

# Planning and Motion Control in Lifelike Gesture: A Refined Approach

Stefan Kopp & Ipke Wachsmuth  
Artificial Intelligence Group  
Faculty of Technology  
University of Bielefeld  
D-33594 Bielefeld, Germany  
{skopp,ipke}@techfak.uni-bielefeld.de

## Abstract

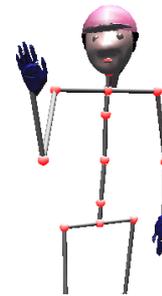
*In this paper an operational model for the automatic generation of lifelike gestures of an anthropomorphic virtual agent is described. The biologically motivated approach to controlling the movements of a highly articulated figure provides a transformation of spatiotemporal gesture specifications into an analog representation of the movement from which the animations are directly rendered. To this end, knowledge-based computer animation techniques are combined with appropriate methods for trajectory formation and articulated figure animation.*

## 1. Introduction

The inclusion of nonverbal modalities into the communicative behaviors of virtual agents has moved into focus of human-computer interface researchers. Humans are more likely to consider computer-generated figures lifelike when appropriate nonverbal behaviors are displayed in addition to speech. This enables the evocation of social communicative attributions to the artificial agent, which are supposed to be advantageous for a natural and efficient communication link between human and machine [1]. Hand-arm gestures and facial expressions, which are an integral part of human-human dialogues and therefore ingredients of practiced communicative skill, are first candidates for extending the communicative capabilities of such virtual agents. A core part in this is the animation and temporal coordination of lifelike gestures, as humans are very sensitive observers of each other's motion, in particular of motions that are supposed to convey meaning like gestures.

In [9], we discussed findings in various fields relevant to the production of gesture in humans and, based on conclusions from these findings, proposed a hierarchical model for planning and generating lifelike gestures of a multimodal

agent. Our approach is based on knowledge-based computer animation techniques and encapsulates low-level motion generation and control, enabling more abstract control structures on higher levels. In this paper, we first discuss related work on synthetic gesture of virtual agents. Then, after presenting our model and providing a succinct description of its planning stages, we focus on the lower level where gestural movements of the articulated figure are explicitly formed and executed. The presented techniques are conceived to drive the kinematic skeleton of a virtual agent (see fig.1) which comprises 43 degrees of freedom (DOF) in 29 joints for the main body and 20 DOF for each hand.



**Figure 1. The articulated structure of the virtual agent comprises 59 joints including the hands.**

## 2. Related work on synthetic gesture

As far as the *automatic* generation of lifelike gestures in virtual agents is concerned, there is very little work besides 2D presentational agents which usually rely on completely predefined gesture animations. The most comprehensive such system is the *Animated Conversation* system, developed by Cassell et al. [6], which generates speech,

intonation, and gesture of a virtual agent in conversational interaction. Gestures are produced by running predefined finite state machines (parallel transition networks, Pat-Nets) which issue gesture requests to the animation system (Jack<sup>TM</sup>). Arm and wrist motions are specified by target positions and orientations while hand motions are given in terms of an expandable library of handshapes, based on the American Sign Language (ASL) alphabet. However, gestural movements are apparently predefined and can only be parametrized in terms of alteration of single gesture phases, e.g., foreshortening the relaxation phase when the prerecorded “canonical” gesture time exceeds actual timing constraints. But there appears to be no means of coherently modifying the gestural movement while preserving natural movement features, e.g. typical velocity profiles. Perlin and Goldberg [14] created lifelike motions of virtual actors by means of script-based animations which are subject to rhythmic constraints and stochastic noise functions yielding a slight but permanently altered appearance. Gibet and Lebourque [11] present a sensori-motor model, restricted to a hand model attached to a three-segment arm, that generates natural hand-arm movements on the basis of successive target end-points annotated with additional synchronization properties. Natural appearance of the produced motion is achieved by incorporating heuristic knowledge into the movement formation process by means of minimization of a cost function during inverse kinematics. Furthermore, a presumed overall bell-shaped velocity profile is scaled in order to fit the movement’s duration which is estimated using closed-form approximations from human movement studies (e.g. Fitts’ law). The recently proposed EMOTE model by Zhao et al. [2] emphasizes the form and execution of qualitative aspects of movement in order to increase naturalness of independently pre-defined human-like motions. Based on the Effort and Shape components of Laban Movement Analysis, operational models for expressive arm and torso movements are presented which provide an intuitive way of controlling the manner of movements, but not the movement itself.

In summary, there is apparent need for convenient techniques of gesture production for virtual humanoids. Although promising approaches exist, most current systems produce movements which are only parametrizable to a certain extent or even rely on predefined motion sequences. As far as we see, a comprehensive model for the automatic generation of lifelike gestures from flexible symbolic descriptions based on spatiotemporal features is still lacking.

### 3. A model for gesture generation

The general goal of our research is an operational model that enables convincingly lifelike gesture animation from adequate representations of spatiotemporal gesture knowl-

edge. To this end, the model must provide sufficient means of motion representation, planning, and control to produce multiple kinds of gestural movements of a highly articulated figure as shown in Fig.1. Furthermore, it must be flexible in the sense that gestures can be parametrized with respect to kinematics, i.e. velocity profile and overall duration, as well as to shape properties. In this section, relevant findings from related disciplines are quoted and a hierarchical model is presented which provides all necessary computational steps from gesture planning to movement execution. In addition, a brief overview of the employed techniques for representing and planning the movements is given. The reader is referred to [9] for a more detailed description.

#### 3.1. Gesturing in humans

The process of gesture production and performance in humans is a complex multi-stage process. According to DeRuiter [4] abstract gesture specifications, generated from spatiotemporal representations of “shape” in the working memory on cognitively higher levels, are transformed into patterns of control signals which can be directly interpreted by low-level motor systems. The resulting gesture exhibits characteristic shape and dynamical properties enabling humans to distinguish them from subsidiary movements and to recognize them as meaningful. In particular, gestural movements can be considered as composed of distinct movement phases which form a hierarchical kinesic structure [7]. In case of coverbal gestures, the *stroke* (the most meaningful and effortful part of the gesture) is tightly coupled to accompanying speech yielding semantic, pragmatic, and even temporal synchrony between the two modalities [12, 4].

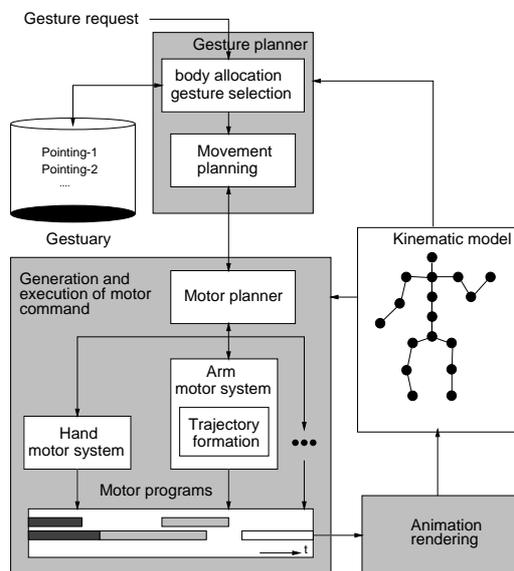
Models of motor control are commonly conceived in a hierarchical structure [16]. At the highest levels of those models, the global aspects of the movement are represented in the form of an abstract goal. Control is passed down through progressively lower levels until all particular choices are made about which motor units to use. However, control may even appear to reside at several levels simultaneously, with processes occurring in parallel at different levels [16].

Latash [10] proposes a general scheme of motor control that incorporates three levels which can be found in most approaches. This scheme accounts for the current view that planning voluntary movements is performed directly in terms of kinematics in the external space, rather than in complex “intrinsic” representations like joint rotation signals [13]. A basis for this is the knowledge about the initial arm position and the target location, suggesting that in gesture production primarily the most significant locations and postures are represented internally. Intermediate movements are generated by the motor control system automat-

ically. Planning of movement, hence, constitutes the *first step* of information processing, where the goal of the movement is expressed in terms of significant properties of its trajectory. By performing an internal simulation which takes into account actual conditions of movement execution, a function is generated that reflects the desired trajectory of the physical movement. The *second step* is to translate the simulated trajectory into motor variables which drive the lower structures. The resulting *virtual* trajectories can be considered reflections of a generalized motor program, suggesting that motor programs partially encode certain properties of movement, including patterns of transition from an initial to a final position. In the *third step*, the execution of these commands at the lowest level leads to a movement that, in the ideal case, exactly follows the simulated trajectory.

### 3.2. Gesture planning

The quoted findings suggest to perform movement planning in extrinsic representation, according to the *principle of smaller complexity* [13]. However, the configurations of an articulated figure eventually have to be described intrinsically in joint angles. Therefore, we propose two major stages of the control scheme as shown in fig.2, namely “gesture planning” and “generation and execution of motor command”, which map a given movement specification eventually onto a set of motor programs from which the movements are directly rendered.



**Figure 2. Schematical overview of the gesture generation process (from [9]).**

In the first gesture planning stage, an image of the movement is created which is internally formed by arranging constraints representing the mandatory spatial and temporal features of the gesture. These are given either as spatiotemporal descriptions from previous stages of gesture production, e.g., location of a referent for deictic gestures, or they are retrieved from representations of common sense knowledge about gestural movements, e.g., the conventionalized hand shape during pointing [4]. Therefore, our model comprises a lexicon (*gestuary*), containing abstract frame-based descriptions of gestural movements along with information about their usage for transferring communicative intent. The entries can hence be considered as defining a mapping from communicative function to explicit movement descriptions. The gestures are described in terms of either postural features (*static* constraints) or significant movement phases (*dynamic* constraints). Features that can be defined independently are described using the symbolic gesture notation system *HamNoSys* [15], while others must be determined for each individual gesture. To this end, the description further accommodates entries which uniquely refer to specific values of the content the gesture is to convey, e.g., quantitative parameters for deictic or iconic gestures. The gesture’s course in time is defined by arranging the constraint definitions in a tree using PARALLEL and SEQUENCE nodes which can optionally be nested. A more detailed description can be found in [9].

The gesture planner first decides which limb to use for performing the gesture. As this choice is partially resembled by the choice of a specific gesture template, the planner collects all gestures from the gestuary which fulfill the required communicative function. Then, taking into account information originating from various sources like proprioceptive feedback about current bodily movements, the planner suggests a particular gesture which suits best the actual movement conditions. The individual movement is, then, completed by deriving concrete values for variable movement constraints and scheduling each constraint of the gesture stroke by applying external temporal constraints (like onset time and duration due to cross-modal synchronization) and traversing the constraint tree [9]. In summary, the gesture planner forms a movement plan, i.e. a tree representation of a temporally ordered set of movements constraints, by (1) retrieving a feature-based gesture specification from the gestuary, (2) adapting it to the individual gesture context, and (3) qualifying the movement constraints to the extent possible by temporal integration with external timing constraints.

### 3.3. Motor planning

In the motor planning stage, motor control is decomposed into more simple modules by distinguishing move-

ment constraints according to the affected body parts. Relying on the assumption that movements of distinct motor systems need to be synchronized only at discrete time points, related constraints are ordered in lists, preserving their temporal relationships, and passed down to dedicated motor subsystems. In particular, hand motion, i.e. the adoption of subsequent hand postures, is controlled solely by a specialized hand model which is able to interpret the *HamNoSys* descriptions of hand shape by translating them directly into hand poses and applying stereotyped transitions between them. In the contrary, alterations of hand location and orientation are caused by arm movements and, thus, transferred to an arm-body control module, which, like the hand model, must be able to interpret *HamNoSys* symbols, convert them into position and orientation constraints with respect to the *egocentric* frame of reference, and generate a movement which lets the arm/hand follow an appropriate trajectory.

## 4. Movement formation and execution

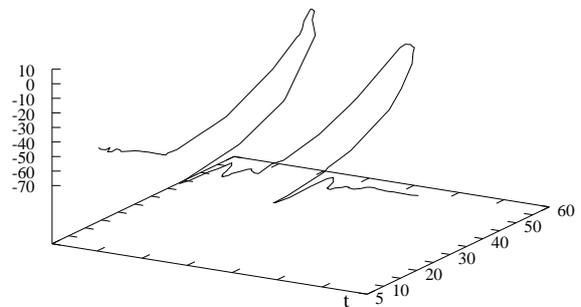
Forming a trajectory for a redundant manipulator (the arm model includes 7 DOF) which is to satisfy a given set of positional, orientational, and temporal constraints while preserving smoothness and naturalness of human motion is a difficult task. Our arm model employs a biologically motivated formation method, described in this section.

### 4.1. Simulation of human movement

In the field of motor control, two kinds of human movements are distinguished [16]. *Ballistic* movements are conceived as being completely prepared in advance and as running their course without allowing any amendment of motor commands, e.g., as in fast movements. Such movements in most cases exhibit symmetrical bell-shaped velocity profiles for the working point. *Controlled* movements on the other hand are performed at moderate speed and are subject to continuous amendment of the movement using feedback information. The resulting closed-loop control process is often reflected in discontinuous movements with several velocity peaks and unsymmetrical acceleration profiles. Coverbal gestures, although being performed spontaneously and mostly unconsciously, may be executed ballistically as well as by controlled movements, e.g., when visual feedback information is required in deictic gestures to a referent of small size.

In both cases considered, the composition of movements by humans is supposed to be a sequential process, which is corroborated by observations that humans generate complex movements by piecing together shorter or simpler movements, commonly referred to as “temporal segmentation”

[5, 10]. According to this view, the often observed non-smooth nature of human movement (“movement intermittency”) results from multiple submovements that humans perform in an attempt to achieve a high degree of accuracy or under unpredictable changes of movement conditions. In contrast, trajectories smooth out considerably for simple movements under consistent conditions [5], which we can assume for the gesture production of a virtual agent.



**Figure 3. Three-dimensional pointing gestures as observed in experimental studies [19]. Only movements in two directions (forward and upwards) are plotted against time.**

A huge number of experiments of human movement under various circumstances can be found in the literature (for an overview see [10]), which report some major observations: First, the trajectory of the effector seems to be more directly related to motor task than trajectories in individual joints. Secondly, there is likely a simple scaling procedure for permitting the same task at different speeds, due to large power reserves that allows implementing an *isochronous* strategy [13]. Finally, complex movements consist of elementary units glued together and speed drops at connection points which frequently correspond to points of maximal curvature of the trajectory. Experiments carried out in our lab [19] indicate that people choose among a great variety of different movements to perform the same gesture and tend to anticipate pending gestures by moving the arm to a preliminary rest position keeping only elbow flexion. However, preparation and retraction movements starting or ending at rest positions exhibit a significantly stereotyped character (see fig.3).

### 4.2. Formation of arm trajectories

From these findings we can draw several major conclusions. First, formation of significant and complex trajectory parts should be done in Cartesian coordinates. As the spatiotemporal features of the coordinated gesture movement are defined solely at distinct time instants, this naturally

leads to an interpolation problem for the hand trajectory in external coordinates. This is reasonable as interpolation in joint angular space for mutable movements, i.e. with considerable alterations of the trajectory's tangent vector, is by no means guaranteed to produce proper external limb trajectories due to the non-linear mapping between Cartesian coordinates and joint angle values. However, for rather simple preparation and retraction movements, a joint angle based formation leads to better results in a more efficient way, provided that a sufficient number of control points is given. In order to generate enough intermediate points to guarantee that the trajectory's deviation stays within prespecified error bounds, inverse kinematics may be applied in a scheme like the BDJP algorithm [18] frequently used in robotics for straight line paths.

Secondly, due to the invariance of movement shape from overall movement duration and speed, the movement's shape can be defined independent of its global kinematics. However, the detailed velocity profile does depend on the movement's shape like in slowing down at turning points and symmetrical acceleration/deceleration phases at straighter paths. Thus, a formation method is needed which on the one hand relates the movement shape to its velocity and, on the other hand, provides flexible means for adjusting gesture timing while preserving its shape. Furthermore, the compositionality of human movement requires a formation method that ensures continuity between consecutive curve segments and provides sufficient local control properties.

In order to fulfill these requirements, we have extended a trajectory formation method proposed by Morasso [13]; it is based on parametric curve composition using *S-splines* which provides the notion of subsequent linear or curved "guiding strokes"  $\vec{C}_i(t)$  with bell-shaped velocity profiles, yielding quite realistic velocity profiles of overall curved movements. An S-spline for  $n + 1$  control points  $\vec{p}_0, \dots, \vec{p}_n$  is defined as follows:

$$\begin{aligned}\vec{x}(t) &= \vec{p}_0 + \sum_{i=1}^{n-2} \vec{C}_i(t) \\ \vec{C}_i(t) &= \vec{\rho}_i S_{i,k}(t) + \vec{\rho}_{i+1} S_{i+1,k}(t) + \vec{\rho}_{i+2} S_{i+2,k}(t) \\ S_{i,k}(t) &= \sum_{j=i}^n N_{j,k}(t) \quad i = 1 \dots n\end{aligned}$$

The  $N_{j,k}$  are the ordinary B-spline blending functions of order  $k$  and  $\vec{\rho}_i = \vec{p}_i - \vec{p}_{i-1}$ . Note that  $\vec{x}$  is defined solely in terms of curved guiding strokes, two of which are simultaneously active; a combined representation consisting of both curved and linear strokes is easily possible. As S-splines are based on classical B-splines, they tend to approximate their control points. Interpolation can best be achieved by generating subsequent strokes towards the goal point, resulting in discontinuity due to the increased mul-

tiplicity of control points. An alternative formulation using the interpolating Kochanek-Bartels splines [8] which provide additional local control parameters of the spline's shape is currently under evaluation.

The gesture's kinematics is modelled explicitly by defining a velocity curve which is a two-dimensional spline of distance travelled, or arclength, against time [20]. It can be obtained either from the absolute value of the S-splines tangent vector  $|\frac{d}{dt}\vec{x}(t)|$  or simply by application of prespecified stereotyped velocity profiles as observed in experiments on unrestrained multijoint movements. The approach is an adapted *double interpolant method* [17] and provides high flexibility with regard to adjusting the movement to preplanned kinematic gesture features, e.g., an abrupt halt at a certain time, as well as the duration of a gesture in order to satisfy temporal constraints.

Taking the sequence of trajectory constraints as input, the position curve is generated by, first, determining the start conditions for the movement. In case the preplanned timing of a movement constraint cannot be satisfied (e.g. due to extensive transitional movements), achievable start and end times for the corresponding constraint are generated and the movement is re-scheduled to preserve the gesture's temporal structure. Once a feasible solution is found, the trajectory is formed using the methods previously described and a typical retraction movement is appended. If the limb is currently subject to another movement, the new motor program is to take over control and a smooth transition is generated [17] such that the position curve has at least second-order parametric continuity.

### 4.3. Rendering animations from motor programs

All motor systems create motor programs which, once fed into the central animation queue, control the execution of a submovement in a strictly timed manner. While motor programs created by the hand motor system directly affect the kinematic model by modifying joint angles, arm movements are defined in terms of trajectories in Cartesian coordinates which are reparametrized in arclength to control the movement's velocity using standard techniques [20]. Then, for each incrementally altered limb position the inverse kinematics problem is solved in real-time by an extension of the Jacobian transpose method [21]. The algorithm translates a composite force/torque  $\vec{F}$  applied to the end-effector into joint angles and takes into account limitations of the human body, as well as biomechanical heuristics in form of cost functions  $C(\vec{q})$ . Based on the observations in [3] that subjects prefer to adopt postures which minimize a superposition of U-shaped cost functions for each rotational degree of freedom, we currently apply a polynomial cost function:

$$C(\vec{q}) = \sum a_i \cdot (\delta q_i)^n$$

where the  $\delta q_i$  are the absolute values of the deviation from rest position of individual degrees of freedom,  $a_i$  being additional weight factors, respectively. Neglecting the effects of inertia, this yields a simplified equation of motion

$$\frac{d}{dt}\vec{q} = K J^T(\vec{q}) F(\vec{q}) - \nabla C(\vec{q}) \quad (1)$$

where  $J(\vec{q})$  is the Jacobian and  $K$  is an optional stiffness matrix. While integrating equation (1), the method converges against a stable state corresponding to a geometrically correct solution. In this sense, the iterative process tracks the position curve of the end-effector.

## 5. Summary

In this paper we have presented a comprehensive model for synthesis of lifelike gestures of various kinds. The model incorporates methods for representing significant spatiotemporal gesture features and planning individual gestural animations, as well as biologically motivated techniques for the formation of arm trajectories. Enhanced methods for rendering animations while executing the resulting motor programs are included which are targeted to enable the gesture performance of an anthropomorphic articulated figure and help to achieve a higher degree of lifelikeness in synthetic agents. While the gesture planning modules, including the gestuary, and the hand motor system were completed in an experimental implementation, the methods for trajectory formation and tracking are currently further elaborated. As the model is particularly conceived to enable a natural cross-modal integration by taking into account temporal synchrony constraints, our mid-range goals include the integration of speech-synthesis techniques as well as run-time extraction of temporal constraints for the control and coordination of gesture and speech.

## References

- [1] J. Cassell. Embodied conversational agents: a new paradigm for the study of gesture and for human-computer interface. In L. Messing and R. Campbell, editors, *Gesture, Speech, and Sign*, chapter 11, pages 203–222. Oxford University Press, 1999.
- [2] M. Costa, L. Zhao, D. Chi, and N. Badler. The EMOTE model for effort and shape. In *Proceedings of SIGGRAPH 2000*, 2000. to appear.
- [3] H. Cruse, E. Wischmeyer, M. Brüwer, P. Brockfeld, and A. Dress. On the cost function for the control of the human arm movement. *Biological Cybernetics*, 62:519–528, 1990. Springer-Verlag.
- [4] J. deRuiter. *Gesture and Speech Production*. PhD thesis, University of Nijmegen, 1998. MPI Series in Psycholinguistics.
- [5] J. Doeringer and N. Hogan. Serial processing in human movement production. *Neural Networks*, 11:1345–1356, 1998.
- [6] J. Cassell, M. Steedman, N. Badler, C. Pelachaud, M. Stone, B. Douville, S. Prevost, and B. Achorn. Modeling the interaction between speech and gesture. In *Proceedings of the Sixteenth Annual Meeting of the Cognitive Science Society*, 1994.
- [7] A. Kendon. Current issues in the study of gestures. In Nespoulous, Perron, and Lecours, editors, *The Biological Foundations of Gestures: Motor and Semiotic Aspects*, pages 23–47. Lawrence Erlbaum Associates, Hillsday N.J., 1986.
- [8] D. Kochanek. Interpolating splines with local tension, continuity and bias control. *Computer Graphics*, 18(3):33–41, 1984. Proc. SIGGRAPH '84.
- [9] S. Kopp and I. Wachsmuth. A knowledge-based approach for lifelike gesture animation. In W. Horn, editor, *ECAI 2000 Proceedings of the 14th European Conference on Artificial Intelligence*, Amsterdam, 2000. IOS Press.
- [10] M. Latash. *Control of Human Movement*. Human Kinetics Publishers, Champaign, IL, USA, 1993.
- [11] T. Lebourque and S. Gibet. Synthesis of hand-arm gestures. In P. Harling and A. Edwards, editors, *Progress in Gestural Interaction: Proceedings of Gesture Workshop '96*, London, 1997. Springer-Verlag.
- [12] D. McNeill. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press, Chicago, 1992.
- [13] P. Morasso. Trajectory formation. In P. Morasso and V. Tagliasco, editors, *Human Movement Understanding*. Elsevier Science Publishers B.V., North-Holland, 1986.
- [14] K. Perlin and A. Goldberg. Improv: A system for scripting interactive actors in virtual worlds. In *SIGGRAPH 96, Proceedings of the 23rd annual conference on Computer Graphics*, pages 205–216, 1996.
- [15] S. Prillwitz, R. Leven, H. Zienert, T. Hamke, and J. Henning. *HamNoSys Version 2.0: Hamburg Notation System for Sign Languages: An Introductory Guide*, volume 5 of *International Studies on Sign Language and Communication of the Deaf*. Signum Press, Hamburg, Germany, 1989.
- [16] M. Sheridan. Planning and controlling simple movements. In M. Smyth and A. Wing, editors, *The Psychology of Human Movement*, chapter 3, pages 47–82. Academic Press, Orlando, Florida, 1984.
- [17] S. Steketee and N. Badler. Parametric keyframe interpolation incorporating kinetic adjustment and phrasing control. In *SIGGRAPH'85 Proceedings*, 1985. Computer Graphics, 19(3).
- [18] R. Taylor. Planning and execution of straight line manipulator trajectories. *IBM J. Res. Devel.*, 23(4):424–436, 1979.
- [19] I. Voss. Empirische Untersuchungen zu sprachbegleitender Gestik im Mensch-Maschine-Dialog. Master's thesis, Faculty of Technology, University of Bielefeld, November 1999.
- [20] A. Watt and M. Watt. *Advanced Animation and Rendering Techniques*. ACM Press, New York, New York, 1992.
- [21] W. Wolovich and H. Elliot. A computational technique for inverse kinematics. In *Proceedings of the 23rd Conference on Decision and Control*, December 1984.