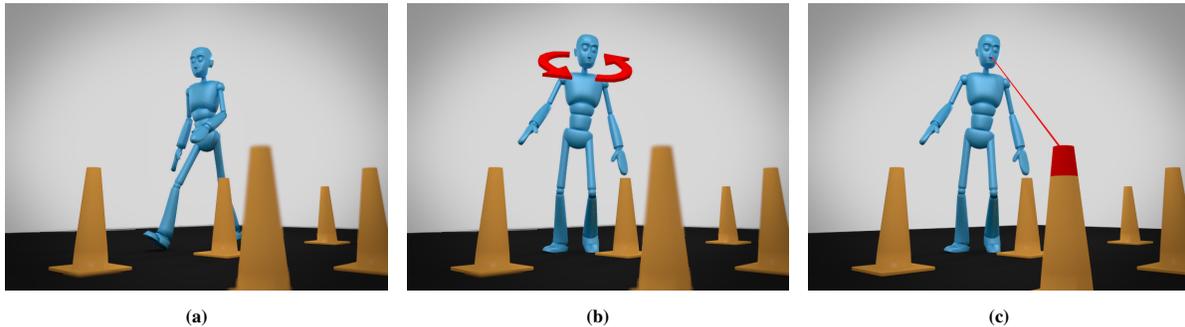
Figure 1: *(a) Our system takes as input a captured body motion and a virtual environment. (b) We identify gaze shifts and fixations in the motion by analyzing the kinematics of head and torso movements . (c) We estimate the gaze target in the environment by casting a ray in the gaze direction*

Keywords: gaze, motion capture, analysis, bilateral filter

1 Introduction

Motion-captured performances seldom include eye gaze, because capturing this motion requires eye tracking technology that is not typically part of a motion capture setup. Yet having eye gaze information is important, as it tells us what the actor was attending to during capture and it adds to the expressivity of their performance.

In this paper, we propose an approach for automated inference of gaze shifts and fixations by analyzing the kinematic properties of a captured body motion. Our focus is on gaze shifts, the intentional movement of the eyes, head, and upper body toward specific targets. Our system accurately infers gaze shift timings and gaze fixation targets . Applications of the inferred information include hand-authoring and automated synthesis of the missing gaze behavior, as well as multimodal analysis of the captured behavior.

2 Our Approach

Our approach takes as input a standard skeletal motion file and optionally a model of a virtual environment. The body motion is a sequence of quaternion orientations specified for each bone in the character's skeleton. The output of our system is a list of gaze annotations, consisting of gaze shift start frames, gaze fixation start frames, gaze fixation durations, and target objects.

Human gaze shifts follow a particular kinematic pattern [Pejsa et al. 2015]. The eyes, head, and torso joints accelerate jointly toward the target, reach some peak angular velocity ω_{max}, and decelerate to a local minimum velocity ω_{min} as they reach the target and begin the fixation. According to [Pejsa et al. 2015], head and torso velocities

start at about $\frac{1}{4}\,\omega_{max}$. In our approach, we infer the timings of gaze shifts and fixations by searching for occurrences of this kinematic pattern in the head and torso velocity signals.

We represent head orientation in the coordinate frame of the topmost torso joint and torso orientation in the coordinate frame of the body root, and we compute their angular velocities in those frames. To provide a stable estimate of the velocity and preserve abrupt changes in gaze movement direction, we filter the rotations using a bilateral filter applied to the rotation vectors [Lee and Shin 2002]. We analyze the velocities of the head and torso to identify matching occurrences of the gaze shift kinematic pattern and infer gaze shift and fixation timings.

When a gaze fixation is detected, the system infers an estimate of the gaze target location. We cast a ray in the forward direction of the head and infer the target location by finding an intersection with the environment or closest point on an environment object.

We have implemented our approach in a testbed application and have successfully used it on a range of example motions. To evaluate our approach, we selected a set of seven motions including locomotion tasks and multi-character interactions. An animator hand-annotated gaze shift timings and targets to provide a baseline for comparison. Our system provides similar results on all examples. We use the the proposed inference approach as a component of an automated gaze synthesis system.

References

LEE, J., AND SHIN, S. Y. 2002. General construction of time-domain filters for orientation data. *IEEE Transactions on Visualization and Computer Graphics 8*, 2, 119–128.

PEJSA, T., ANDRIST, S., GLEICHER, M., AND MUTLU, B. 2015. Gaze and Attention Management for Embodied Conversational Agents. 1–34.