Vision-Based Reaching for Autonomous Virtual Humans

Christopher Peters, Carol O' Sullivan
Image Synthesis Group, Trinity College, Dublin 2, Republic of Ireland
email: {christopher.peters, carol.osullivan}@cs.tcd.ie

Abstract

A method for the generation of realistic real-time goal-directed virtual human arm motion is presented. Agents are endowed with a rudimentary synthetic vision and memory system that is used to gather and store data about objects in the vicinity. Agents then use the perceived object data rather than global database object data for the planning of reaching arm motions. Our method differs from previous attempts at goal-directed motion generation in that it uses sensory information available to the agent in order to distinguish between movements towards those objects visible to the agent, and movements towards memorised object locations. The generation of appropriate arm configurations under these circumstances is based on results from neurophysiology.

1 Introduction

This paper combines a synthetic vision and memory model, incorporating results from neurophysiology, to compute realistic arm configurations for reaching towards objects and the remembered locations of objects. We distinguish between objects that are out of view and those that are in view since neurophysiological results have shown that the computation performed by humans varies under these two conditions (see Soechting and Flanders 1989).

The method presented in this paper is applicable to real-time virtual environments for the realistic animation of agent-object interactions. For example, consider a bar scene. An agent is sitting at a table enjoying a drink. Because the glass is visible to our agent, reaching motions towards the glass are accurate. Suddenly a friend walks over, and our agent becomes distracted in a riveting conversation. During the conversation, our agent decides to reach for the glass, which is no longer inside its visual field (our agent is looking at his/her friend). Under these circumstances, the agent's reach is less accurate, since the agent is reaching towards a remembered location. It is also possible that the barman has taken the glass, or that the glass has been moved. It would be more realistic to have our agent look back at the remembered position of the glass, in order to visually establish that it is still there, and also to provide a correction to the reach in order to guide the hand towards the glass with precision.

2 Related Work

The generation of natural arm motions and implementation of synthetic vision have been applied to the field of computer graphics, albeit separately.

Lee at al. (1990) simulated lifting motions based on comfort, strength and perceived exertion factors. Tolani and Badler (1996) propose an analytic inverse kinematics based approach. Their basic strategy is to reduce the degrees-of-freedom of the arm by one in order to obtain the closed-form equations for solving the inverse kinematics problem. As the authors have pointed out, a shortcoming of this method is that it does not have a well-defined theory for generating natural looking configurations. Kondo presents an inverse kinematics algorithm based on a sensorimotor model to generate natural looking arm postures. Since the sensorimotor model is approximate, an iterative pseudo-inverse Jacobian method is used to perform a final adjustment. However, no distinction is made between visible objects and virtual (remembered) objects: the sensorimotor model for remembered objects is used in all circumstances. Nebel (2000) uses a sensitivity model from neuroscience for realistic "cautious" collision free motion of the arm.
Numerous researchers have suggested the use of a virtual model of perception in order to permit agents to

perceive their environment (see Renault et al. 1990; Tu 1996). An early example applies group behaviours to simulated creatures (Reynolds 1987). Tu and Terzopoulos (1994) implemented a realistic simulation of artificial fishes. Noser et al. (1995) proposed a navigation system for animated characters using synthetic vision and memory. Kuffner and Latombe (1999) provide real-time synthetic vision, memory and learning, and apply it to the navigation of animated characters.

3 Synthetic Vision and Memory

Our synthetic vision module is based on the model described by Noser et al. (1995). This model uses *false-colouring* and *dynamic octrees* to represent the visual memory of the character. We adopt a similar system to Kuffner and Latombe, where the speed of the memory component is increased by removing the octree structure, replacing it with a vector containing object observation information.

The process is as follows. Each object in the scene is assigned a single, unique colour that identifies it. The rendering hardware is then used to render the scene from the perspective of each agent. The frequency of this rendering may be varied, and could be used to implement attention levels for agents. In this mode, objects are rendered with flat shading in their unique false-colour. No textures or other effects are applied. The agent's viewpoint does not need to be rendered into a particularly large area: our current implementation uses 128x128 renderings. The false-coloured rendering is then scanned, and the object false-colours are extracted. These colours are then used to do a look-up of the objects in the scene, and object state information is extracted and stored with the agent in the form of observations. Essentially, an agent's visual memory consists of a list of these observations.

In our implementation, the precise position of an object in the environment is not stored as part of an observation. Rather, an approximation of the object's location in spherical coordinates with respect to the agent's viewing frame is used. During the scanning process, bounding boxes are assembled for each object based on the object's minimum and maximum x and y coordinates extracted from the view specific rendering, and the object's minimum and maximum z coordinates extracted from the z-buffer for that view. The object's position is then estimated to be the centre of this bounding box. This process has the overall effect of making accurate judgements about the positions of partially occluded objects more difficult. Also, estimates made about the distance to the centre of the object will vary depending on the obliqueness of the object with respect to the viewer.

Observations, then, are represented as tuples that are composed of the following components:

objID	globally unique identifier of the object
objAzi	azimuth of object
objEle	elevation of object
objDis	distance to object
t	time stamp

A specific object will have at most a single observation per agent. The observation will match the last perceived state of the object, although it must be noted that this may not correspond with the actual current state of the object. Also, since object coordinates are stored with respect to the viewing frame of the agent, they need to be updated whenever the direction of the view frame changes. This is not necessarily a problem; in our implementation observations are assumed to be composed of something analogous to iconic memory. Based on a number of heuristics, objects of particular interest would be converted from egocentric coordinates to world-space coordinates and placed in something more similar to a short-term or long-term memory system.

4 Reaching

The term 'sensorimotor transformation' refers to the process by which sensory stimuli are converted into motor commands (Pouget and Snyder 2000). Reaching towards a visual stimulus with the hand is an example of a sensorimotor transformation.

Research in neurophysiology suggests that there is a reasonably accurate internal representation of a target's location in a body-centred frame of reference (see Soechting and Flanders for references). The coordinates of this representation are referred to as extrinsic coordinates. There also appears to be an internal representation for the orientation of the upper arm and the forearm. These are referred to as intrinsic coordinates, and allow the calculation of the hand position.

Under normal, visually guided circumstances, the mathematically exact relationships between intrinsic and extrinsic coordinates were found to be highly non-linear (Soechting et al. 1986). However, research shows that the relations between intrinsic and extrinsic coordinates are close to linear when subjects point at virtual target locations that are based on visual memory (Soechting and Flanders 1989). It was hypothesised that these movements would be less accurate than those made towards visible objects because of the use of a linear approximation to the more accurate, but non-linear, sensorimotor transformation.

Our reaching module uses this research and a human arm inverse kinematics algorithm that is based on it (see Kondo). However, the method presented here distinguishes between reaching that takes place toward a visible object, and reaching that takes place towards a virtual (remembered) object.

The shoulder coordinate frame is centred on the shoulder. The x-axis of the frame is along the line connecting both shoulders, the y-axis goes outward from the chest, and the z-axis points upward toward the head. Arm posture is encoded in the shoulder frame by four parameters: θ is the upper arm elevation, β is the forearm elevation, η is the upper arm yaw, and α is the forearm yaw. The extrinsic coordinates are defined in terms of spherical coordinates with the origin at the shoulder. These coordinates consist of the target azimuth χ, the target elevation ψ, and the radial target distance R.

Given these definitions, it was found that the following relations between intrinsic and extrinsic object coordinates describe a pointing movement towards a remembered target location:

$$\begin{aligned} \vartheta &= -4.0 + 1.10\,R + 0.90\,\Psi \\ \beta &= 39.4 + 0.54\,R - 1.06\,\Psi \\ \eta &= 13.2 + 0.86\,\chi + 0.11\,\Psi \\ \alpha &= -10.0 + 1.08\,\chi - 0.35\,\Psi \end{aligned} \qquad \text{(Eq. 1)}$$

The following relations describe the approximation for visually guided movements toward a target location:

$$\begin{aligned} \vartheta &= -6.7 + 1.09\,R + 1.10\,\Psi \\ \beta &= 47.6 + 0.33\,R - 0.95\,\Psi \\ \eta &= 67.7 - 0.68\,R + 1.00\,\chi \\ \alpha &= -11.5 + 1.27\,\chi - 0.54\,\Psi \end{aligned} \qquad \text{(Eq. 2)}$$

The relations for visually guided reaches (Eq. 2) approximate a highly non-linear function. Because of this, the wrist will not be in the desired position in the final computed arm configuration. In order to move it into its final position, an iterative inverse kinematics algorithm is applied in a final correction phase. The inverse kinematics technique that we use is a heuristic search technique called cyclic coordinate descent (Welman 1993). At each iteration, the algorithm attempts to minimize position and orientation errors by varying one joint at a time. This technique is numerically stable and computationally efficient, since the Jacobian matrix does not need to be inverted. Computational efficiency is an important requirement since the reaching algorithm must work in a real-time environment.

5 Implementation

Our implementation of the reaching process is summarised as follows:

First, an agent is requested to reach for a specific object. If the object was originally within the agent's view, or falls within his/her view, the agent will reach towards the target using the accurate movement model (Eq. 2) and the correction phase.

In the case where the object is not within the agent's current view, but is in the memory, the agent will initiate a reach toward the remembered target location using (Eq. 1). As it does so, it will also attempt to bring the target location into view. If the target location is in view, but the object is not, the agent assumes the object has moved, retracts the arm and continues searching. If the object appears in view then the arm will be moved to the object location from its current position (i.e. mid-movement toward the inaccurate arm position) using the accurate movement model (Eq. 2).

If the object is not in the agent's view or its memory, then the agent will start a search routine. This search involves the agent moving its head around and scanning so that the view frustum falls within an area of 90 degrees either side of the forward vector of the his/her chest. Note that during a search, the agent's attention and observation abilities will be very active (essentially, the interval between internal view updates will be short). This is not necessarily always the case: under normal circumstances, the agent will have a lower attention level, characterised by more passive update intervals. Passive updates are useful for situations involving multiple agents, since the number of scene renderings from different viewpoints will be minimized.

This scenario was implemented on the ALOHA animation system (Giang et al. 2000), which uses the OpenGL API on a Windows platform.

6 Conclusions and Future Work

This paper combines a synthetic vision and memory model with results from neurophysiology to compute realistic arm configurations for reaching towards objects and the remembered locations of objects.

One drawback of the method presented is that the reaching algorithm does not take obstacles into account. Future work will need to implement motion planning into the system. Also, because an approximation of the accurate non-linear sensorimotor transformation is used, a correction phase is implemented. It is envisaged that a more appropriate sensorimotor transformation model will be used in the near future for the accurate reaching motions. Further work must also take place on the perception of events through the vision system. If an object is no longer at a remembered location, it is possible that the object has moved, or also that the object has been occluded by another object. In the current implementation, we assume that only the former situation is possible. We also intend to look at the idea of synthesizing different memory models, and implementing more types of virtual senses.

References

Giang T, Mooney R, Peters C, O'Sullivan C. ALOHA: adaptive level of detail for human animation, *Eurographics 2000*, Short Presentations, 2000.

Kondo K. Inverse kinematics of a human arm, Technical Report CS-TR-94-1508, Robotics Laboratory, Department of Computer Science, Stanford University, Stanford, USA.

Kuffner J, Latombe JC. Fast synthetic vision, memory, and learning models for virtual humans, *Proc. of Computer Animation*, IEEE, pages 118-127, 1999.

Lee P, Wei S, Zhao J, Badler NI. Strength guided motion, *Computer Graphics*, Vol. 24, No. 4, pages 253-262, 1990.

Nebel JC. Realistic collision avoidance of upper limbs based on neuroscience models, *Computer Graphics Forum* (EG2000 proceedings), Vol. 19, No. 3, 2000.

Noser N, Renault O, Thalmann D, Thalmann NM. Navigation for digital actors based on synthetic vision, memory and learning, *Computer Graphics*, 19, pages 7-19, 1995.

Pouget A, Snyder LH. Computational approaches to sensorimotor transformations, *Nature Neuroscience supplement*, Volume 3, pages 1192-1198, 2000.

Renault O, Thalmann NM, Thalmann D. A vision-based approach to behavioural animation, *Visualization and Computer Animation*, Vol. 1, pages 18-21, 1990.

Reynolds CW. Flocks, herds and schools: A distributed behavioural model, *Computer Graphics*, 21(4), pages 25-34, 1987.

Soechting JF, Flanders M. Errors in pointing are due to approximations in sensorimotor transformations, *Journal of Neurophysiology*, Vol. 62, No. 2, pages 595-608, 1989.

Soechting JF, Lacquaniti F, Terzuolo CA. Coordination of arm movements in three-dimensional space. Sensorimotor mapping during drawing movement, *Neuroscience*, 17, pages 295–311, 1986.

Tolani D, Badler NI. Real-time inverse kinematics of the human arm, *Presence* 5.4, pages 393-401, 1996.

Tu X. Artificial animals for computer animation: biomechanics, locomotion, perception, and behaviour, PhD thesis, University of Toronto, Toronto, Canada, 1996.

Tu X, Terzopoulos D. Artificial fishes: Physics, locomotion, perception, behaviour, *Proc. SIGGRAPH '94*, pages 43-50, 1994.

Welman C. Inverse Kinematics and Geometric Constraints for Articulated Figure Manipulation, Masters Thesis, 1993.