

Computer Puppetry: An Importance-Based Approach

HYUN JOON SHIN, JEHEE LEE, and SUNG YONG SHIN

Korea Advanced Institute of Science & Technology

and

MICHAEL GLEICHER

University of Wisconsin—Madison

Computer puppetry maps the movements of a performer to an animated character in real-time. In this article, we provide a comprehensive solution to the problem of transferring the observations of the motion capture sensors to an animated character whose size and proportion may be different from the performer's. Our goal is to map as many of the *important* aspects of the motion to the target character as possible, while meeting the online, real-time demands of computer puppetry. We adopt a Kalman filter scheme that addresses motion capture noise issues in this setting. We provide the notion of dynamic importance of an end-effector that allows us to determine what aspects of the performance must be kept in the resulting motion. We introduce a novel inverse kinematics solver that realizes these important aspects within tight real-time constraints. Our approach is demonstrated by its application to broadcast television performances.

Categories and Subject Descriptors: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—*animation*

General Terms: Algorithm

Additional Key Words and Phrases: Human-figure animation, motion retargetting, performance-based animation, real-time animation

1. INTRODUCTION

Computer puppetry [Sturman 1998] transforms the movements of a performer to an animated character in real-time. The immediacy of computer puppetry makes it useful for providing live performances and as a visualization tool for traditional cinematic animation. However, this immediacy creates a number

The work of H. J. Shin, J. Lee, and S. Y. Shin was supported by the NRL (National Research Laboratory) program of KISTEP (Korea Institute of Science and Technology Evaluation and Planning) and KBS (Korean Broadcasting System).

The work of M. Gleicher was supported by National Science Foundation grant CCR-9984506, Microsoft Research, and equipment donations from IBM and Intel.

Authors' addresses: H. J. Shin, J. Lee, and S. Y. Shin, Division of Computer Science, Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, Taejeon, Korea, e-mail: {joony,jehee,syshin}@jupiter.kaist.ac.kr; M. Gleicher, Department of Computer Sciences, University of Wisconsin—Madison, Madison, WI, e-mail: gleicher@cs.wisc.edu.

Permission to make digital/hard copy of part or all of this work for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee.

© 2001 ACM 0730-0301/01/0400-0067 \$5.00

of challenges, as solutions to animation issues must be handled in an online real-time manner. A computer puppetry system must capture the movements of the performer, interpret the important aspects of this motion, and determine the movements required to make the character reproduce these important aspects of the performance.

The challenges of mapping a motion from the performer to the target character become more difficult when the target character is of a different size and proportion than the performer [Badler et al. 1993; Bodenheimer et al. 1997; Choi and Ko 2000, Gleicher 1998]. In such cases, the resulting motion of the character cannot exactly duplicate the original performer's. For example, we cannot simultaneously match the original joint angles and end-effector positions. Generally, to preserve the important aspects of the original motion we must alter the unimportant aspects of the motion. This process of adapting a motion for a new character is called *retargetting* [Gleicher 1998; Lee and Shin 1999].

To date, solutions to computer puppetry issues have been limiting: either restricting the range of puppets that can be used, or providing restrictive notions of what is important in motions. The latter implicitly limits the range of puppets since artifacts are introduced as the differences of the puppet from the performer are increased.

In this article we provide techniques that address the challenges of computer puppetry when the target character is different from the performer. The following major animation issues are addressed in a manner that fits within the online real-time nature of computer puppetry.

- (1) The sensors used to capture the performer's motion are often noisy. Therefore, we provide a filtering technique that operates in an online manner with the efficiency required to process whole body motions in real-time. We apply a Kalman filter to rotation vectors, providing an orientation smoothing technique that is more efficient than previous methods.
- (2) The important aspects of the original performance must be determined such that these details can be reproduced in the resulting motion. We provide the notion of a dynamic importance measure that allows us to account for changing situations even when the future is unknown.
- (3) The resulting pose of the target character must be computed in a way that recreates the important aspects of the original. We provide a fast inverse kinematics solver that provides the necessary real-time performance and predictability.

Our solutions have been used to realize a computer puppetry system that has been used successfully to create animated television broadcasts.

We begin our discussion of computer puppetry by providing an overview of our approach. We examine previous solutions with respect to the issues raised in the overview. The components of our approach are then detailed in Sections 3 through 5. An analysis in Section 6 reviews why our approach avoids introducing unwanted artifacts such as temporal discontinuities. Our experimental results are provided to support this approach. We conclude with a summary and discussion of future directions.

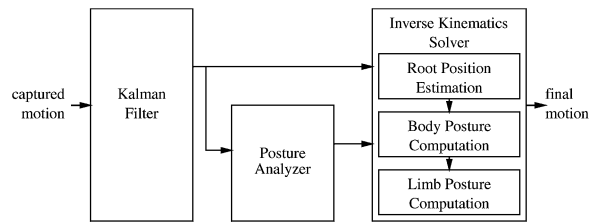


Fig. 1. Overall structure.

2. OVERVIEW

Computer puppetry requires the captured movements of the performer to be mapped to the target character in real-time. As shown in Figure 1, our approach for online motion retargeting divides the task into phases. First, the filtering phase “cleans” the sensor data to remove artifacts of the motion capture device. The second phase examines this filtered motion and determines the importance value of every end-effector in relation to its environment. The final phase computes a pose for the target character that achieves as many of the important aspects as possible. In this section, we provide an overview of these components and survey their relationship to previous work.

2.1 Online Filtering of Orientations

In general, captured motion data are noisy. The real-time sensors required for computer puppetry are particularly problematic in this regard. However, because of the dense sampling rates and signal characteristics of motion capture data, low-pass filtering is an effective tool to suppress noise in the captured data. This is challenging for several reasons.

- (1) Because computer puppetry is an online application, standard offline filters cannot be employed.
- (2) Because the orientation space is highly nonlinear, standard signal processing methods cannot be applied directly.
- (3) Because of the real-time demands, filtering should be performed on the entire body very efficiently.

A Kalman filter predicts the future values and corrects them in accordance with the actual input data to create a filtering scheme that operates in an online manner. The technique is common in online applications, and was first introduced to the graphics community by Friedman et al. [1991]. Such a filter cannot be directly applied to rotation data without accounting for the nonlinearity of the orientation space. To address this problem, Welch and Bishop [1997] linearized the orientation space by locally parameterizing the incremental orientation change with Euler angles, based on the result in Azarbayejani and Pentland [1995] and Broida and Chellapa [1986]. Because they were interested only in tracking the head motion, they were less concerned with efficiency than we are and therefore addressed only issues 1 and 2 above. In Section 3 we provide a modified Kalman filter. To achieve real-time performance, we locally

parameterize the incremental orientation changes with rotation vectors instead of the Euler angles used in Welch and Bishop [1997].

2.2 Importance Determination

The goal of computer puppetry is to create the movements of a target character based on the performer's movements. If the target character is quite different from the performer, there may not be a direct mapping. Indirect mappings are common in traditional puppetry; for example, a marionette is controlled by strings that pull on its end-effectors. Computer equivalents may create arbitrary mappings from sensor input to character parameters. For example, the Alive system from Protozoa [1999] allows arbitrary Scheme¹ functions [Hanson 2000] to be written to perform mapping.

Our interest is in recreating characters with human form, so the target character has equivalent degrees of freedom as a simplified model of a human performer. In this article, we consider characters that are articulated figures with identical connectivity, so that it is possible to transfer the captured joint angles directly to the target character. Despite this structural equivalence, the resulting motion will not match the performer's unless the character has an identical size and proportion. There will be some level of mismatching even for characters that have the same size and proportion as the performer, since we simplify the real human by a hierarchy of rigid bodies. One approach to performance animation, described by Molet et al. [1996, 1999], models the character to be as similar to the performer as possible. Bodenheimer et al. [1997] presented a way to determine the segment lengths of a character that best fit the captured motion data while discarding outliers in these data by a robust estimation technique. If the segment proportions of the character are kept the same as those of the performer, a motion adaptation can often be achieved by scaling the position data according to the size difference and then by translating the character globally. Restricting the proportions of the character precludes the use of stylized cartoon characters, unless we can find similarly proportioned performers.

When the virtual character and performer have different sizes and proportions, not all aspects of the motions can be preserved during mapping. At the lowest level, it is simply not possible to mimic both the locations of the end-effectors and the joint angles. A system must make choices as to which aspects of the motion should be preserved and which should be allowed to change. We call an approach to motion retargeting that makes this choice explicitly an *importance-based approach*.

Nonimportance-based approaches make implicit choices as to what should be preserved during retargeting. For example, the most naive implementation of retargeting simply transfers the parameter (joint angles and root position) values from the performer to the target character. Such a scheme implicitly selects the values of the parameters to be important and, therefore, the positions of the end-effectors to be unimportant. This is a poor choice when the character must interact with other objects in the world, such as the floor.

¹Scheme is a statically scoped and properly tail-recursive dialect of the Lisp programming language invented by Guy Lewis Steele, Jr. and Gerald Jay Sussman.

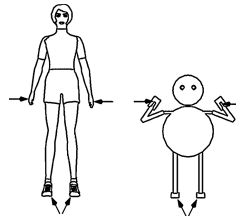


Fig. 2. Artifacts of position-based approach.

A common approach to motion retargetting matches the end-effector positions of the character to those of the performer. Such an approach has the advantage that it preserves the interactions between the character and its environment. Badler et al. [1993] used only the position data of hands and feet to adopt them to a virtual character with an inverse kinematics technique. Residual degrees of freedom are fixed by exploiting biomechanical knowledge. Choi et al. [2000] adopted the idea of inverse rate control [Whitney 1969] to compute the changes in joint angles corresponding to those in end-effector positions while imitating the captured joint angles by exploiting the kinematic redundancy.

Implicit in end-effector schemes is the notion that end-effector positions are more important than joint angles; that is, joint angles should be changed to achieve end-effector positioning goals. While this prioritization is often preferable to the reverse, it is not without its flaws. Consider the example of Figure 2. In this example, the importance of the foot positions is properly reflected, while that of the hand positions is overstated.

The central observation of an importance-based approach is that what is important can only be determined by the context of the motion. At each instant, a system must somehow select among the many possible things that are important, so it can change the aspects that are not important.

Constraint-based approaches to motion adaptation explicitly represent details of the motion that are important as geometric constraints. The space-time motion editing and retargetting system of Gleicher [1997, 1998] proposed the notion of preserving the important qualities of the motion by changing unimportant ones, where the important qualities were defined by constraints. Lee and Shin's [1999] hierarchical motion editing provided similar results using a different underlying implementation. Popovic and Witkin [1999] demonstrated results that made the kinetic aspects of the original motion important to preserve.

The methods mentioned in the previous paragraph are all offline in that they examine the entire motion simultaneously in processing. This offline nature is also implicit in the problem formulation, as well as in the solution method. All of the methods require the constraints to be identified before the motion can be processed. The decisions as to what is important in a motion must be known before processing can occur in these previous constraint-based approaches. This is infeasible in online applications. Bindiganavale and Badler [1998] introduced

a scheme to generate constraints automatically. However, their motion adaptation is done in an offline manner.

For computer puppetry, we must decide what is important in a given motion in an online manner. We analyze the importance of each end-effector position based on several factors discussed in Section 4. For example, the proximity of an end-effector position to its surrounding environment can be used as a predictor of its importance. The importance of an end-effector is inversely proportional to its distance to the nearest object in the environment. A key notion of this work is that the power of an importance-based approach, already demonstrated in offline constraint-based systems, can be brought to the online domain of computer puppetry.

2.3 Inverse Kinematics

We employ an inverse kinematics (IK) solver to compute the pose of the target character. IK has become a standard technique in animation systems to control the pose of a character based on the positions of its end-effectors.

IK solvers can be divided into two categories: analytic and numerical solvers. Most industrial robot manipulators are designed to have analytic solutions for efficient and robust control. Paden [1986] divided an IK problem into a series of simpler subproblems each of which has closed-form solutions. Korein and Badler [1982] showed that the IK problem of a human limb allows an analytic solution, and Tolani et al. [2000] derived their actual solutions. A numerical method relies on an iterative process to obtain a solution. Girard and Maciejewski [1985] generated the locomotion of a legged figure using a pseudoinverse of a Jacobian matrix. Based on neurophysiology, Koga et al. [1994] produced an experimentally good initial guess for a numerical procedure. Gullapalli et al. [1996] reduced the dimensionality of the redundant control system using synergies as a basis control set. Zhao and Badler [1994] formulated the IK problem as a constrained nonlinear optimization problem. Rose et al. [1996] extended this formulation to cover constraints that hold over an interval. To prevent the figure from making unnatural motions and reduce the redundancy of the IK problem, Badler et al. [1993] incorporated biomechanical information.

For computer puppetry, we make a number of demands on IK that require the development of a novel solver. First, we must achieve real-time performance on the entire body of the character. Second, we need the solver to provide predictably consistent solutions: small changes to the problems should provide similar answers. Finally, the solver must be able to account for the importances of the end-effectors that are determined dynamically in our system.

Our IK solver is discussed in Section 5. To solve an IK problem in real-time, we divide it into three subproblems: root position estimation, body posture computation, and limb posture computation. First, the root position of a virtual character is computed to provide a good initial guess for the body posture computation. If needed, we then adopt numerical optimization to refine the body posture, which consists of the root position, the root orientation, and the posture of the upper body. Finally, we use an analytic IK solver to compute the

limb postures and blend them with the captured limb postures. Our solution for each of these subproblems is designed to incorporate the importance value of each end-effector so that it tries to preserve the end-effector position when its importance value is high, while trying to preserve the captured joint angles of the corresponding limb, otherwise.

3. MOTION FILTERING

In general, motion capture devices capable of providing real-time performance are particularly susceptible to noise. Magnetic motion capture systems, which are widely used for real-time motion capture, suffer from the interference of low-frequency current-generating devices such as a CRT-type display. Thus, there always exists some level of jitter, that is, rapid random changes in reported positions and orientations that do not correspond to actual movements [Ascension Technology 1996]. Since online motion retargeting requires a high quality input motion as the reference of an output motion, filtering is an essential part. In the context of computer puppetry, filtering must be real-time, online, and performed on orientations as well as positions.

For online filtering, Kalman filters [Azuma and Bishop 1994; Friedmann et al. 1991; Welch and Bishop 1997] are often employed because of their capability of prediction and correction, that is, predicting future input data from their history and correcting them by incorporating actual input data. In a standard (extended) Kalman filter, its state would completely describe the position of a sensor and its velocity. However, because of the nonlinearity of the orientation space, this scheme can hardly be applied directly to orientation data. Adopting the results in Azarbajani and Pentland [1995] and Broida and Chellappa [1986], Welch and Bishop [1997] parameterize an incremental orientation change with Euler angles which are regarded as a three-vector to filter. The filtered Euler angles are transformed back to an incremental orientation change in the nonlinear space to update the target orientation at each time step. However, the conversion between an incremental orientation change and its equivalent Euler angles is inefficient. Moreover, recent motion capture devices measure orientations directly in unit quaternions. Therefore, differently from Welch and Bishop, we parameterize incremental orientation changes with rotation vectors.

To facilitate our scheme, we maintain the target orientation \mathbf{q}_e externally to the Kalman filter together with the internal state vector \mathbf{x} . In particular, \mathbf{q}_e is represented by an unit quaternion:

$$\mathbf{q}_e = (w \ x \ y \ z),$$

where $w^2 + x^2 + y^2 + z^2 = 1$. The internal state \mathbf{x} consists of the position \mathbf{p} , the rotation vector \mathbf{r} , and their derivatives $\dot{\mathbf{p}}$ and $\dot{\mathbf{r}}$:

$$\mathbf{x} = (\mathbf{p}^T \ \dot{\mathbf{p}}^T \ \mathbf{r}^T \ \dot{\mathbf{r}}^T)^T. \quad (1)$$

Here the rotation vector \mathbf{r} parameterizes the incremental orientation change of the actual sensor input $\mathbf{q}(t)$ at the current frame with respect to the target orientation $\mathbf{q}_e(t - \Delta t)$ at its previous frame. Therefore, $\mathbf{r}(t)$ can be measured

through the logarithmic map [Kim et al. 1995]:

$$\mathbf{r}(t) = \ln(\mathbf{q}_e^{-1}(t - \Delta t)\mathbf{q}(t)). \quad (2)$$

At each filter update step, $\mathbf{r}(t)$ in the state is converted into its incremental orientation change equivalent $e^{\mathbf{r}(t)}$ through the exponential map to update the external target orientation \mathbf{q}_e and then reset to zero. Therefore, incremental orientation changes are linearized for our (extended) Kalman filter, centered about zero.

Our dynamic model predicts the current position and the rotation by first-order approximations. Therefore, the prediction $\hat{\mathbf{x}}^-(t)$ of the state through the *state transition matrix* \mathbf{A} can be described:

$$\hat{\mathbf{x}}^-(t) = \mathbf{A} \hat{\mathbf{x}}(t - \Delta t) = \begin{bmatrix} \mathbf{I}_3 & \Delta t \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \Delta t \mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \hat{\mathbf{x}}(t - \Delta t), \quad (3)$$

where \mathbf{I}_3 and $\mathbf{0}_3$ are, respectively, 3×3 identity and zero matrices. Similarly, the *error covariance matrix* $\mathbf{P}(t)$ is predicted:

$$\mathbf{P}^-(t) = \mathbf{A} \mathbf{P}(t - \Delta t) \mathbf{A}^T + \mathbf{Q}. \quad (4)$$

Here, $\mathbf{P}(t) = E[(\hat{\mathbf{x}}^-(t) - \mathbf{x}(t))(\hat{\mathbf{x}}^-(t) - \mathbf{x}(t))^T]$, which models estimation uncertainty. The *process noise covariance matrix* \mathbf{Q} characterizes the accuracy of the dynamic model. In our implementation, we simplify \mathbf{Q} as

$$\mathbf{Q} = \begin{bmatrix} q_1 \mathbf{I}_3 & q_2 \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ q_3 \mathbf{I}_3 & q_4 \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & q_5 \mathbf{I}_3 & q_6 \mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & q_7 \mathbf{I}_3 & q_8 \mathbf{I}_3 \end{bmatrix}. \quad (5)$$

When the values of q_i s are small, the filter tends to suppress the details of the captured motion. On the other hand, if they are large, it tends to preserve the captured motion. Therefore, q_i s should be tuned interactively for a good filter response.

In practice, we sample motion signals at a higher frame rate (120 Hz) than that actually required for animation to avoid the overshooting which occasionally occurs in constant velocity models, especially when the velocity changes suddenly. Our measurement consists of the position of a sensor and its incremental orientation change represented by a rotation vector; that is, $\mathbf{z} = (\mathbf{p}^T \mathbf{r}^T)^T$ which can be obtained from the state vector directly. Therefore, our measurement can be estimated from the predicted state:

$$\hat{\mathbf{z}}(t) = \mathbf{H} \hat{\mathbf{x}}^-(t) = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \end{bmatrix} \hat{\mathbf{x}}^-(t). \quad (6)$$

Now, we are ready to compute the Kalman gain $\mathbf{K}(t)$:

$$\mathbf{K}(t) = \mathbf{P}^-(t) \mathbf{H}^T (\mathbf{H} \mathbf{P}^-(t) \mathbf{H}^T + \mathbf{R})^{-1}, \quad (7)$$

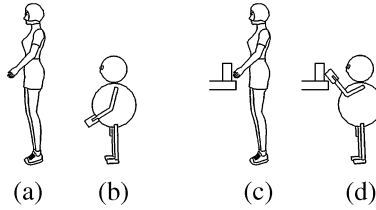


Fig. 3. Two different situations.

where \mathbf{R} is the *measurement noise covariance matrix*. That matrix is either given from the device manufacturer or acquired by offline measurement. In practice, we measure the noise while holding the sensor stationary to compute its noise covariance matrix \mathbf{R} .

The residual between the actual sensor measurement $\mathbf{z}(t)$ and its estimation $\hat{\mathbf{z}}(t)$ from Eq. (6) is:

$$\Delta \mathbf{z}(t) = \mathbf{z}(t) - \hat{\mathbf{z}}(t). \quad (8)$$

Then, the predicted state and the error covariance matrix are corrected as follows,

$$\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}^-(t) + \mathbf{K}(t)\Delta \mathbf{z}(t),$$

and

$$\mathbf{P}(t) = (\mathbf{I} - \mathbf{K}(t)\mathbf{H})\mathbf{P}^-(t). \quad (9)$$

We finish filtering at each frame by updating the external target orientation using the rotation vector $\hat{\mathbf{r}}(t)$. Taking the exponential map of the rotation vector and postmultiplying it with the external target orientation $\hat{\mathbf{q}}_e(t - \Delta t)$ at the previous frame, we can find the final target orientation $\hat{\mathbf{q}}_e(t)$ at the current frame:

$$\hat{\mathbf{q}}_e(t) = \hat{\mathbf{q}}_e(t - \Delta t)e^{\hat{\mathbf{r}}(t)}. \quad (9)$$

The rotation vector $\hat{\mathbf{r}}(t)$ is reset to zero for filtering at the next frame.

4. IMPORTANCE ANALYSIS

When the performer and the target character do not have the same size and proportion, not all aspects of the original motion can be preserved. A system must determine which aspects of the motion are important to preserve, so that other less important aspects may be changed to preserve them.

For an articulated figure, differing segment lengths means that the joint angles and end-effector positions cannot be recreated simultaneously. There are three obvious choices of motion aspects to preserve:

- (1) the position of the root of the character,
- (2) the joint angles, and
- (3) the positions of the end-effectors.

There exist situations under which any of these three might be most important. For example, observe the arm postures in Figure 3. Figure 3(a) shows a captured arm posture from the performer that does not touch any object. Retargeting

this motion to a virtual character, we prefer the posture in Figure 3(b) that preserves the joint angles. However, the position of a hand needs to be preserved when it touches an object as shown in Figures 3(c) and 3(d).

Our system must choose which of the three choices above is most important in a dynamic online way. To make this decision, we employ a number of heuristics.

- (1) **The position of the root is most likely *not* important.** This heuristic comes from the observation that the choice of making the root is arbitrary: we could have just as easily chosen any point as the root. In fact, preserving the root position may change some important parameters that characterize a posture itself. Because of this, the importance of the root position is downplayed in many approaches that consider importance. As with our solver, described in Section 5, the retargetting system of Gleicher [1998] uses a heuristic that attempts to satisfy the constraints (generally on the end-effectors) as much as possible by moving the root position.
- (2) **If an end-effector is interacting with another object (such as the floor), then its position is likely to be important. Therefore, proximity to objects in the environment should increase the importance of an end-effector.**
- (3) **If an end-effector will be interacting with another object in the near future, then its position is important (as it is likely to be getting ready for the interaction). Therefore, we incorporate the predicted proximity of an end-effector to an object in the measure of its importance.**
- (4) **If an end-effector has just finished interacting with another object and is moving away from it, its position may not be as important as its proximity suggests.**
- (5) **If the end-effector is not in proximity to another object, it is likely that its position is unimportant.**

In order to measure the interactivity of an end-effector with its environment, we introduce the notion of *importance* of an end-effector, which can be determined by analyzing the posture of the character in relation to the environment. In particular, the distance from the end-effector to objects in the environment is a good measure of interaction possibility. That is, the end-effector is more likely to interact with the environment when it is closer to objects in the environment. Therefore, as the end-effector approaches an object, its importance value should be increased continuously to enforce the geometric constraints created by the object. As the end-effector moves away from the object, the importance value should be decreased continuously to preserve the captured posture of the corresponding limb. Moreover, it is desirable to develop the distance measure to reflect the trajectory of an end-effector and its dynamic nature.

Given end-effector e_i of the performer and object o_j in the real space, let $d_{ij}(t)$ be Euclidean distance between them at time t . o_j has its corresponding object in the virtual space. The new distance function $d_{ij}^+(t)$ is defined as

$$d_{ij}^+(t) = \frac{d_{ij}(t) + d_{ij}(t + \kappa \Delta t)}{2} \quad (11)$$

for small positive κ and Δt . $d_{ij}^+(t)$ represents the average of the current

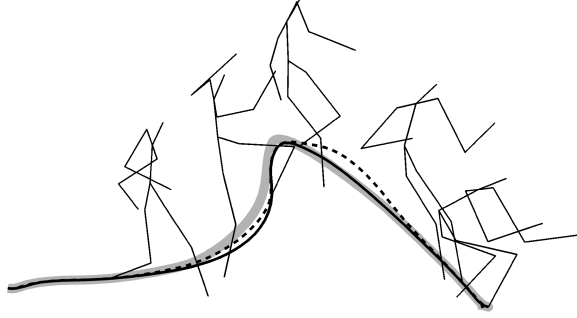


Fig. 4. Trajectories of the left foot generated by varying importance measure.

distance and the predicted distance after $\kappa \Delta t$ time. For small Δt , $d_{ij}^+(t)$ can be approximated as

$$\begin{aligned} d_{ij}^+(t) &\approx \frac{d_{ij}(t) + (d_{ij}(t) + \kappa \Delta t \dot{d}_{ij}(t))}{2} \\ &= d_{ij}(t) + \frac{\kappa \Delta t}{2} \dot{d}_{ij}(t) = d_{ij}(t) + \lambda \dot{d}_{ij}(t), \end{aligned} \quad (12)$$

where $\dot{d}_{ij}(t)$ is the first derivative of $d_{ij}(t)$. $d_{ij}^+(t)$ reflects both the distance at t from e_i to o_j and its changing rate $\dot{d}_{ij}(t)$. By varying λ we can control the degree of prediction for $d_{ij}^+(t)$.

For an example, Figure 4 exhibits a jumping motion adapted with $\lambda = 0$ and $\lambda = 0.15$. The legs of the character are shorter than the performer's. For $\lambda = 0$, the left foot trajectory of the character (dashed line) agrees with that of the performer (thicker line) only near the floor. For $\lambda = 0.15$, the former follows the latter while approaching down to the floor (solid line). The foot is moving off the captured trajectory to preserve the captured joint angles, either near the peak ($\lambda = 0$) or approaching to the peak ($\lambda = 0.15$).

Let D_{ij} denote the maximum distance within which e_i is influenced by o_j . Then, the normalized distance \bar{d}_{ij} is defined as

$$\bar{d}_{ij} = \frac{d_{ij}^+}{D_{ij}}. \quad (13)$$

An animator assigns D_{ij} for the pair of end-effector e_i and object o_j in the environment in accordance with a given animation context. A wider range of D_{ij} shows a sensitive interaction of end-effector e_i with object o_j . On the other hand, a narrower range exhibits that e_i moves independently of o_j unless e_i is close to o_j .

The importance is zero when the normalized distance \bar{d}_{ij} is greater than or equal to one; that is, e_i is out of the influence of o_j . As the distance decreases to zero, the importance increases to one. Thus, the importance function p of the normalized distance \bar{d}_{ij} can be designed with the condition of $p(1) = 0$ and $p(0) = 1$. In addition, we set its derivatives there to be zero, that is, $p'(0) = 0$ and $p'(1) = 0$, to reduce the rate of change of the function p at both extreme points. Thus, the importance of e_i with respect to o_j is represented by the cubic

polynomial function p satisfying those conditions. That is,

$$p(\bar{d}_{ij}) = \begin{cases} 2\bar{d}_{ij}^3 - 3\bar{d}_{ij}^2 + 1, & \text{if } \bar{d}_{ij} < 1, \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

The importance value w_i of end-effector e_i over all external objects can be defined as the maximum of them:

$$w_i = \max_j(p(\bar{d}_{ij})). \quad (15)$$

It requires much time to compute the distance \bar{d}_{ij} from each end-effector e_i of a virtual character to every object o_j in the environment, in particular, for a complex surrounding environment. To achieve a real-time performance, we need to minimize the number of possible objects that interact with each end-effector in accordance with an animation context. An object that is hardly touched during the animation may be eliminated in importance value computation. Moreover, objects may also be described approximately with simpler geometry for easy distance computation.

5. REAL-TIME INVERSE KINEMATICS SOLVER

For computer puppetry, we must position the character such that the important aspects of a captured motion are preserved while providing real-time performance. For our application, this demands computing the character's posture 30 times per second. Therefore, we need an IK solver that not only can incorporate the importance measures of the previous section, but also has real-time performance even in the worst case.

As discussed in Section 2.3, previous IK solution methods do not meet the demands of computer puppetry. Analytic methods provide guaranteed performance but cannot incorporate importance measures required for retargeting. Numerical solvers can include the importance metrics, but they hardly guarantee real-time performance. To meet these two conflicting demands, we have developed a hybrid solver.

In this section, we present a fast IK algorithm that is specialized for humanlike articulated characters. We divide the IK process into three subproblems: root position estimation, body posture computation, and limb posture computation. For each step, we give a method that is specialized to achieve high performance. This leads us to employ inexpensive closed-form solutions if applicable, and reserve numerical optimization for the case in which it is absolutely required.

5.1 Root Position Estimation

In order to position the end-effectors of a character, an IK solver may change the root position of the character or adjust its joint angles. As mentioned in Section 4, the root of the character has been arbitrarily chosen, and thus is rarely the most important aspect to preserve. Therefore, our solver first attempts to make the character satisfy the constraints as much as possible by moving the root position. This strategy was demonstrated for retargeting by Gleicher [1998].

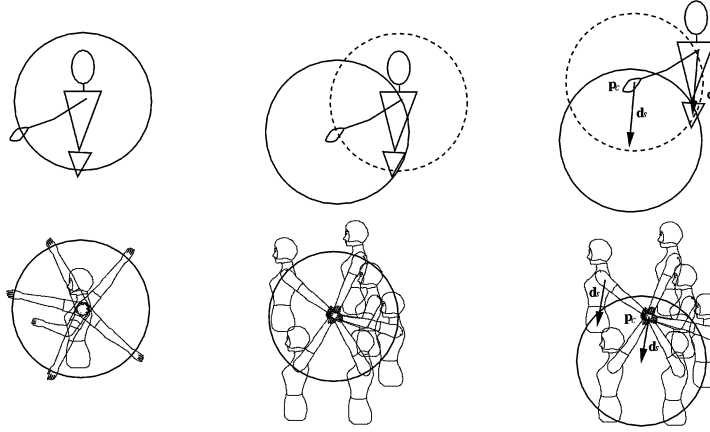


Fig. 5. Reachable ranges: range of hand, shoulder, and root position.

Beginning with the positional offset has an important advantage: unlike angular changes that cause nonlinear equations to compute, positional offset computation is trivial and therefore efficient. Let \mathbf{p}_i^e represent the position of the i th end-effector when the character is posed with the captured joint angles, and \mathbf{p}_i^g denote the goal position for that end-effector. The displacement vector $\mathbf{d}_i = \mathbf{p}_i^g - \mathbf{p}_i^e$ measures how much the solver must move an end-effector to reach its goal. If there were only one end-effector with a specified goal position, this constraint could be met by simply moving the character's root position by the displacement vector, where the joint angles would not need to be changed.

In the event that multiple end-effectors are to be positioned, we compute the weighted average of the displacements to find an initial offset \mathbf{d} as follows:

$$\mathbf{d} = \frac{\sum_i^n w_i \mathbf{d}_i}{\sum_i^n w_i}, \quad (16)$$

where w_i is the importance of the i th end-effector. In the (unlikely) event that all end-effectors require the same displacement, this displacement will solve all of the constraints. More likely, the joint angles will need to be adjusted so that all of the end-effector goals can be met.

While the weighted averaging attempts to position the root to meet all of the goals simultaneously, it does not necessarily guarantee that all goals can be met. Once the root position is fixed, the character will try to meet the goals by straightening its joints. Therefore, the root position must be chosen such that all end-effector goals are "reachable," that is, close enough that straightening limbs will be sufficient, while being as close to the initial estimate as possible.

As shown in the left column of Figure 5, the reachable range of the hand can be represented as a 3-D ball centered at the shoulder, and its radius is the length of the arm. The middle column of Figure 5 shows that the same ball centered at the goal position represents the range of the shoulder joint position. Finally, with the orientations of the pelvis and the waist fixed as in the captured posture, we compute the range of the root position as illustrated in the right column of Figure 5. Let \mathbf{d}_s denote the vector from the shoulder to

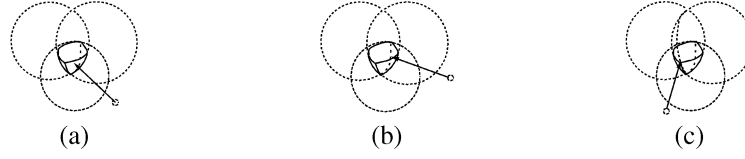


Fig. 6. Closest points: (a) on a sphere; (b) on a circle; (c) at a vertex.

the root position. The translation of the ball at the goal position \mathbf{p}_c by the vector \mathbf{d}_s yields the ball that gives the range of the root position. If the root is in this ball, the character can reach the goal position by stretching the arm only.

When the importance value of an end-effector is low, the root position does not need to be modified to make this end-effector reachable at its goal. Therefore, the range corresponding to this end-effector may be larger than the actual reachable range. To avoid an unnecessary offset of the root position, we enlarge the size of the ball, so that its size is inversely proportional to the importance value. The increased radius r_i corresponding to the i th limb is given as

$$r_i(l_i, w_i) = \frac{l_i}{w_i}, \quad (17)$$

where l_i is the length of the i th limb and w_i is its importance value.

Since the virtual character has four end-effectors, we have four corresponding 3-D balls. The common intersection of these balls is the range of the root position that makes all of the end-effectors reachable to their goal positions. As an initial guess for the root position, we choose the closest point from the offset root position to this intersection to preserve the posture of the performer as much as possible. Thus, the root position estimation is formulated as the problem of finding the closest point from a given position to the common intersection of four 3-D balls.

The intersection of 3-D balls consists of three surface elements as shown in Figure 6: spherical regions, circular edges, and vertices. A spherical region is a part of a sphere bounded by a sequence of spherical arcs. A circular edge is a part of a circle that is the intersection of two spheres. A vertex is determined by the common intersection of three spheres.

There are two cases depending on the offset root position with respect to the intersection. If this position is contained in the interior of the intersection, then the position itself is the closest point to the intersection. Suppose that it is not contained in the interior. Then the closest point must lie on the boundary of the intersection. Therefore, we may enumerate all possible surface elements due to the intersection of the four spheres corresponding to the bounding surfaces of the balls, respectively.

Three spheres determine at most two vertices. Since there are four ways of choosing a triple out of four spheres, we have a maximum of eight vertices. Every pair of vertices can possibly admit a spherical edge, and thus we have at most 24 edges. However, these are completely included in a maximum of six circles. Moreover, each spherical face is completely contained in one of four spheres. Instead of enumerating all surface elements, we equivalently check those spheres, circles, and vertices.

We first compute the closest point to each sphere from the offset root position. Among these points, if any, we choose the point that is contained in the intersection and the closest to the root position. If such a point does not exist, then we compute the set of points; each is the closest from the root position to each circle. Out of them, we choose the one that is closest to the root position and in the intersection. Suppose that such a point does not exist. Then one of vertices may be the solution. We choose the one closest to the root position among those contained in the intersection. For more details in estimating the initial root position, refer to the Appendix. If no common intersection of the balls exists, we discard the spheres that do not intersect the one whose corresponding end-effector has the largest importance value and repeat this process for the remaining balls.

5.2 Body Posture Computation

If the initial root position estimate does not allow all limbs to be reachable to the goal positions, we need to adjust the body posture consisting of the root position, the root orientation, and the posture of the upper body. Since those segments are tightly coupled, a numerical method is adopted to find their configurations. Numerical methods hardly guarantee a real-time response for computing an inverse kinematics solution of an entire human figure, while it is practical to solve only a small part of the IK problem numerically, and to employ analytic methods for the rest of the task. Such a hybrid solver was demonstrated in Lee and Shin [1999].

We formulate a restricted version of the IK problem for determining the posture of the body separately from the problem of computing the postures of the limbs. The body posture of a character can be written as $\mathbf{v} = (\mathbf{p}_0, \mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_n)$, where \mathbf{p}_0 and \mathbf{q}_0 are the position and the orientation of the root, respectively. \mathbf{q}_j , $1 \leq j \leq n$, are the orientations of body segments such as the waist and the upper body. When the character has a rigid torso, \mathbf{v} is simply reduced to $(\mathbf{p}_0, \mathbf{q}_0, \mathbf{q}_1)$, since $n = 1$.

The objective function consists of two terms:

$$E = E_g + \alpha E_p, \quad (18)$$

where the first term E_g is for making the end-effectors reachable to their goals and the last term E_p is to preserve the captured posture. We explain those two terms in detail.

E_g is the sum of E_i s, each of which is a function of the distance from the i th end-effector e_i to its goal position. Provided with the shoulder (or the coxa) position \mathbf{p}_i^s of the i th limb and its goal position \mathbf{p}_i^g , E_i is given as

$$E_i = \begin{cases} 0, & \text{if } \|\mathbf{p}_i^s - \mathbf{p}_i^g\| < l_i, \\ (\|\mathbf{p}_i^s - \mathbf{p}_i^g\| - l_i)^2, & \text{otherwise,} \end{cases} \quad (19)$$

where l_i is the length of the i th limb when it is maximally stretched. E_i is zero when the end-effector e_i is able to reach its goal position. Recall that an end-effector of a low importance value has no need to preserve its captured position. Thus, to relax the constraint on such an end-effector we enlarge the range of

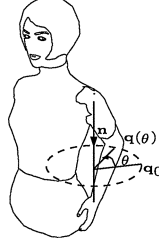


Fig. 7. Residual degree of freedom of shoulder.

the shoulder. By substituting the length l_i of each limb with the new radius $r_i = l_i/w_i$ as mentioned in Section 6.1, we have

$$E_i = \begin{cases} 0, & \text{if } \|\mathbf{p}_i^s - \mathbf{p}_i^g\| < r_i, \\ (\|\mathbf{p}_i^s - \mathbf{p}_i^g\| - r_i)^2, & \text{otherwise.} \end{cases}$$

Note that with the importance value w_i of one, E_i plays a role of pulling the end-effector to reach the goal position exactly. On the other hand, as importance value w_i approaches zero, the i th end-effector keeps the original posture by preserving the joint angles.

Letting \mathbf{q}_j^* and \mathbf{p}_0^* be the captured orientation of the j th segment and the estimated position of the root, respectively, E_p is a weighted sum of the squared geodesic distances between \mathbf{q}_j and \mathbf{q}_j^* for all $0 \leq j \leq n$, and the squared distance between \mathbf{p}_0 and \mathbf{p}_0^* :

$$E_p = \sum_{j=0}^n \beta_j \|\ln(\mathbf{q}_j^{-1} \mathbf{q}_j^*)\|^2 + \gamma \|\mathbf{p}_0 - \mathbf{p}_0^*\|^2. \quad (20)$$

Minimizing E_p preserves the captured motion as much as possible. We find a solution that minimizes the objective function by employing the conjugate gradient method. Here, we use the captured joint angles and the root position estimated in Section 5.1 as the initial guess for our optimization.

5.3 Limb Postures Computation

Given the position of a shoulder and that of the goal together with a hand orientation, we present the way our IK solver computes the configuration of an arm. The configuration of a leg can be computed similarly from the hip position, foot position, and orientation. As pointed out by Tolani et al. [1996] and Lee et al. [1999], the angle between the brachium of the arm and its forearm can be computed uniquely from the distance between the shoulder and the goal. We adjust the shoulder joint to locate the wrist at the goal position. Even with the wrist position fixed at the goal position, the shoulder joint still has one residual degree of freedom that rotates the elbow about the axis passing through the shoulder and the wrist. Korein and Badler [1982] have parameterized that degree of freedom by the swivel angle θ . As illustrated in Figure 7, the elbow traces a circle called the *elbow circle* as θ varies. Once θ is chosen, the shoulder joint is uniquely determined, and the joint angle of the wrist is also fixed to satisfy the given hand orientation.

This swivel angle θ can be described with a unit quaternion formulation. The unit quaternion $\mathbf{q}(\theta)$ representing the rotation by θ about the axis \mathbf{n} is $e^{\theta\mathbf{n}/2}$ for $-\pi < \theta \leq \pi$, where θ is measured from an arbitrarily chosen reference point on the circle. Denoting this point by a unit quaternion \mathbf{q}_0 , we have

$$\mathbf{q}(\theta) = e^{\theta\mathbf{n}/2} \mathbf{q}_0. \quad (21)$$

Unlike Lee and Shin [1999], who determine θ by a numerical optimization, we solve for θ analytically so that the arm posture deviates as little as possible from the captured posture. Thus, we choose θ that minimizes the geodesic distance ϕ in $[0, \pi]$ from the captured shoulder joint orientation \mathbf{q}^* to $\mathbf{q}(\theta)$. Note that both \mathbf{q}^* and $-\mathbf{q}^*$ represent the same orientation, due to the antipodal equivalence of the unit quaternion space. Therefore, we need to find θ that minimizes the function $\phi(\theta)$:

$$\phi(\theta) = \min(\arccos(\mathbf{q}^* \cdot \mathbf{q}(\theta)), \arccos(-\mathbf{q}^* \cdot \mathbf{q}(\theta))) = \arccos(|\mathbf{q}^* \cdot \mathbf{q}(\theta)|). \quad (22)$$

$\phi(\theta)$ is minimized when $|\mathbf{q}^* \cdot \mathbf{q}(\theta)|$ is maximized. By definition, $e^{\theta\mathbf{n}/2} = (\cos \theta/2, \mathbf{n} \sin \theta/2)$. Let $\mathbf{q}^* = (w^*, \mathbf{v}^*)$ and $\mathbf{q}_0 = (w_0, \mathbf{v}_0)$. Then, the absolute value of their inner product is given as

$$\begin{aligned} |\mathbf{q}^* \cdot \mathbf{q}(\theta)| &= \left| (w^*, \mathbf{v}^*) \cdot \left\{ \left(\cos \frac{\theta}{2}, \mathbf{n} \sin \frac{\theta}{2} \right) (w_0, \mathbf{v}_0) \right\} \right| \\ &= \left| a \cos \frac{\theta}{2} + b \sin \frac{\theta}{2} \right| \\ &= \sqrt{a^2 + b^2} \left| \sin \left(\frac{\theta}{2} + \alpha \right) \right|, \end{aligned} \quad (23)$$

where

$$\begin{aligned} \alpha &= \tan^{-1} \frac{a}{b} \\ a &= w^* w_0 + \mathbf{v}^* \cdot \mathbf{v}_0 \\ b &= w_0 \mathbf{n} \cdot \mathbf{v}^* - w^* \mathbf{n} \cdot \mathbf{v}_0 + \mathbf{v}^* \cdot (\mathbf{n} \times \mathbf{v}_0). \end{aligned}$$

The absolute value of the sine function has the maximum at $\pi/2$ and $-\pi/2$. Thus, $|\mathbf{q}^* \cdot \mathbf{q}(\theta)|$ is maximized either at $\mathbf{q}(-2\alpha + \pi)$ or at $\mathbf{q}(-2\alpha - \pi)$. Since the distance of two points in the unit quaternion space is inversely proportional to their dot product, $\mathbf{q}(-2\alpha + \pi)$ is the closest from \mathbf{q}^* if $\mathbf{q}^* \cdot \mathbf{q}(-2\alpha + \pi) > \mathbf{q}^* \cdot \mathbf{q}(-2\alpha - \pi)$; otherwise, $\mathbf{q}(-2\alpha - \pi)$ is the closest.

Now, with both captured and computed limb postures available, we blend them to obtain a realistic motion. For this purpose, we perform spherical linear interpolation between each captured joint orientation of a limb with its corresponding IK solution. Let q_{ik} and q_{ik}^* be the orientation of the k th joint in the i th limb obtained from the IK solver and that from the captured posture. Then, the blended joint angle q'_{ik} can be described by spherical linear interpolation as

$$\begin{aligned} \mathbf{q}'_{ik} &= \text{slerp}(q_{ik}^*, q_{ik}, w_i) \\ &= e^{w_i \ln(\mathbf{q}_{ik} \mathbf{q}_{ik}^{*-1})} \mathbf{q}_{ik}^*, \end{aligned} \quad (24)$$

where w_i is the importance value of the i th end-effector. That is, the limb with a high importance value can preserve the end-effector position, and that with a low importance value can preserve the captured joint angles.

The nonpenetration condition may be violated since the posture is blended regardless of the constraints. Thus the blended posture has to be adjusted explicitly to prevent unwanted penetration. Provided with the predefined external objects for each end-effector, this violation can be detected easily. Before penetrating an object, the end-effector touches the boundary of the object. Thus, the preferable position of the end-effector is the intersection point of the object boundary and the ray from the shoulder to the end-effector during penetration. This position moves continuously on the object in accordance with the end-effector movement. The penetration problem can be effectively eliminated by adjusting the limb posture using the IK solver for limbs.

6. ANALYSIS OF TEMPORAL CONSTRAINTS

In retargetting motions, we must preserve important temporal aspects of the motions along with spatial aspects. Gleicher [1998] emphasizes the importance of avoiding the introduction of high frequencies during adaptation. Both this work and the work of Lee and Shin [1999] provide approaches for avoiding the addition of discontinuities during adaptation. Unfortunately, both schemes rely on examining durations of motions and therefore can only be applied in offline applications. In this section, we show that the approach presented in this article does not introduce unwanted discontinuities into the resulting motion.

To begin, we must assume that the initial motion is free of unwanted discontinuities. This assumption is not restrictive because the movement of the performer is continuous. Discontinuities may be introduced by noise in the capture process, but these are generally removed by the filtering process described in Section 3. The continuity of the initial motion applies to both the captured joint angles and the end-effector positions.

Given continuous paths for the end-effectors, our IK solver will provide continuous trajectories for the parameters. Achieving this requires the solver to make consistent changes. That is, similar inputs to the solver must provide similar outputs. To guarantee this consistency, our IK solver tries to find the solution in an online manner so that it is close to the filtered input posture, while satisfying the kinematic constraints.

For our IK solver, the only kinematic constraints are the positions of end-effectors. These constraints are specified at every frame as temporal constraints. As an end-effector is approaching an object in the environment, its distance to the object is monotonically decreasing. Similarly, the distance is monotonically increasing as the end-effector is departing from the object. When the end-effector touches (or passes by) the object, the monotonicity changes but the distance function is still continuous at that instance.

For any continuous distance function, our importance function gives continuous importance values as described in Section 4. In other words, the importance values are consistently changed to reflect the temporal proximity of end-effectors to the environment. Therefore, the importance values have interframe

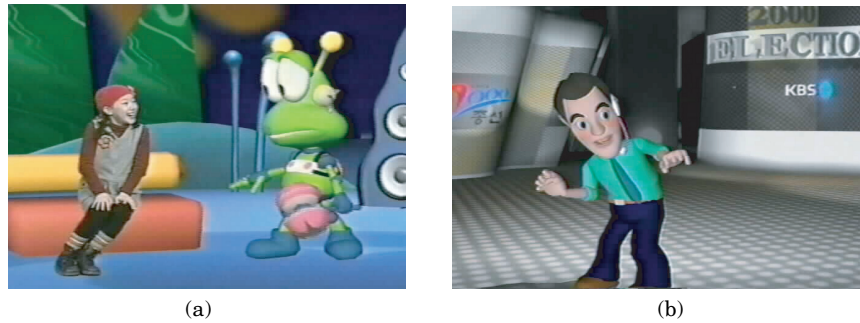


Fig. 8. Virtual characters on air controlled by our prototype system: (a) Pang-Pang; (b) Aliang.

coherence. Since our IK solver utilizes as input the reference motion data and the importance values, we can exclude unexpected motion artifacts such as unwanted jerkiness. That is, enforced to minimize the change from the reference motion, our IK solver tries to find an intended motion. Moreover, guided by the importance values for interaction with the environment, it also predicts the future temporal constraints and continuously pays attention to them for motion coherence.

7. EXPERIMENTAL RESULTS

For puppetry performance, we have used a MotionStar Wireless motion capture device from Ascension Tech, Inc. with 14 sensors and 2 extended range transmitters. Each of sensors detects the magnetic field emitted by a transmitter to report its position and orientation up to 144 times per second.

Our prototype system has been deployed for production and used successfully to create a virtual character for a children's television program as well as a virtual news reporter. Both have been shown on Korean national television, called KBS. The froglike creature ('Pang-Pang') given in Figure 8(a) has demonstrated his comic performance in a daily TV show for children. Thanks to the capability of our system for synthesizing realistic motion in real-time, Pang-Pang and a real actor can interact with each other. Figure 8(b) shows a virtual character ('Aliang') who has performed the role of a news reporter for the election of the Korea National Assembly. Even in a time-critical situation such as reporting interim election results, Aliang can accomplish his role successfully.

The skeleton used in our system has 43 degrees of freedom including 11 revolute joints of 3 degrees of freedom, 4 hinges on elbows and knees, and the position of the root and its orientation. The floor is modeled as a plane for all of the uses of our system to date.

To test our system's performance, we have created two puppets specifically designed to provide challenging retargetting problems. The character named *long tall Sally* has long arms and legs, while a ball-shaped man called *Blubby* has extremely short legs. To perform experiments, 61 prerecorded motion clips were used as the input for motion retargetting.

Table I. Number of Iterations in Numerical Solver With and Without Root Position Estimation

Motion	# Frames	Number of Iterations			
		Without		With	
		Blubby	Sally	Blubby	Sally
Walk	39	47	0	0	0
Throw	157	244	0	0	0
Jump	88	111	0	0	0
Handstand	211	266	38	0	0
Dance	591	1253	0	1	0
Total (61 Clips)	9692	15634	429	8	0

Table II. Performance Data

Motion	# Frames	Blubby		Sally	
		Elapsed Time (msec)	Per Frame (msec)	Elapsed Time (msec)	Per Frame (msec)
Walk	39	203	5.2	206	5.3
Throw	157	864	5.5	876	5.6
Jump	88	466	5.3	474	5.4
Handstand	211	1135	5.4	1139	5.4
Dance	591	3188	5.4	3201	5.4
Total (61 Clips)	9692	52543	5.4	52732	5.4

Table I shows the number of iterations in numerical optimization with and without initial root position estimation. Statistics for five selected motion clips are given in the first five rows of the table. The total figures for 61 clips are shown in the last row. Since Sally has long arms and legs, she can reach the captured end-effector positions without moving its root position. Thus, the number of iterations for Sally is small even without initial root position estimation. However, with estimated initial root positions, the number of iterations decreases to zero for our test motion clips. The effect of initial root position estimation is more apparent for Blubby with short legs. In most cases, our estimation algorithm finds the root position that makes the end-effectors reachable to their goal positions without any help from the numerical solver given in Section 5.2.

Table II gives an overall performance of our online motion retargetting algorithm excluding rendering time. Timing information has been obtained on a SGI Indigo2 workstation with an R10000 195 MHz processor and 128 Mb memory. The execution time for each example mainly depends on the number of iterations in numerical optimization. The table shows real-time performance for each example.

In Figure 13, a captured walking motion is applied to Blubby with various methods. The upper figure (Figure 13(a)) reveals artifacts due to the geometric inconsistency between the performer and the puppet. Since the positions of the feet are not incorporated into the motion retargetting, the supporting foot is sliding. In contrast, the middle figure (Figure 13(b)) preserves the positions well. However, the motions of the arms look unrealistic, since the joint angles of the arms are overadjusted to preserve the positions of the hands. The bottom

figure (Figure 13(c)) is generated by our motion retargetting. The supporting foot is fixed at the proper position without sliding, and the joint angles of the arms are preserved as the original ones.

With conventional approaches based on joint angle preservation, there would also exist foot-sliding artifacts when the character has longer limbs, as given in the top figure (Figure 14(a)). The middle image (Figure 14(b)) exhibits unintended bending of legs due to position preservation and an ill-selected initial root position. By assigning low importance values to the hands and offsetting the root position, we have a better result in which the legs are not bent as shown in the bottom figure (Figure 14(c)). More examples are given in Figures 15 and 16. In particular, Figure 16 shows the motions such as crawling and picking up which exhibit interaction of hands with objects in addition to that of feet.

8. CONCLUSIONS AND FUTURE WORK

We have presented a new approach for online motion retargetting that transforms motions of a performer to a virtual character of a different size and proportion. Introducing the notion of the importance of an end-effector, we have been able to generate realistic motion for a character in real-time while preserving the characteristics of captured motions as much as possible. KBS (Korean Broadcasting System), the largest public television broadcasting company in Korea, has been adopting our online motion retargetting algorithm to control the virtual character Pang Pang in a daily TV show for children. This show has become a children's favorite partly due to Pang Pang's successful performance. KBS also successfully showed the performance of the virtual reporter Aliang for a real election using this algorithm.

Our inverse kinematics solver is specialized for humanlike characters to ensure real-time performance, although it can be easily adapted to other types of creatures with limbs. The notion of importance gives reasonable lookahead capability useful for avoiding jerkiness in motion. However, unlike full-scale space-time optimization [Gleicher 1998], our approach has a limited lookback capability implicitly achieved by the Kalman filter, and allows only a restricted repertoire of constraints for real-time performance.

Our approach addresses only the interaction between the end-effectors of a character and objects in the environment. However, there may also be interaction among the segments of a character. For example, due to the geometric difference between the character and the performer, an end-effector of the character may penetrate its body, and also it is difficult for its hands to touch each other for clapping without interpenetration. Thus, to resolve these artifacts, the IK solver should be enhanced to efficiently handle self-interactions.

We focus on handling only the geometric discrepancy between a performer and a puppet. To generate more realistic motions, retargetting should also incorporate the characteristics of the puppet. Anthropomorphized animals such as cartoonlike birds and monkeys have their unique motion characteristics. Those motions can hardly be captured directly from a human performer, and thus give an additional barrier to overcome.

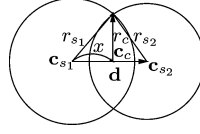


Fig. 9. Intersection of two spheres.

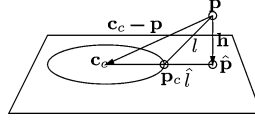


Fig. 10. Closest point from a point to a circle.

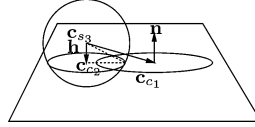


Fig. 11. Intersection of a sphere and a plane.

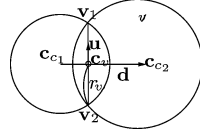


Fig. 12. Intersection of two circles.

APPENDIX

A. FINDING THE CLOSEST POINT ON THE INTERSECTION OF SPHERES

As given in Section 5.1, there are three types of surface elements: spheres, circles, and vertices. We describe how we find the closest point on each type of element to a given point \mathbf{p} . It is trivial to find the closest point on a sphere to the given point. Therefore, we proceed directly to the other cases.

Now, consider the closest point on a circle to \mathbf{p} . We start with constructing the circle C , which is the common intersection of the two spheres S_1 and S_2 (see Figure 9). The radius r_c of C can be computed with the Pythagorean theorem. Let \mathbf{c}_{s_i} and r_{s_i} for $i = 1, 2, 3$ be the center of the sphere S_i and its radius, respectively. The radius r_c of C satisfies the equations:

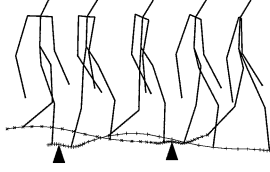
$$r_c^2 + x^2 = r_{s_1}^2, \quad (25)$$

and

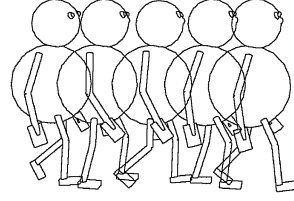
$$r_c^2 + (\|\mathbf{d}\| - x)^2 = r_{s_2}^2, \quad (26)$$

where x is the distance between the center \mathbf{c}_c of C and that of S_1 , and \mathbf{d} is the vector from \mathbf{c}_{s_1} to \mathbf{c}_{s_2} . Solving those equations for r_c^2 , we get

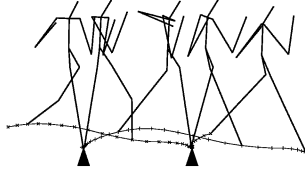
$$r_c^2 = r_{s_1}^2 - \frac{(r_{s_1}^2 - r_{s_2}^2 + \|\mathbf{d}\|^2)^2}{4\|\mathbf{d}\|^2}. \quad (27)$$



(a)



(b)



(c)

Fig. 13. Walking motion of *Blubby*: (a) captured joint angles only; (b) conventional IK solution with Kinematic constraints on end-effectors; (c) proposed algorithm combining captured joint angles and IK solution.

Here S_1 and S_2 intersect unless r_c^2 is negative. From Eqs. (25) and (26),

$$x = \frac{r_{s_1}^2 - r_{s_2}^2 + \|\mathbf{d}\|^2}{2\|\mathbf{d}\|}. \quad (28)$$

Thus,

$$\mathbf{c}_c = \frac{r_{s_1}^2 - r_{s_2}^2 + \|\mathbf{d}\|^2}{2\|\mathbf{d}\|} \cdot \frac{\mathbf{d}}{\|\mathbf{d}\|} + \mathbf{c}_{s_1}. \quad (29)$$

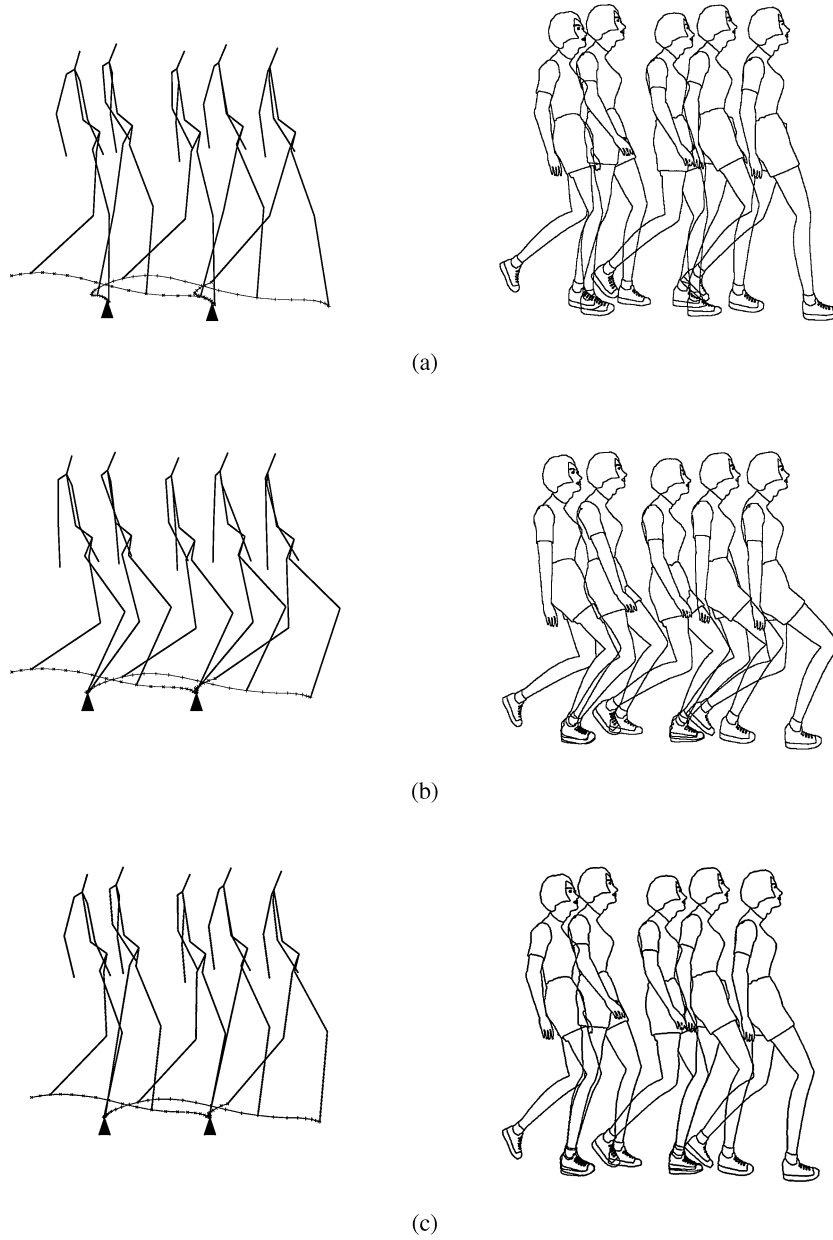


Fig. 14. Walking motion of *Sally*: (a) captured joint angles only; (b) conventional IK solution with Kinematic constraints on end-effectors; (c) proposed algorithm combining captured joint angles and IK solution.

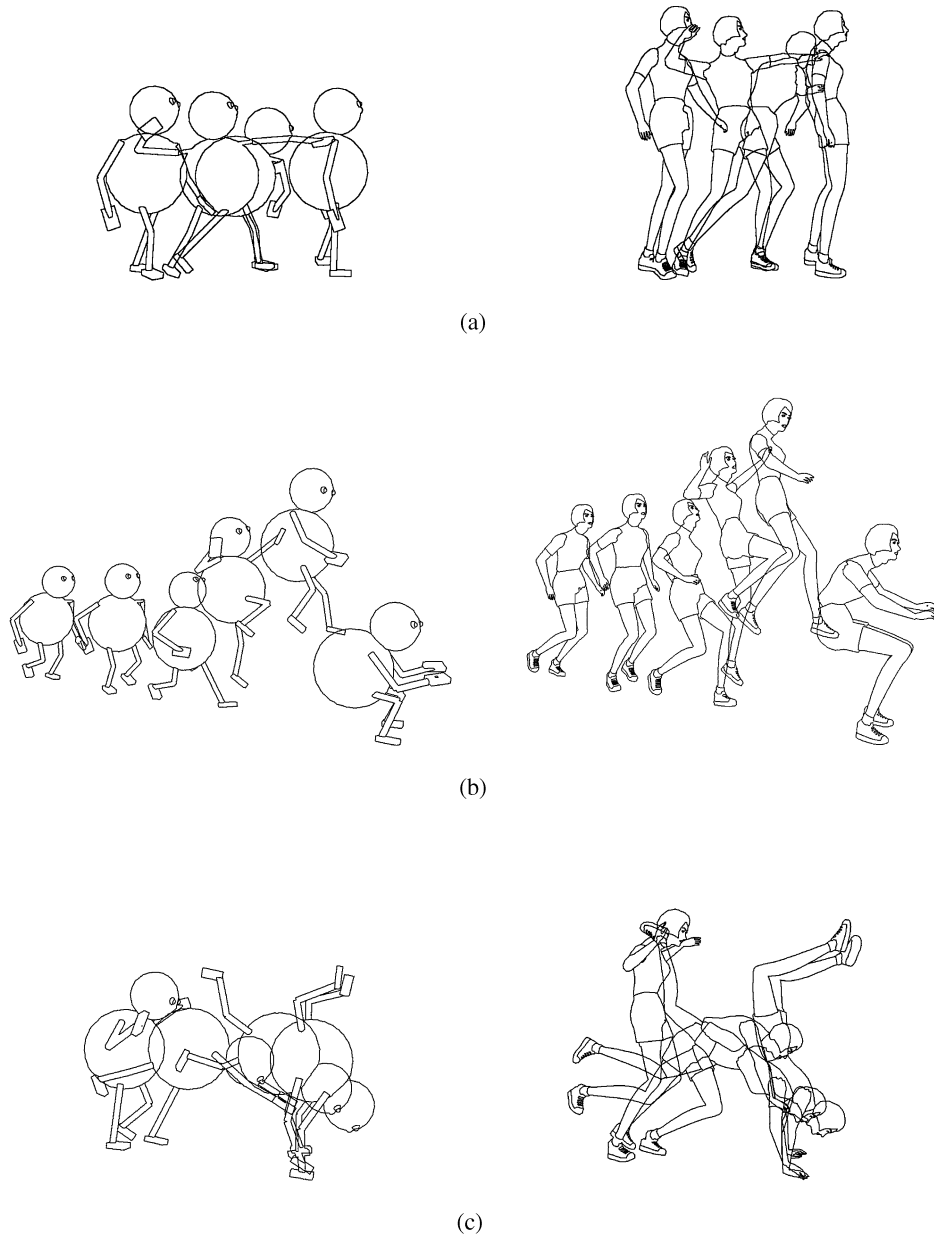


Fig. 15. Example motions of *Blubby* and *Sally*: (a) throwing; (b) jumping; (c) handstand.

Let \mathbf{n} be the normal vector of the plane where the circle lies. Then,

$$\mathbf{n} = \frac{\mathbf{d}}{\|\mathbf{d}\|}. \quad (30)$$

We are ready to find the closest point on the circle C to the given point \mathbf{p} . Let \mathbf{h} be the projection of the vector $\mathbf{c}_c - \mathbf{p}$ onto the normal vector \mathbf{n} of the plane;

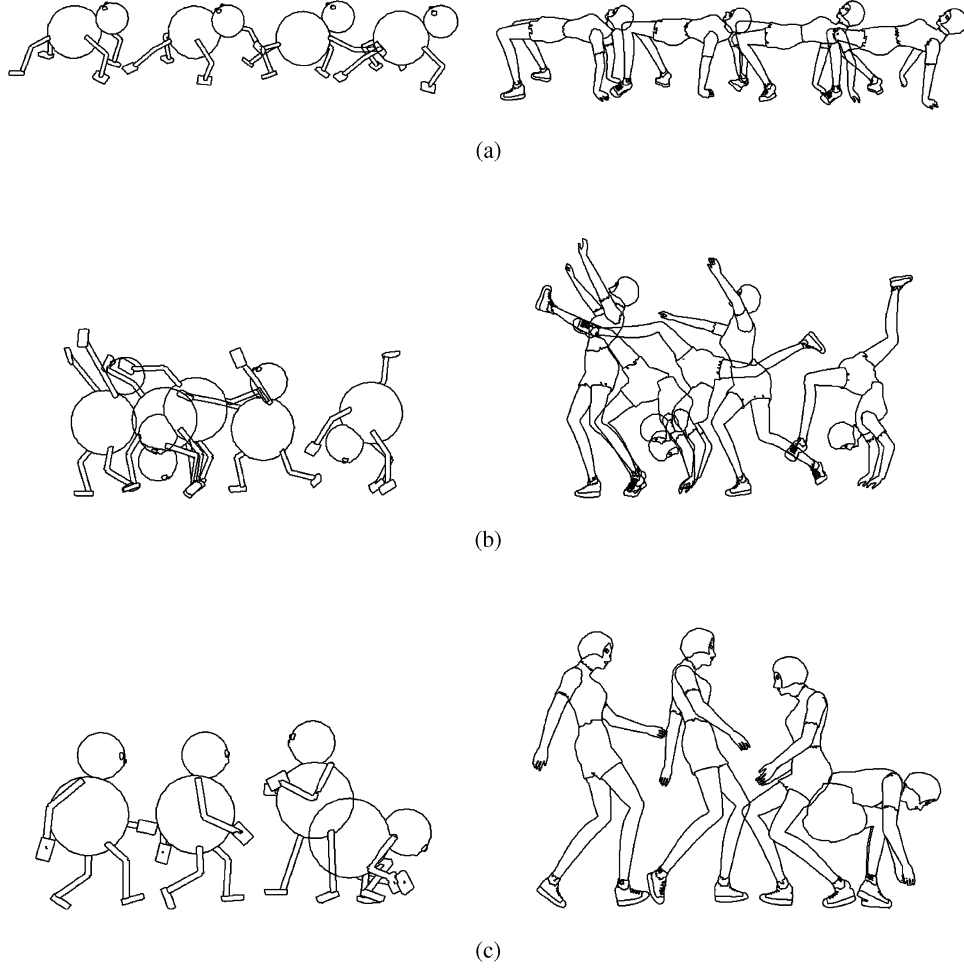


Fig. 16. Example motions with interaction of hands: (a) crawling; (b) backflipping; (c) picking up.

that is, $\mathbf{h} = [\mathbf{n} \cdot (\mathbf{c}_c - \mathbf{p})]\mathbf{n}$. Then, the closest point \mathbf{p}_c on C to \mathbf{p} is

$$\mathbf{p}_c = \mathbf{c}_c + \frac{\hat{\mathbf{p}} - \mathbf{c}_c}{\|\hat{\mathbf{p}} - \mathbf{c}_c\|} r_c, \quad (31)$$

where $\hat{\mathbf{p}} = \mathbf{p} + \mathbf{h}$; that is, $\hat{\mathbf{p}}$ is the projection of \mathbf{p} onto the plane containing C . As shown in Figure 10, the distance l from \mathbf{p} to \mathbf{p}_c is $\sqrt{\|\mathbf{h}\|^2 + \hat{l}^2}$, where \hat{l} is the distance from $\hat{\mathbf{p}}$ to \mathbf{p}_c ; that is, $\hat{l} = \|\hat{\mathbf{p}} - \mathbf{c}_c\| - r_c$.

Finally, we show how to find the closest vertex, if any, to the given point \mathbf{p} . Given those vertices, it is trivial to find the closest. Thus, we focus on explaining how to compute the vertices lying at the corners of the common intersection of three spheres, S_1 , S_2 , and S_3 . We first calculate the intersection circle C_1 of two spheres S_1 and S_2 . Cutting the sphere S_3 with the plane containing C_1 , we have the circle C_2 . Provided with the center point \mathbf{c}_{c_1} of C_1 and the normal vector \mathbf{n} of the plane containing the circle C_1 , the center point \mathbf{c}_{c_2} of C_2 is the projection

of the center point \mathbf{c}_{s_3} of the sphere S_3 onto the plane (see Figure 11). Thus,

$$\mathbf{c}_{c_2} = \mathbf{c}_{s_3} + \mathbf{h}, \quad (32)$$

where \mathbf{h} is the vector from \mathbf{c}_{s_3} to \mathbf{c}_{c_2} on the plane; that is, $\mathbf{h} = [\mathbf{n} \cdot (\mathbf{c}_{c_1} - \mathbf{c}_{s_3})] \mathbf{n}$. The radius r_{c_2} of C_2 is given as follows:

$$r_{c_2}^2 = r_{s_3}^2 - \|\mathbf{h}\|^2, \quad (33)$$

where r_{s_3} is the radius of the sphere S_3 . The sphere S_3 does not touch the plane if $r_{c_2}^2$ has a negative value.

Two vertices determined by three spheres are the intersection of the circles C_1 and C_2 . To compute the intersection of C_1 and C_2 , we evaluate the midpoint \mathbf{c}_v of the vertices \mathbf{v}_1 and \mathbf{v}_2 (see Figure 12). Similarly to the sphere–sphere intersection, the midpoint \mathbf{c}_v , and the distance r_v from each of vertices to \mathbf{c} are given as follows:

$$r_v^2 = r_{c_1}^2 - \frac{(r_{c_1}^2 - r_{c_2}^2 + \|\mathbf{d}\|^2)^2}{4\|\mathbf{d}\|^2}, \quad (34)$$

and

$$\mathbf{c}_v = \frac{r_{c_1}^2 - r_{c_2}^2 + \|\mathbf{d}\|^2}{2\|\mathbf{d}\|} \cdot \frac{\mathbf{d}}{\|\mathbf{d}\|} + \mathbf{c}_{c_1}, \quad (35)$$

where the \mathbf{d} is the vector from the \mathbf{c}_{c_1} to \mathbf{c}_{c_2} . The unit direction vector \mathbf{u} from \mathbf{c}_v to \mathbf{v}_1 is obtained from the cross-product of \mathbf{n} and \mathbf{d} ; that is, $\mathbf{u} = (\mathbf{n} \times \mathbf{d}) / \|\mathbf{n} \times \mathbf{d}\|$. Hence, we have the vertices $\mathbf{v}_1 = \mathbf{c}_v + r_v \mathbf{u}$ and $\mathbf{v}_2 = \mathbf{c}_v - r_v \mathbf{u}$.

ACKNOWLEDGMENTS

The authors would like to thank House of Moves Studios for providing sample motion data and Min G. Choi and Lucas Kovar for their help with paper production.

REFERENCES

- Ascension Technology Corp. 1996. *Motion Star Installation and Operation Guide*. Ascension Technology Corporation, Burlington, Vt.
- AZARBAYEJANI, A. AND PENTLAND, A. P. 1995. Recursive estimation of motion structure, and focal length. *IEEE Trans. Pattern Anal. Mach. Intell.* 17, 6, 562–575.
- AZUMA, R. AND BISHOP, G. 1994. Improving static and dynamic registration in an optical see-through hmd. In *Proceedings of SIGGRAPH'94*. ACM, New York, pp. 197–204.
- BADLER, N., HOLLICK, M. J., AND GRANIERI, J. P. 1993. Real-time control of a virtual human using minimal sensors. *Presence* 2, 1, 82–86.
- BINDIGANAVALA, R. AND BADLER, N. I. 1998. Motion abstraction and mapping with spatial constraints. In *Proceedings of International Workshop, CAPTECH'98*. Springer-Verlag, Wein, Austria, pp. 70–82.
- BODENHEIMER, B., ROSE, C., ROSENTHAL, S., AND PELLA, J. 1997. The process of motion capture: Dealing with the data. In *Proceedings of the Eurographics Workshop on Computer Animation and Simulation'97*. Springer-Verlag, Wein, Austria.
- BROIDA, T. J. AND CHELLAPPA, R. 1986. Estimation of object motion parameters from noisy images. *IEEE Trans. Pattern Anal. Machine Intell.* 8, 1, 90–99.
- CHOI, K.-J. AND KO, H.-S. 2000. On-line motion retargeting. *J. Visual. Comput. Animation* 11, 223–243.

- FRIEDMANN, M., STARNER, T., AND PENTLAND, A. 1991. Synchronization in virtual realities. *Presence* 1, 1, 139–144.
- GIRARD, M. AND MACIEJEWSKI, A. A. 1985. Computational modeling for the computer animation of legged figures. In *Proceedings of SIGGRAPH 85*. ACM, New York, pp. 263–270.
- GLEICHER, M. 1997. Motion editing with spacetime constraints. In *Proceedings of 1997 Symposium on Interactive 3D Graphics*. ACM, New York, pp. 139–148.
- GLEICHER, M. 1998. Retargeting motion to new characters. In *Proceedings of SIGGRAPH 98*. ACM, New York, pp. 33–42.
- GULLAPALLI, V., GELFAND, J. J., LANE, S. H., AND WILSON, W. W. 1996. Synergy-based learning of hybrid position/force control for redundant manipulators. In *Proceedings of the 1996 IEEE International Conference on Robotics and Automation*. IEEE Computer Society Press, Los Alamitos, Calif.
- HANSON, C. 2000. Scheme. <http://www-swiss.ai.mit.edu/projects/scheme/>.
- KIM, M.-J., SHIN, S. Y., AND KIM, M.-S. 1995. A general construction scheme for unit quaternion curves with simple high order derivatives. In *Proceedings of SIGGRAPH'95*. ACM, New York, pp. 369–376.
- KOGA, Y., KONDO, K., KUFFER, J., AND LATOMBE, J. 1994. Planning motions with intentions. In *Proceedings of SIGGRAPH'94*. ACM, New York, pp. 395–408.
- KOREIN, J. U. AND BADLER, N. I. 1982. Techniques for generating the goal-directed motion of articulated structures. *IEEE Comput. Graph. Appl.* 2, 71–81.
- LEE, J. AND SHIN, S. Y. 1999. A hierarchical approach to interactive motion editing for human-like figures. In *Proceedings of SIGGRAPH'99*. ACM, New York, pp. 39–48.
- MOLET, T., BOULIC, R., AND THALMANN, D. 1996. A real-time anatomical converter for human motion capture. In *Proceedings of the Seventh Eurographics Workshop on Animation and Simulation*. Springer-Verlag, Wein, Austria.
- MOLET, T., BOULIC, R., AND THALMANN, D. 1999. Human motion capture driven by orientation measurements. *Presence* 8, 2, 187–203.
- PADEN, B. 1986. Kinematics and control robot manipulators. Ph.D dissertation, University of California, Berkeley, Berkeley, Calif.
- POPOVIC, Z. AND WITKIN, A. 1999. Physically based motion transformation. In *Proceedings of SIGGRAPH'99*. ACM, New York, pp. 11–20.
- Protozoa. 1999. Technology information. http://www.protozoa.com/Page_2/info_index.html.
- ROSE, C. F., GUENTER, B., BODENHEIMER, B., AND COHEN, M. F. 1996. Efficient generation of motion transitions using spacetime constraints. In *Proceedings of SIGGRAPH'96*. ACM, New York, pp. 147–154.
- STURMAN, D. J. 1998. Computer puppetry. *IEEE Comput. Graph. Appl.* 18, 1, 38–45.
- TOLANI, D. AND BADLER, N. I. 1996. Real-time inverse kinematics of the human arm. *Presence* 5, 4, 393–401.
- TOLANI, D., GOSWAMI, A., AND BADLER, N. 2000. Real-time inverse kinematics techniques for anthropomorphic limbs. *Graph. Models* 62, 5.
- WELCH, G. AND BISHOP, G. 1995. An introduction to the Kalman filter. Tech. Rep. TR95-041. Department of Computer Science. University of North Carolina at Chapel Hill, Chapel Hill, NC.
- WELCH, G. AND BISHOP, G. 1997. Scaat: Incremental tracking with incomplete information. In *Proceedings of SIGGRAPH'97*. ACM, New York, pp. 333–344.
- WHITNEY, D. J. 1969. Resolved motion rate control of manipulators and human prostheses. *IEEE Trans. Man-Mach. Syst.* 47–53.
- ZHAO, J. AND BADLER, N. I. 1994. Inverse kinematics positioning using nonlinear programming for highly articulated figures. *ACM Trans. Graph.* 13, 4, 313–336.

Received November 2000; revised March 2001; accepted June 2001