Intelligent CG Making Technology and Intelligent Media

Masayuki Nakajima[†] (member)

Abstract In this invited research paper, I will describe the Intelligent CG Making Technology, (ICGMT) production methodology and Intelligent Media (IM). I will begin with an explanation of the key aspects of the ICGMT and a definition of IM. Thereafter I will explain the three approaches of the ICGMT. These approaches are the reuse of animation data, the making animation from text, and the making animation from natural spoken language. Finally, I will explain current approaches of the ICGMT under development by the Nakajima laboratory.

Keywords: Intelligent, CG, reuse, agent, language understanding

1. Introduction

Computer Graphics (CG) is a technique used to create images and animation using computer and graphics peripheral devices. CG has become a popular medium for science, industrial manufacturing, animation, games, communication and medical visualization. Formally, Large-scale investments in technology, skilled personnel and large time allotments were necessary to produce CG animation. However, due to developments in Motion Capturing Systems, the KINECT and other motion acquisition systems, the production of CG animation has become faster, easier and less expensive.

Furthermore, as a result of recent advances in rendering technologies, it is now possible to produce high quality CG animation that is difficult to distinguish from live-action footage. We can conclude that CG animation is nearing technical completion as it becomes indistinguishable from reality.

It can be concluded that in a post-realism era, CG engineers are struggling to find new ways to further develop the field. This paper proposes that the key aspects for the future development and facilitation of CG production are AI (Artificial Intelligence) and Pattern Recognition. This invited paper introduces a new AI-like trial as proposed primarily by Nakajima laboratory.

2. What is Intelligence in CG

In this chapter, I introduce the technology supporting the Intelligence in CG and thereafter I will define Intelligent Media, (IM).

2.1 Intelligent CG Making Technology (ICGMT)

The term of "Artificial Intelligence" or AI is often used to describe the intelligence of machines and the branch of computer science that aims to create intelligence in ways that emulate some functions of the human brain. However "Intelligence" in CG animation differs from this. Intelligence in CG means that the computer aids production efficiently, comfortably and more practically¹⁾. Furthermore, intelligence in CG includes the domain of KANSEI, which adds sensitivity and feelings. The following five aspects provide a definition of the intelligent CG image production concept.

- (1) Production is more readily available to wider user groups without previous animation skills giving them the ability to make high quality CG animations.
- (2) We can generate animation in more ergonomically designed environments providing pleasant work-flows and generally a more pleasant experience thus allowing animators to concentrate on other elements such as directing.
- (3) Correspondence with diversification.

The system corresponds with 2D, 3D and four-dimensional aspects of representation on both mobile devices and large high-resolution screens.

- (4) Delivery of high resolution graphics providing increased artistic freedom with high fidelity anima-

Received September 19, 2012; Revised October 4, 2012; Accepted October 12, 2012

[†] Gotland University (Sweden) and Director of Center for the Study of World Civilization in Tokyo Institute of Technology

tion for abstraction or realism.

- (5) Optimized and semi-automated production is providing shorter production periods at less cost.

This allows non-skilled users to make CG animation cheaply and easily.

Demands in hardware and software specifications require faster computer processing, user-friendly graphics and high-level peripheral devices. In software technology the following should be taken into consideration:

- (1) Advanced image information processing technologies.
- (2) Advanced image expression and direction technology incorporating KANSEI processing, (aesthetic expression and entertainment).

2.2 Intelligent CG Production Technology

The following (1)-(7) technologies are the important image information processing.

- (1) Database Technology.

Database Technology is fast enough to retrieve adequate image information.

- (2) Computer Vision Technology.

Computer Vision (CV) Technology is effective for modeling of real human and animal movement and for the generation of corresponding animations.

- (3) Automatic Language understanding Technology.

Technical capability of generating animation effectively from both written and spoken language.

- (4) Human Interface Technology.

The development of a user-friendly CG making system.

- (5) Standardization Technology.

The use of standardization technology, allows us to operate equally between multiple platforms and projects.

- (7) The Expert System.

We can make animation, by reusing animation assets, which are designed and made by experts in the field.

We can state "Intelligent Media" (IM) is created through the development of the Intelligent CG Making Technology (ICGMT) production methodology. These technologies offer new advances towards the development of future trends of media production providing optimized high-level animations.

3. Trial of the Intelligent CG Making System

In this chapter, I will describe examples of the intelligent CG (image and animation) making research papers. 3.1 describes the reuse of animation data, 3.2 describes making animation from Text and 3.3 describes the making of animation through the use of natural language.

It should be pointed out that the reuse of animation

data is low-level intelligence and the use of natural language is high-level intelligence due to the use of language recognition technologies.

3.1 Reuse in the Animation Production

This section introduces advancements in the reuse of data and assets in animation production, as developed by the Nakajima laboratory²⁾. We have two approaches in animation production. One is making animation though the production of new assets and another is making animation by reusing existing animation sequences and assets.

Making animation through the asset production has been a branch standard, however reuse is new concept in CG production. There are two ways of reuse. The first is reusing the movement data of characters, the other is the reuse of the animation sequence itself. These are described as follows.

- (1) Reuse of Motion Data.

The main approach for the reuse of motion data is the reuse of the actual motion data itself.

- (a) Reuse of MOCAP Data.

The most standard approach for the reuse of MOCAP(Motion CAPture) is the application of the MOCAP data for reuse in several characters³⁾. And reference⁴⁾ is the interactive and hierarchical motion editing system of humanoid character movement using MOCAP data. We proposed MOTION BELT⁵⁾ for the reuse of MOCAP data effectively and visually.

- (b) Using Motion Graph.

When we reuse MOCAP data, Motion Graph is used for the connection of several groups of MOCAP data. Kovar⁶⁾ is proposed as an automatic Motion Graph generation method from the MOCAP data. This detects the adequate movement from the Motion Graph and generates the connection pass between two Motion Graphs.

- (c) Combination with the Key Frame Animation Method.

One of the approaches is the generation of Key frame animation from several kinds of MOCAP data⁷⁾. In this paper, we can also add exaggeration of movement in selected key frames of the animation to improve the expressive quality.

- (2) Reuse of Animation Scenes and Image Data.

A widely used method in production for weekly television programs is the pre-vis checking of the character motion in the early stages of production. This allows for adjustments to be made early on, preventing irregularities later in production. In consideration of this, it is important for the industry to be able to reuse existing

animation sequences to produce new animations directly and effectively without irregularities of motion.

(a) Method of Making Reusable.

To make existing sequences reusable, it is necessary to extract the motion at the same time as the extraction of the model shape from the animation. To achieve this, we used pattern matching and structural matching algorithms for previous animation sequences²⁾. The matching process is as follows:

(I) Binarization and Line Approximation.

The bitmap image from a frame of a 2D animation sequence is finalized and the lines of model shape are filamented. Because the image is drawn using lines, these lines can be transferred and reformed in several characters.

(II) Extracting Lines.

The line of a sequence is extracted by detecting intersection points and start and end points.

(III) Matching the Correspondence Point.

The correspondences of the cross and terminal points in the target image to the original image are set automatically.

(IV) Adjustment of the Amount of Transformation and Correlation is Possible Between Frames.

(b) Reusable Animation Database.

It is necessary to create the reusable animation sequences as a production standard. We create the reusable animation sequence and store them in a database. Our proposed animation reuse method is then ready for applying stored character motion sequences to the user's character.

The developed animation database system has the following functionality.

- (I) Database Registration
- (II) Animation Generation
- (III) Animation Retrieval.

The database allows for the production of a wide variety of different animations²⁾.

3.2 Automatic Animation Making from Text

We consider automatic and intelligent animation making methods from a text scenario as a more advanced approach than reuse.

(1) TVML

Hayashi⁸⁾ has proposed and developed TVML (TV program Making Language) for several years. The TVML has the ability of making TV programs from simple natural language to describe the script of the TV program. The animation for the TV program is then generated immediately when we enter the program script language into the TVML player. A real-time CG character then

Fig.1 The structure of the system implemented with the use of the 'Reuters' website.

speaks by the use of a synthetic voice. In addition, the studio shot can control all camera positioning and movement as well as all lighting. In addition, all postproduction processing and compositing used in TV programs, BGM, and movie productions, are also facilitated in the TVML player in real-time.

(2) T2V System

We proposed T2V (Text-To-Vision) technology capable of generating animation from simple text. We apply T2V to Automatic and Intelligent News Broadcasting System⁹⁾.

We have constructed functioning software that generates TV news shows from Internet news sites using TVML and T2V. **Fig.1** shows the structure of the test system implemented with the use of the 'Reuters' website. This system extracts HTML data from the top page of the site and analyzes it and divides it into the corresponding number of news articles. It then extracts the title, the main body and the main jpeg image for each article respectively. The system then creates a T2V script from the HTML news text. The script is then converted to a TVML script, which includes and formats visual and audio effects such as the CG announcer, the news show setup, sound effects, superimposed graphics etc. Finally the TVML engine plays back the TVML script and delivers the full-CG news show animation without pre-render waiting. It can also support multi-language operation capable of speaking virtually every language.

3.3 Automatic Animation Making from Natural Language

There is much research on making 3D scenes from texts written in natural language or animated agent systems which can interact with users through natural language¹⁰⁾¹¹⁾.

There are two approaches. One is generating still images and another is generating CG animation. When we produce a still image from natural language, it is necessary to generate a depiction of a scene more precisely. Namely the identification of the noun to point to the object of the sentences and an analysis of the predi-

cate expression including position relations become more important. On the other hand, the case of the generation of the animation by natural language, analysis of the verb and adverb in the sentence become more important for automatic generation of real action in the character.

The Pennsylvania University Group has written several papers proposing animation production instructed by natural languages which can be found at this URL.

<http://www.cis.upenn.edu/~hms/publications.html> named "The Center for Human Modeling & Simulation."

(1) SPRINT

SPRINT(SPatial Representation INTerpreter) is a system that makes 3D scenes from natural language¹⁰⁾. This system focuses on the spatial constraints in a text that describe a scene and determines the location of objects. A potential model is used to express the vagueness of spatial constraints. The potential model used in SPRINT is becomes very complicated when several spatial constraints are combined. The potential model can treat several spatial constraints at the same time like a Boolean expression.

(2) WordsEye System

WordsEye makes 3D scenes from natural language proposed by ATT laboratory¹¹⁾. Namely, WordsEye generates 3 dimensional animations according to English text entered into the system. At first, the input text is grammatically and semantically analyzed using "The Natural Language Analyzer". Next, 3D animation is generated using prepared tagged 3D polygon objects according to the analyzed text. WordsEye also gives spatial tags, which assign a function to a part of the object. For example, the "top surface" tag is assigned to the seat of a chair. The spatial tags are used to determine the location of the objects.

(3) Smart Avatars

A natural language interface should be powerful enough to express conditional instructions and hypothetical situations. Smart Avatars are virtual human representations controlled by real people¹²⁾. Given instructions interactively, Smart Avatars can act as autonomous or reactive agents. During a real-time simulation, a user should be able to dynamically refine his or her avatar's behavior in reaction to simulated stimuli without having to undertake a lengthy off-line programming session. One promising and relatively unexplored option for giving runtime instructions to virtual humans is a natural language based interface. After all, instructions for real humans are given in natural language, augmented with graphical diagrams and, occasionally,

animations. In the paper¹²⁾, the Badler group introduce an architecture, which allows users to input immediate or persistent instructions using natural language and to then see the agents' resulting behavioral changes in the graphical output of the simulation. They have therefore implemented an architecture, which allows users to input instructions using natural language sentences. These instructions can range from specific instantaneous commands, like "Sit down," to very general standing orders, like "Drive abandoned vehicles to the parking lot," affording various degrees of autonomy to the avatar/agent.

(4) Animated Agent System

We are also developing an animated agent system, which can interact through Japanese natural language¹³⁾. In recent years, there has been considerable interest in simulating human behavior in both real and virtual world scenarios. If simulated agents or robots could understand and carry out instructions expressed in natural language, they could vastly improve their utility and extend their area of application. However in general, linguistic expressions have ambiguity and vagueness. It is thus often hard to resolve the ambiguity in an automatic manner. In this work, we are focusing on the problem of using natural language for command driven motion generation¹⁴⁾. At first, to express the constraints specified explicitly by the user, with those implied by the virtual character's body and the surrounding environment, into a uniform representation; and second to develop a system that uses this representation in order to generate smooth agent animation consistent with the constraints.

(b) System Overview.

Using speech, a user can command the agents to manipulate objects in the space. The current system accepts simple Japanese commands, such as "Tsukue no mae ni ike" (Walk to the table) or "Motto" (Further). The agent's behavior and the subsequent changes in the virtual world are displayed to the user as a three-dimensional animation.

Fig.2 illustrates the architecture of the system. The speech recognition module receives the user's speech input and translates it to a sequence of words. The text/discourse analysis module analyzes the word sequence to extract a case frame, thus extracting the user's goal and passing it over to the planning modules, which then build a plan to generate the appropriate animation. We separate the planning into two stages; macro and micro planning to account for the differences in repre-

Fig.2 System architecture.

Fig.3 Experimental result of "Natchan ha aoi tsukue no mae ni ike" (Natchan, go in front of the blue table).

sensation. During the macro planning, the planner needs to know the qualitative properties of the involved objects depending on their size, location and so on.

(b) Experimental Result

Using our character agent system, we generated a variety of different user command-driven animations. **Fig.3** shows the result when "Natchan ha aoi tsukue no mae ni ike" (Natchan, go in front of the blue table.) is given as a user's input. The gradation coloring of the floor shows the value of the potential field and the line signifies the generated trajectory.

4. Current Approach in Nakajima Laboratory

Finally in this paper, novel approaches by the use of the Intelligent CG Making Technology (ICGMT) production methodology as proposed by Nakajima laboratory are reported.

(1) Agent Movement in Accordance with Social Relationships to One Another.

The analysis of the non-verbal communication between people via the management of their Personal Spaces (PS) shown in **Fig.4**, gives an idea about the nature of their relationship. We propose a mathematical model for the concept of Personal Space and demonstrate its application in simulating the non-verbal communication between agents in virtual worlds and also in Human Computer Interaction¹⁵. We focus on the communication between two virtual agents. We assume three different types of relationships: business relationship, friendly relationship and relationships between strangers. Two virtual agents behave under our proposed PS. Our method can simulate the behavior of virtual agents according to the relationship between them. We use the Personal Space model to: 1) automatically control the speed of an agent when it is moving towards another agent, which it is going to meet, 2) automatically find a natural stopping distance in front of the target. The proposed method enables the modeling of the agent's mobile territory and his relationship with others.

Fig.4 Definition of the Personal Space.

Results of this work can be applied to modeling the behavior of autonomous virtual agents and avatars in virtual worlds, as well as individual movement in social groups and crowds.

(2) Learning System Approach in NPR

We proposed a highly practical framework in Non Photo-realistic Rendering, (NPR) for painterly rendering that is an automatic and intelligent approach to creating artistic paintings¹⁶⁾. We mainly focused on creating oriental ink painting ("Sumi-e"), which is one of the oldest artistic brushworks and particularly popular in Asian countries. The main research challenge in oriental ink painting is stroke placement and how to distribute strokes with realistic brush textures in desired shapes. However, this process tends to cause unsatisfactory defects such as non-natural stretching in textures and undesirable folds or creases appearing inside corners or curves. To address these classic problems, we introduced an intelligent learning agent theory for the art of painting. This work contains the design of a brush agent and the development of two agent's learning algorithms for automatic stroke drawing.

(a) Design of Brush Agent

We designed the brush as an intelligent agent for deciding behaviors of drawing strokes in the framework of reinforcement learning (RL), and formulate this sequential decision-making problem as a Markov decision process (MDP)¹⁷⁾. We then provided an elaborate design of an environment, actions, states, and rewards specifically tailored to the Sumi-e agent. Under this framework, the stroke generation is formulated as an

Fig.5 Results of automatic photo conversion into an oriental ink style.

optimization problem to find an optimal policy of controlling the brush so as to maximize a cumulative reward during the process of drawing strokes. RL methods help to build an artificial agent that learns how to optimize its behavior in an unknown environment, without requiring prior knowledge.

(b) Model-based Learning of Sumi-e Agent

Model-based methods require an explicit model of the Markov decision process, including the transition dynamics and the reward function. Model-based methods work offline to produce a policy, which is then used to control the process. Transition dynamics is the model of the environment, which guides the agent in how to move inside the desired shapes. To construct transition dynamics, we begin by defining the space around a shape of a desired stroke by sampling locations. An optimal policy describes a mapping from states to actions to form the optimal brush trajectory, which obtains the maximum cumulative reward. Since this model-based method requires the transition dynamics of a specific shape, this results in a limitation that the optimal policy on a desired shape may not be directly applied to new shapes.

The effectiveness of our proposed learning approaches was demonstrated through simulated Sumi-e experiments shown in **Fig.5**. Statistical comparison with mainstream commercial software through the user study showed that the performance of our methods is closer to the real paintings than the commercial software.

5. Conclusion

In this invited paper, I introduce the Intelligent CG Making Technology (ICGMT) system as a new trend in CG fields. However there are other papers related to "Intelligent CG media" which are not included in this paper due to format constraints. I hope to introduce these in coming papers. Furthermore I aim to develop our Agent System providing increased AI capabilities in the animated characters for future use in applications such as Robotics.

I strongly hope many researchers and engineers are active in the development of the Intelligent Media fields.

Thanks to Prof. Steven Bachelder and Prof. Masaki Hayashi, Gotland University for assistance and support in this paper.

[References]

- 1) M. Nakajima: "Computer and AI", Journal of Artificial Intelligent Society, **19**, 1, pp.10-14 (2004)
- 2) F. Sumi, M. Nakajima "A Production Method of Reusing Existing 2D Animation Sequences", CGI 2003 pp.282-287(2003)
- 3) A. Witkin and Z. Popovic: "Motion Warping", Proceedings of SIGGRAPH (1995)
- 4) J. Lee and S.Y. Shin: "A Hierarchical Approach to Interactive Motion Editing for Human-Like Figures", Proceedings of SIGGRAPH(1999)
- 5) H. Yasuda, R. Kaihara, S. Saito, M. Nakajima: "Motion Belts", Visualization of Human Motion Data on a Timeline IEICE Transactions, **91-D**, 4, pp.1159-1167 (2008)
- 6) L. Kovar, M. Gleicher and F. Pighin: "Motion Graphs", Proceedings of SIGGRAPH2002 (2002)
- 7) K. Pullen and C. Bregler: "Motion Capture Assisted Animation: Texturing and Synthesis", Proceedings of SIGGRAPH2002 (2002)
- 8) M. Hayashi, H. Ueda, T. Kurihara and M. Yasumura: "TVML (TV program Making Language) - Automatic TV Program Generation from Text-based Script -" Imagina'99 proceedings (1999)
- 9) M. Hayashi, M. Nakajima and S. Bachelder: "International Standard of Automatic and Intelligent News Broadcasting System", Nicograph International in Indonesia (2012)
- 10) Yamada and Nishita: The analysis of the Spatial Descriptions in Natural Language and the Reconstruction of the Scene, IPSJ, **31**, 5, pp.660-672 (1990)
- 11) B. Coyne, R. Sproat: "WordsEye: An Automatic Text-to-Scene Conversion System", SIGGRAPH 2001 proceeding, pp.487-496 (2001)
- 12) A. Bindiganavale, W. Schuler, J.M. Allbeck, N.I. Badler and A.K. Joshi: "Dynamically Altering Agent Behaviors Using Natural Language Instructions", the 4th International Conference on Autonomous Agents 2000 (AGENTS 2000), pp.293-300 (2000)
- 13) M. Nakajima: "Autonomous Agent Action and Semantics", 1st International Symposium on Shape and Semantics, pp.1-5 (2006)
- 14) S. Funatsu, T. Koyama, S. Saito, T. Tokunaga and M. Nakajima: "Action Generation from Natural Language", Advances in Multimedia Information Processing-PCM 2004, pp.15-22 (2004)
- 15) T. Amaoka, H. Laga, M. Yoshie, M. Nakajima: "Personal Space-based Simulation of Non-Verbal Communications", Entertainment Computing, **14**, pp.1-36 (2011)
- 16) N. Xie, H. Laga, S. Saito, M. Nakajima: "Contour-driven Sumi-e rendering of real photos", Computers & Graphics, **35**, 1, pp.122-134 (2011)
- 17) N. Xie, H. Hachiya and M. Sugiyama. "Artist Agent: A Reinforcement Learning Approach to Automatic Stroke Generation in Oriental Ink Painting", Proceedings of 29th International Conference on Machine Learning (ICML2012), pp.153-160 (2012)

Masayuki Nakajima received Dr. Eng. degree from the Tokyo Institute of Technology, Tokyo, Japan in 1975 and has been Professor at the Department of Computer Science, Graduate School of Information Science & Engineering, Tokyo Institute of Technology during 1997-2012 March. He began working at the Department of Game Design, Technology and Learning Processing, Gotland university, (Sweden) in April 2012.