ADVANCED TOPICS IN SCIENCE AND TECHNOLOGY IN CHINA

Yueting Zhuang
Yunhe Pan
Jun Xiao

A Modern Approach to Intelligent Animation

Theory and Practice

ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

Springer

图书在版编目 (CIP) 数据

智能动画的现代方法: 理论和实践= A Modern Approach to Intelligent Animation: Theory and Practice / 庄越挺，潘云鹤, 肖俊著. —杭州：浙江大学出版社，2008.4
ISBN 978-7-308-05454-6

Ⅰ.智… Ⅱ.①庄…②潘…③肖… Ⅲ.计算机应用-动画-制作-研究 Ⅳ.TP391.41

中国版本图书馆 CIP 数据核字 (2007) 第 182268 号

智能动画的现代方法: 理论和实践

庄越挺　潘云鹤　肖俊　著

责任编辑　钱欣平　冯　骏
出版发行　浙江大学出版社
网址：http://www.zjupress.com
Springer-Verlag GmbH
网址：http://www.springer.com
排　　版　浙江大学出版社电脑排版中心
印　　刷　杭州富春印务有限公司
开　　本　787mm×960mm　1/16
印　　张　20
字　　数　719 千
版 印 次　2008 年 4 月第 1 版　2008 年 4 月第 1 次印刷
书　　号　ISBN 978-7-308-05454-6 (浙江大学出版社)
ISBN 978-3-540-73759-9（Springer-Verlag GmbH）
定　　价　110.00 元

浙江大学出版社发行部邮购电话 (0571)88072522

Preface

With the explosive evolution of computer hardware and software technology, computer animation has been no longer a mystery in our daily life. The special effects, photorealistic scenes or vivid characters in movies, video games and commercials are all owed to the modern computer animation techniques. In recent years, computer animation is always a very hot topic in the research field of computer graphics. Researchers, engineers and artists all around the world are seeking ways to broaden computer animation's applications, to make animator's work more convenient and to make the world more colorful by computer animation.

By the title of this book, A Modern Approach to Intelligent Animation: Theory and Practice, the authors intend to give readers two impressions: first, intelligent techniques are the tide of modern computer animation; second, the marriage of theory and practice is quite crucial in computer animation, because neither theory nor innocent effort can produce perfect animation works alone.

In recent animation industry, some widely used commercial tools, such as Maya, 3D Max, Softimage and Motion Builder, have produced abundant wonderful results. Yet to some extent, producing animation with these tools still needs many interactions among experienced animators and lacks intelligent assistance, leading to low efficiency and high cost. On the other hand, many new intelligent theories and techniques that could be applied in the animation production are constantly reported in the top international graphics conferences (i. e. SIGGRAPH, EuroGraphics, CGI, SCA, CASA, etc.). But most of these novel theories and techniques have not been transferred into industrial applications. So from this point of view, the theories of animation are apart from the practices.

Motivated by these phenomena and inspired by the rich experiences and research results in computer animation, the authors think that it is essential to write a book that combines both the theories and the practices in intelligent computer animation. In this book, the authors introduce some original theories and algorithms of intelligent computer animation. Moreover, these abstract theories and algorithms are demonstrated by some prototype systems developed by the authors, which might help readers not only understand but also practice these theories and algorithms as well.

The authors plan to achieve two goals by this book: first, for engineers, this book could be a guidance for developing intelligent computer animation applications; second, this book could be a helpful reference for researchers and a shortcut for new arrivals to the research field of intelligent computer animation.

This book contains 8 chapters. Each chapter is designed to introduce an independent topic and stands alone, which is convenient for readers to select the chapter they are interested in. Chapter 1 introduces the preliminaries of computer animation techniques and the background of intelligent computer animation. Chapters 2, 3 and 4 discuss in details video-based human motion capture techniques, with each chapter focusing on a different variety of methods. Video-based facial animation techniques are discussed in Chapter 5. Intelligent motion data preprocessing and management techniques are presented in Chapter 6. In Chapter 7, intelligent motion data reusing methods are introduced. Chapter 8 describes intelligent approaches for character animation.

This book presents our researches in recent 10 years in the area of intelligent animation. Our sincere thanks go to the National Natural Science Foundation of China who supported our research under Grants No. 60525108, No. 60533090.

The authors are extremely thankful to the following gifted colleagues and students who have contributed significantly to the publication of this book. Prof. Chun Chen, Executive Dean of College of Computer Science, Zhejiang University, gave us numerous valuable suggestions and supports during the whole process. Prof. Fei Wu, the key member of our group, has made great contributions in this research, owing to his creative idea and his high sense of responsibility. Dr. Xiaoming Liu, before his PhD study at CMU, did a lot of excellent work in video-based tracking as his master degree work at ZJU. Dr. Zhongxiang Luo did many talented work in tracking and animation during his PhD study at ZJU. We also wish to thank Dr. Qiang Zhu, Dr. Cheng Chen, Dr. Jian Zhang, Dr. Feng Liu, Dr. Yushun Wang, Dr. Jun Yu, Xin Zhou and etc. We have got so much help and support, consequently it is impossible for us to list all of them. Without their contributions, this book could not have been written.

Last, but not least, our thanks go to the Zhejiang University Press, Dr. Qiang Fu, the head of the press, Ms. Xiaojia Chen, Dr. Xinping Qian, Mr. Jun Feng and so many editors. We are especially thankful to those reviewers who did a lot of work for this book. Their responsibility and encouragement help the birth of this book.

Contents

1

Introduction

In recent years, computer animation has been a highly active research topic and is widely applied in various fields such as movie special effects, advertisements, cartoon, computer games and computer simulation, etc. From a traditional perspective of view, researchers categorize computer animation techniques into the field of computer graphics; however, with the fast development of computer animation techniques and enrichment of animation producing facilities, computer animation is no longer restricted to traditional computer graphics category but rather refers to many research areas, such as image processing, digital signal processing, machine vision and artificial intelligence, etc. , to become an interdisciplinary subject.

The subject of computer animation can date back to the 1960s. As the fast growing of computer hardware and theory of computer graphics, computer animation has penetrated to every aspect of life, including television, movie, education, industry, science, etc. Computer animation techniques can be categorized into two domains: model animation growing with the traditional computer animation, and motion capture animation rising in recent years. In this chapter, we will introduce some commonly used computer animation techniques briefly.

1.1 Traditional Computer Animation Techniques

1.1.1 Key-frame Animation

Key-frame animation can be categorized into key-frame interpolation and spline driven animation. The problem solved in these two methods is to compute the position in a certain frame given by the trajectory of a moving object. The motion trajectory is often represented by parametric splines. Interpolation with equidistance will arouse the problem of ill-proportioned motion. In order to obtain the well-proportioned motion, we have to parameterize the spline. Assume that the length of arch is $s=A(\mu)$, where μ is parameter. In order to calculate the value of parameter for the given arch, we need to solve the function $A^{-1}(s)$, which has no analytic solution. Accordingly we solve it numerically, usually by dichotomy algorithm. In the above solution process, we need to compute the length of an arch with a certain parameter value using Simpson method. Guenter and Parent [1] proposed a Gaussian numerical integration method, which used Newton-Raphson iteration method instead of dichotomy method and stored the pre-computed parameter and arch length values to accelerate computation, to replace Simpson method for computing length of arch in order to build correspondence between arch length and parameters effectively. Watt [2] employed forward difference plus search table to speed up interaction. In situation of non-high precision, this method is very effective.

Key-framing can be regarded as a parameter interpolation problem. For the time control in the interpolation, Steketee and Badler [3] used positional splines and motion splines to control motion parameters, where the positional spline is a function of position with respect to key-frames and the motion spline is a function of key-frames with time. Kochanek and Bartels [4] adopted cubic interpolation spline suitable for key-framing system, where they segmented the tangent vector into incoming vector and outgoing vector and introduced three parameters: tensor $\boldsymbol{t}$, continuity $\boldsymbol{c}$ and offset $\boldsymbol{b}$ to control the spline. The tensor $\boldsymbol{t}$ is used to control the curve degree, continuity $\boldsymbol{c}$ for continuity control of key-frames and offset $\boldsymbol{b}$ for overshooting or undershooting control of key-frames. This method allows animators to adjust the movement of object without adjusting the key-frames.

In the key-framing animation, the interpolation of key-frame parameters is usually independent. In this way, the connections between parameters will result in unnatural motion. Brotman and Netravali [5] employed a method of differential functions from classical mechanics to describe the motion constraints from key-frames. Extra forces will be added in the control to satisfy these constraints. The smooth and natural motion could be obtained by energy control in minimizing the coarseness in trajectory. This method is applied for optimal control in inter-connected key-frames, where the parameters are position, face direction, linear and angular velocity, etc.

In the key-frame interpolation for facing directions of moving objects, Eulerian angle is often used. The rotation matrices of Eulerian angle is unexchangeable, therefore the rotations need to be performed in a certain sequence. Besides, equal variation of Eulerian angle will not lead to equal rotation variation, which leads to the asymmetry of rotation. Furthermore, Eulerian angle may result in lose of freedoms. Aiming at the limitations of

Eulerian angle, researchers introduce the concept of quaternion in the following papers.

Shoemake[6] first introduced quaternion to animation, and proposed to apply Béizer spline on unit quaternion to perform interpolation. This method can be used for local control but is opaque to users. It is difficult for users to control vertices by adjusting quaternion.

Duff[7] used B-spline instead of Bézier spline for smoothing rotation. This method can achieve C^2 continuity, but the generated curves do not pass the control vertices and are hard to control.

In order to solve these above problems, Pletincks[8] implement interpolation by four-point normal curves in space of quaternion to control vertices. This method has advantage in solving the control problem of quaternion.

Barr, et al.[9] used a quaternion with an Eulerian angle constraint to smoothly interpolate the facing directions, which allows users to apply constraint on the end points of trajectory. First they transformed the rotation angle to the quaternion, and then minimized the tangent acceleration of quaternion path in non-Euclidean space, and finally used finite difference and optimization algorithms to solve the energy equations.

The main idea of key-framing is to interpolate the transition frames between several given key-frames. Key-framing can be categorized into two classes such as key-frame interpolation and spline driven animation. In the traditional computer 2D cartoon and 3D animation production, key-framing techniques are widely used. Commercial animation software such as 3D MAX, Maya, and Motion Builder, etc., are all facilitated with key-framing to generate animation sequence.

The key-framing can be effective in scenarios which requiring less accuracy. However, considering the demand of reality and efficiency in modern animation production, key-frame has several deficiencies as below.

- Low efficiency: the key-frames need to be handcrafted by animators to enable computer interpolation.
- High experience requirement: The choice of key-frames will to a large extent affect the final production and efficiency, which requires abundant experience of animators.
- Low reality: In the object movement especially character animation, animation generated by hand-adjusting key-framing are less realistic in behavior or facial expression.

1.1.2 Articulated Animation

In 3D computer animation, character animation is one of the most challenging topics because character body has more than 200 freedoms and very complicated motion, character body is of irregular shapes, muscles are deformed with motion and there are abundant facial expressions. Besides, inconsistent motion will be easily recognized due to the human familiarity with their own motions.

Articulation animation is often represented by articulation model. Articulation model is n-tuple tree, where each inner node is an articulation with a freedom of translation or rotation. For human body, there is only rotation articulation. The motion of articulation is controlled by kinematic or dynamic methods. Forward kinematics sets key-frames by rotation angle to obtain the positions of connected limbs. Inverse kinematics calculates the position of intermediate joints by assigning the position of end joints. Their basic idea is shown below.

Kinematic Methods

Kinematic methods include forward kinematics and inverse kinematics as shown in Fig. 1. 1. Forward kinematics sets key-frames by rotation angle to obtain the positions of connected limbs. In the animation system Pinocchico by Maiocchi and Pernici [10], the authors segmented the real human motion information and stored them in database, and then used animation script to guide the motion of character. Inverse kinematics calculates the position of intermediate joints by assigning the position of end joints, which can be very complicated as the growth of articulation complexity. The cost of solution will be higher and higher, where numerical methods become a feasible solution.

Korein and Badler [11] proposed an intuitive method using hierarchical working space for each articulation segment, which tries to minimize the displacements of articulation positions. The drawback is that users cannot control the obtained results. For a complex articulation structure, the results may not be natural.

Girard and Maciejewski [12,13] used inverse kinematics to generate articulation animation. In this method, users assign the coordinates of feet. The rotation angle from feet to hip will be obtained by solving pseudo inverse Jacobian matrix. This is one of the best methods for generating realistic articulation movement.

An advantage of kinematic methods is that we can set constraints for some key positions of an articulation. For example, when a performer bends one's knees, the feet can be restricted on the floor to lean one's body.

Forward: $A=f(\alpha, \beta)$ Inverse: $\alpha, \beta=f^{-1}(A)$

Fig. 1. 1 Forward kinematics (left) and inverse kinematics (right)

When applying kinematics, we can set a constraint from dynamics. Kinematics and dynamics combination allow animators to be flexible. Isaacs and Cohen [14] proposed a kinematics and dynamics combined system DYNAMO, which has three original characteristics: (1) embed traditional key-framing system into dynamic analysis as a kinematic constraint; (2) able to represent the reaction behavior by surrounding environment through a behavior function; (3) generate specific motion forces through inverse kinematics.

Boulic, et al [15] combined forward and inverse kinematics together to edit articulation motion, where animators could revise the motion towards an object based on user interaction. The key idea of this method is to plug the desired spatial motion into the inverse kinematic mechanism.

Philips and Badler [16] proposed a bipod animal motion control system using interactive kinematic constraints, which could both grasp the motion characteristics and provide balance and stable results.

In summary, inverse kinematic methods are simpler than forward kinematic method, but with high computation cost. Articulation animation requires not only tedious labor of animators but also high computation power. Because it is only a simulation of real human motion, the obtained motion is usually not realistic enough. Therefore some researchers proposed dynamic methods to control articulation motion.

Dynamic Methods

Compared with kinematic methods, dynamic methods can produce more complicated and realistic motions and need less specification of parameters. However, dynamic methods are of high computation cost and hard to control.

Wilhelms and Barsky [17-19] proposed a matrix method applying a generalized force of a freedom only to consider the actual freedom of motion. Therefore the reduced articulation constraint needs not to be adopted in separate equations. The drawback of this method is that the matrix is not sparse, which requires high computation cost to obtain acceleration. Therefore this method is not often used.

Armstrong and Green [20,21] adopted a recursive method from graphic simulation to avoid the reconstruction of matrix. The time complexity of this recursive method is linear with respect to the number of freedoms, which is fast and stable.

Besides the computation complexity, another important problem in dynamic methods is motion control. If there are no effective control methods, users have to provide control instructions such as force or moments, which are almost impossible. Therefore, it is necessary to provide high level control and coordination facilities. High level control depends on lower level control, such as response from collision, impact of friction and damp, articulation constraints to avoid structural artificiality, avoiding unnatural motion, position constraint of a certain joint, etc. A method satisfying the above requirements is pre-processing method, which transforms the required constraints and control to proper forces and moments and plugs into dynamic systems. Witkin, et al [22] proposed a temporal and spatial constraint method by minimizing an object function, which can be solved by conjugate gradient method.

An advantage of dynamics is to simulate the reciprocity between objects, which refers to two problems: occurrence time and response after interaction. Moore and Wilhelms [23] proposed an equation set to describe the momentum conservation and used analytical method to solve the new position and velocity after collision. The collision detection and response increase the reality of simulation but also the computation cost.

Motion Control

The early study of motion control can be found in Zeltzer's work [24]. In his object oriented system, some human motion like walking and jumping was implemented. However, when computing the rotation angle of an articulation, he used kinematics and interpolation between test data, therefore could not realize motion control like alternation of velocity and step size.

Bruderlin and Calvert [25] proposed a mixture method for human walking, which combined object-based and kinematic control techniques. They integrated the cycle motion into a hierarchical control procedure, where the required motion could be specified on the upper level conveniently (such as walking at speed v), and split into low level small tasks, which could be solved by dynamic models. In their experimental system KLAW, after users specify some parameters such as speed, step length and step frequency, the large scale human walking could be generated almost in real-time. In the procedural control method [26], cubic and linear interpolations replace the original dynamic method, keeping nearly all the reality. Therefore, animators could almost control human motion interactively in real-time.

In the real-time human walking model proposed by Boulic and Thalmann [27], the walking model comes from experimental data divided into two levels. The first level was to generate spatio-temporal parameters and the second level used the parameterized trajectory to generate the spatial position of human articulations. Their kinematic method includes the dynamic characteristics of human walking.

Raibert and Hodgins [28] proposed a dynamic control method for motion with legs, where animals could move at different speed and step (running, jogging, galloping and hopping).

McKenna and Zelter [29] proposed a forward kinematic simulation algorithm, whose complexity varied linearly according to the number of articulations and a motion coordination strategy for hexapod animals.

Specifying the motion of articulated animals to achieve an object with the reality of physical laws is one of the purposes of animators. Witkin and kass [30] proposed a new spatio-temporal constraint method, where animators specified content of motion, physical structure of character and physical resources for characters to complete the motion. Based on the description below, plus the Newton law, an optimization problem with constraints is solved to obtain the motion complying with physical laws. The realistic motion generated by this method conforms to some principles of traditional animation. The spatio-temporal constraint leads to a nonlinear problem with constraints, which often has no single result. One of the solutions is to reduce possible trajectories by cubic B-spline basis functions and use constraint optimization to solve the coefficients of B-splines. However, the general solution of these nonlinear optimization problems is often unknown. Therefore Cohen [31] used a symbolic and numerical comprehensive method to realize interactive control, where users could guide the iterative numerical process to convergence. However, as the number of articulations and time complexity increase, the computation cost is still high. This complexity comes from the choice of generalized freedom finite basis. Liu, et al. [32] used wavelet basis to represent the generalized freedom function with time, which had the advantage that we could increase the motion detail as needed to reduce the discrete variables to minimum to get a faster convergence.

In computer animation, it is difficult to build interesting and realistic virtual objects and to keep their control. We need to compromise between complexity and effectiveness of control. For a fixed articulation structure, commonly used optimization methods come from automatic dynamical control systems, e. g. Ngo and Marks'[33] stimulation-reaction algorithm, Van de Panne's sensor network algorithm. These algorithms successfully generated the 2D rigid model motion. For non-stationary 3D objects, Sims [34] obtained autonomous 3D virtual animals, which did not require tedious designing instructions. The morphology and nurse system to control muscles are generated automatically by algorithms.

1.1.3 Facial Expression Animation

Since the pioneering work of 3D human facial animation in 1972, many researches have been done. However, due to the complex structure of facial anatomy and subtle non-rigid motions which are hard to be modeled mathematically, and also due to the human familiarity with facial appearance, this research topic is very difficult. Currently realistic 3D facial animation can be categorized into the following classes.

Interpolation method is the basis for the earliest facial animation. First animators designed several key-frames of facial expressions. Then linear interpolation was used to create transitions between these key-frames. This method mainly depends on the capability of animators and is time-consuming.

Free-form deformation method is to transform facial animation to a surface deformation problem. Several control vertices form a bounding volume, whose motion directly drives facial mesh deformation. This method has less input constraints but requires manual segmentation of face, which is difficult and tedious for ordinary users. Besides, free form deformation does not consider the topology of facial mesh, which leads to some distortions.

Physical model based method is to approximate the anatomical structures of the face, i. e. skull, muscles and skin. The animation from physical models reflects the underlying tissue stresses. Due to the complex topology of human faces, it requires tedious tuning to animate a new face.

Example based method uses motion vectors to estimate the deformation parameters or blending shape coefficients. Predefined morph targets are then blended with respect to the estimated parameters or coefficients. These approaches are attractive for their stable and accurate results. Nevertheless, the processes of handcrafting of morph targets themselves are expensive and time-consuming.

Expression cloning method is to map the displacements of vertices on source model to target meshes. The motion displacements are scaled and rotated with respect to the local detail geometry of source and target meshes for pre-processing. This method works well with the mapping between dense source models and similar target meshes. However, animation by mapping sparse facial motion data to static face models remains unsettled.

Real-time 3D capture method captures both 3D geometry and texture information for facial animations. Photorealistic facial expressions can be reconstructed by deforming the underlying face model with respect to the captured data. The generated head highly resembles the performer. Therefore it is not suitable for animating a given static face model.

1.2 Motion Capture Based Animation Techniques

1.2.1 Definition of Motion Capture

Motion capture is one of the hottest research directions in computer animation field. It includes measuring position and direction in physical space and recording by computers. It captures human body, facial expressions, camera and lighting positions and other elements in the scene. Once the information is recorded in computers, animators can use it as material to generate character animation and virtual scene.

In the synthesized animation, animators usually need to control the path and properties of scene elements by visually simulation. Motion capture animation can synthesize the motion by specifying object path, event time and property control. In the computer generated scene, pure motion capture animation uses the position and direction of real objects to generate movements for synthesized objects. However, due to the constraints by geometric volume matching, precision of motion capture data and requirements of creativity, there is no pure motion capture animation. Even some strict behavior driven systems such as behavior animation have some pre-programming techniques.

1.2.2 Introduction of Motion Capture Techniques

The use of motion capture for computer character animation is relatively new, having begun in the late 1970's, and only now beginning to become widespread. Sturman [35] introduced the brief history of motion capture for computer character animation in detail. Animation by motion capture is mapping captured motion to computer generated virtual objects. Usually, the object of motion capture is the motion of human and animals, where special markers are attached on the joints of objects and tracked position and direction by special hardware.

A motion capture system often includes perception and processing, whose complexity is co-related and compromised. Perception includes active one and passive one. The active perception based motion capture system, applied in fields like sports performance analysis and human computer interface, is simple and widely used, but has high requirement of environment control. The passive perception based on natural signal sources, such as natural light or electro-magnetic waves, does not require wires. The passive perception is applied for intelligent surveillance and human machine interface control. The choice between two perception techniques depends on compromise in complexity.

Motion capture systems usually make some assumptions on capture conditions, such as motion restrictions of body or camera and appearance restrictions of environment and body. Some commercial motion capture systems ranging from simple mechanical system to complicated optical system can be divided into following domains.

Mechanical system consists of potentiometer to measure the position and direction of body articulations. The drawback is that the obtained reality largely depends on the capability and perseverance of animators.

Magnetic system may be one of the most popular systems nowadays. The 3D positions and the relative angles can be acquired by measuring the electro-magnetic field using magnetic sensors. The disadvantages include: (1) sensitive to metal materials in the capture region; (2) restrictions by