F-024

Using CLIPS and C++ to Realize the Rule-based Reasoning in Computer Animation Creation

JINHONG SHEN† SEIYA MIYAZAKI‡ TERUMASA AOKI† HIROSHI YASUDA†

1. Introduction

This paper describes the rule-based reasoning in computer animation creation as a part of system EMM (Electronic Moviemaker) we are implementing realized by using language C++ and expert system language CLIPS.

Current advanced broadband network techniques already enable us to access multimedia streams and display them on personal computer. A great challenge on Web communication lies in giving such an environment that enables any people to make his presentation and deliver it uncomplicatedly. A low-cost easy-to-use electronic moviemaker has good entertainment and education marketplace. After analyzing the feasibility of realization, we came to the conclusion that it is reasonable to automatically visualize a verbal screenplay using relevant sound motion picture with visual effects like real image, 3D animation, or their composition, where real images are extracted from digital video (movie, animation, TV programs, etc.) library [1], [2].

EMM aimed at visualizing user's inputted screenplay words by sound motion picture, which is belonging to the project DMP (Digital Movie Producer). If there are suitable video clips in video database or video web library, the required clips will be extracted from the database/library, otherwise, 3D animation will be created based on cinematic knowledge, so that at present it is feasible to automatically make motion picture with visual effects of computer animation and real images, and their simple composition, where content-based retrieval techniques are used to automate the procedure of video retrieval.

Those works on interpreting text-based input into dynamic visualized presentations are in progress, like *Virtual Director* in [3] and *Mario* in [4]. Virtual Director aimed to visualize simple scenario in virtual scene and animation. Mario focused on automatic camera control to create 3D animation from annotated screenplay. They were both designed through KB approach but not systems for home movie making usage. Other methodologies employing AI for computer animation have been put forward. In [5], [6], domain knowledge base was applied in automatically generating animation focusing on camera shot design while in [7] animation creation focused on human gesture. Cognitive modeling for intelligent agent was employed by John Funge et al to solve the same cinematic problem [8].

By AI approach, the digital filmmaking procedure in production system EMM can be automated dependent on cinematic knowledge. The filmmaking knowledge base contains domain knowledge about objects, color, lighting, scene, shot, also contains spatial-temporal knowledge. Automated digital moviemaking may generate shot sequence from screenplay that describes abstract relationship between objects (such as *two talk*) or concrete actions of characters (such as *stand*) in various shots (such as *close up*). A virtual director achieves user's intentions by knowledge-based (KB) approach through setting a scene, determining the corresponding shot types and shot sequence, and planning virtual camerawork dependent on the cinematic expertise stored in a domain knowledge base. We model the filmmaking knowledge and rule-based reasoning strategies in CLIPS and embedding CLIPS into C++.

The next section will have an overview of EMM system architecture. In Section three, the design of screenplay user interface is introduced. Section four expounds how to realize the automation of video retrieval and animation generation by using AI approaches from a film director's point of view. Finally, I will have a discussion about my present work.

2. EMM System Structure

Virtual Director in Fig. 1 is embedded as a subsystem in the integrated system environment to realize the automation of 3D animation and video retrieval. He is responsible for the visual aspect of screenplay dependent on knowledge of plot structure in KB, giving commands for the dramatic structure, pace, and directional flow elements of the sounds and visual images to visualize the event. Supposing the existence of a library that stores 3D models and actions mentioned in the script, it is possible to combine objects and actions according to the screenplay and to choose optimal placement for the camera automatically. Composition, the location of characters, lighting styles, depth of field and camera angle are all determinant factors in the formulation of the visual information. Movie player assembles the resultant plan created by inference engine into images. *Virtual camera* records the frames that are to be played as a still or a sequence of images.

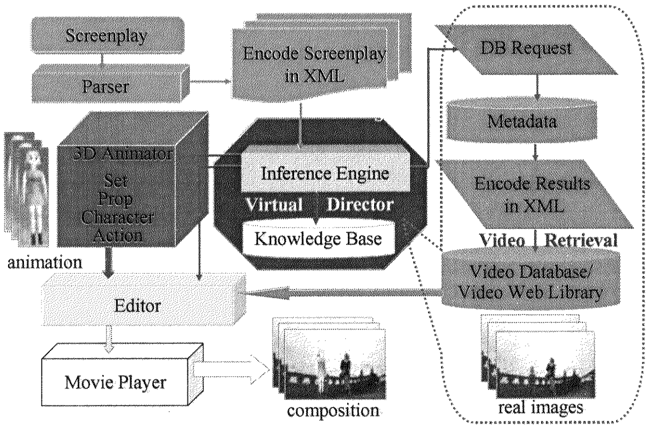


Figure 1. System Architecture Diagram

† School of Engineering, the University of Tokyo
{j-shen, aoki, yasuda}@mpeg.rcasu.u-tokyo.ac.jp
‡ Matsushita Electric Industrial Co., Ltd.
miyazaki.seiya@jp.panasonic.com

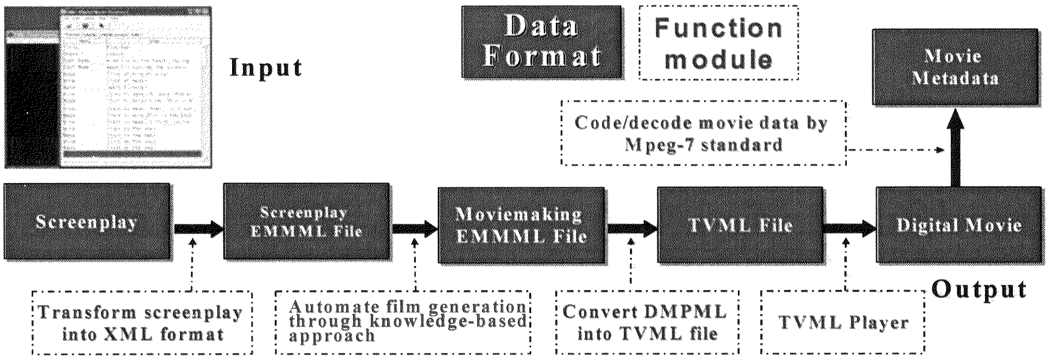


Figure 2. EMM System Data Flow Diagram

The pipeline production drawn in Fig. 2 shows the data flow beginning from screenplay input to the generated movie output, covering all main data conversion step by step.

3. Screenplay GUI

Some works on translating verbal presentations into visualized presentations are in progress. At & T' is making a system named WordsEye [11] for automatically converting text into representative 3D static scenes. As they said, "Natural language is an easy and effective medium for describing visual ideas and mental imaginary." However, fully capturing the semantic content of language in movies is infeasible because linguistic descriptions tend to be at a high level of abstraction and there will be a certain amount of unpredictability in translating the script into the visual effects.

We chose screenplay as input because it is a formal language for filmmaking, implying the lots of rules of film that are almost invisible by audience. When we design the form of screenplay, the first feature to consider is *common user access*: it should be easy-to-learn and easy-to-use for non-professionals and non-artist users such as school-age children. On the other hand, 3D motion picture is far more difficult to be realized than 3D static scene, decided by synthetic techniques involving the fields of Linguistics, Artificial Intelligence, Computation, and Computer Vision so that the screenplay design is also based on current technology. Based the above considerations, we designed two types of screenplay formats - EventSp and MarkupSP.

3.1 EventSP (Event ScreenPlay)

When we carefully analyze a visual message, we find images become real property the mind and remember only when language express them. Symbols are used to retain events and ideas in our memory so that one kind of screenplay we designed is called EventSP that can describe abstract relationship between objects (such as *two talk* event).

There were some typical works on applying film theory for computer graphics generation. Christianson et al. adopted the notion of *film idioms* from film theory and formalized them into a sequence of shots [12]. He et al. encoded the film idioms into hierarchically organized finite state machine applied in real-time system [13]. Amerson & Kime proposed a system *FILM* (Film Idiom Language and Model) for real-time camera control in interactive narratives [14]. In our system, intelligent rule-based reasoning is employed which will be demonstrated in the next section.

3.2 MarkupSP (Markup ScreenPlay)

The mind's picture is a combine of the perceptual elements of color, form, depth and movement combined with the verbal thoughts. To describe their imagery concretely, user should be allowed to add their controls in screenplay such as actions of characters (e.g. *stand*) or layout (e.g. on the left) in various shots (e.g. *close up*). These controls are included in filmmaking techniques involving the four aspects: 1) *mise-en-scène* (*what to shoot*) which involves setting, lighting, figures, 2) *cinematograph* (*how to shoot it*) which involves camerawork –

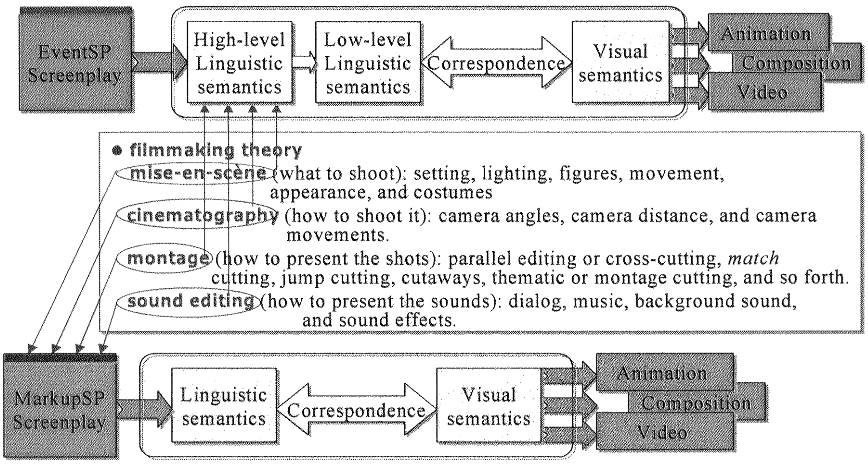


Figure 3. Principle of Visualization for the Two Different Screenplay Formats by Sound Motion Picture

camera angle, camera movement and camera distance, 3) *montage* (how to present the shots), e.g., fade in/out, parallel editing and 4) *sound edition* (how to present the sounds), e.g., dialog, music, background sound from film theory.

When designing the screenplay format, an important issue that must be considered is the possibility of automatically generated visual effects made of various media (e.g. animation, video) and modalities (e.g. music, talk). Human beings perceive the world via the five senses of touch, hearing, sight, smell and taste. Film creates a five-dimensional world in the two-dimensional screen of sight and sound modes composed of different modalities. A modality indicates a particular form of a communication mode. For example, noise, music, and speech are modalities of the sound mode. For modalities of smell and taste, their expressions in sound motion picture may be realized by speaking ("rotted apple") or image (rotted apple).

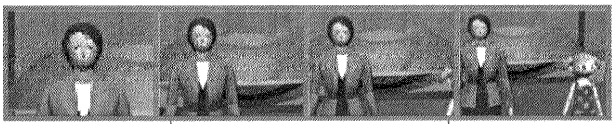
Figure 3 shows the principle of visualization by sound motion picture beginning from the two different screenplay formats. Since the most important function of movie is to rightly express user's feelings, meanings and emotion toward audience, photorealism (realistic style in two respects: realistic picture or moving in realistic fashion) is not required.

4. Visualize Screenplay

4.1 Rules Presentation about Shooting

Dynamic picture generation concerns object description, objective movement, and camerawork. CLIPS provides a cohesive tool for handling a wide variety of knowledge with supports for three different programming paradigms: rule-based, object-oriented and procedural so that it can fulfill the above mentioned needs for high-level tool to program the generation of digital movie.

To write rules about objects (e.g. characters or props) in 3D graphics space, it is necessary to describe all the scene information in COOL (CLIPS Object-Oriented Language) [15]. CLIPS deals with the motion of object in the way of pattern matching. To encode shot sequence, an example is showed in Fig. 4 composed of shots transformed from medium shot to two-shot through dolly shot (concerning camera movment).



Medium shot　Dolly shot　Two-shot

Figure 4. A Shot Sequence.

In rule-based language, a rule is a concise description of a set of conditions and a set of actions to take if the conditions are true. Film rules about above shooting can be written in *defrule* construct of CLIPS as:

```
(defrule shot-type "A rule of intention shot"
   ;track two persons in half length size
   ;make a medium shot and a two-shot
  (track-two-half -front)                           ; If
  =>                                                ; Then
  ;assert (medium-shot) (dolly-shot) (two-shot)
  (assert (action MS) (action DS) (action TS)))
```

where two parts are separated by the '=>' symbol (means 'then'). The first part consists of the LHS left-hand side *pattern* (track-two-half-front) which is used to match facts in the knowledge base while the second part consists of the RHS right-hand side *action* (MS) (DS) (TS) that contain function calls. The rule of shot selecting will be activated when the fact (track-two-half-front) appear in the knowledge base. When the rule executes or *fires*, the functions (action MS), (action DS), and (action TS) are called. Annotation begins with symbol ';'.

4.2 Reasoning Shot Sequence from EventSP

Heuristics about making a sequence of shots involves the techniques of montage and sound related to image, and another unit in film named event. *Event* is an important primitive action unit in camera planning procedure such as "a private conversation between two characters". This section will demonstrate how to visualize EventSP by making sequence of two-talk event.

Place: park
Time: day
Character: a boy, a girl
Prop: trees
Event: talk (or two talk)
Talk: the boy
  Why did not you wear that yellow shirt that your sister gave you for your birthday.
Talk: the boy
  It looks terrific on you.
Talk: the girl
  I love the shirt, but it missed two buttons.

Figure 5. Two-talk Event Described in EventSP

To stage face to face two-talk described in figure 5, the virtual director determines five basic shots from nine camera positions, selects shots (Fig. 6, Tab. 1) from the set and arranges them in order dependent on dialogues according to the following planning rules (involving continuity cutting):

(a) If character A and B have a private conversation, five basic shots could be used: *two-shot* (default size: *very long shot*), *profile shot* (default size: *close-up)*, *over-the-shoulder shot*, *point-of-view shot* (default size: *close-up)*, and *angular shot* (default size: *close-up)*.

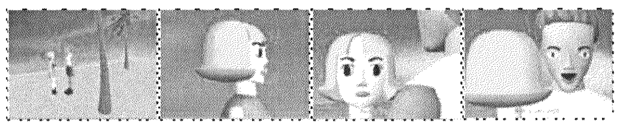
(b) If both character A and B are silent, use two-shot.

(c) If character A talks, select one least used shot by A from the set of basic shots.

(d) If character B talks, select one least used shot by B from the set of basic shots.

(e) If character talks, OTS should be selected first.

(f) If the selected shot is not OTS, it should be set before OTS in the shot sequence.



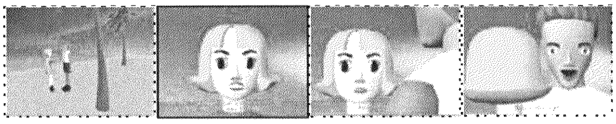
1.Two-shot VLS 2. Profile-shot CU 3. OTS (facing A) 4.OTS (facing B)

Figure 6. Two-talk Shot Sequence

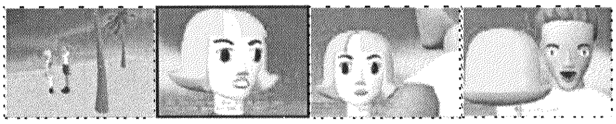
Inference procedure		
Premises	Actions	
	Set of basic shots	Rule
Two talk	Two-shot & Very large shot (VLS) Profile-shot & Close-up (CU) Over-the-shoulder shot (OTS) Point-of-view shot (POV) & Close-up Angular shot & Close-up	(a)
	Shot sequence	Rule
Silence	**1. Two-shot VLS**	(b)
	Shot sequence	Rule
A talks	1. Two-shot VLS **2. OTS (facing A)**	(c) (e)
	Shot sequence	Rule
A talks	1. Two-shot VLS **2. Profile-shot CU** 3. OTS (facing A)	(c) (f)
	Shot sequence	Rule
B talks	1. Two-shot VLS 2. Profile-shot CU 3. OTS (facing A) **4. OTS (facing B)**	(d) (e)

Table 1. Staging Dialogue Sequence for Two Characters

As showed in Fig 7, there are other possible results according to the above planning rules (a) – (f) because of some stochastic process in shot selection. If adding new shot sizes to the shots maybe used, there will be more possible shot sequences resulted for the same event. For example, when POV shot and angular shot in figure 7 are changed into medium shot size, another two results will be obtained.



1.Two-shot VLS 2. POV shot CU 3. OTS (facing A) 4. OTS (facing B)



1.Two-shot VLS 2.Angular shot CU 3. OTS (facing A) 4.OTS (facing B)

Figure 7. Other Possible Two-talk Shot Sequences

5. Conclusion

In recent years, filmic techniques have been extended to a degree possible with live actors shot in real time, but the commercial software tools used to make CG movie are not common user access. The filmmaking knowledge base of our digital moviemaking system EMM contains domain knowledge about objects, color, lighting, scene, shot, also contains spatial-temporal knowledge. The cinematic knowledge-based environment instead of programming enables nonprofessionals to make their own digital movies easily. Moreover, the virtual director can intelligent generate the visual scene not lonely dependent on film rule of thumb, but also stochastically "image" the 3D virtual world the same as real humans, so that the results are more plentiful and interesting, not just pre-scripted contents.

Reference

[1] Shen Jinhong, Seiya Miyazaki, "Terumasa Aoki, Hiroshi Yasuda, The Framework of an Automatic Digital Movie Producer". 2002 AVM Conference of IEICE, IEICE Technical Report, 102, 517, Nagoya, Japan, Dec 2002, p.p. 15-18.

[2] Shen, Jinhong; Miyazaki, Seiya; Aoki, Terumasa; Yasuda, Hiroshi, "A Prototype of Cinematic Rule-based Reasoning and Its Application". The 9th International Conference on Information Systems Analysis and Synthesis: ISAS '03 (CCCT2003), VI, Florida, USA, Aug 2003, p.p. 60-365.

[3] Konstantinos Manos, Themis Panayiotopoulos, George Katsionis, "Virtual Director: Visualization of Simple Scenarios". 2nd Hellenic Conference on Artificial Intelligence, SETN-02, Thessaloniki, Greece, April 11-12, 2002

[4] Doron Friedman, "Yishai Feldman, Knowledge-Based Formalization of Cinematic Expression and its Application to Animation". Proc. Eurographics 2002, Saarbrucken, Germany, Sept. 2002, p.p.163-168,

[5] Kevin Kennedy, Robert. E. Mercer,"Planning animation cinematography and shot structure to communicate theme and mood". Proceedings of the 2nd international symposium on Smart graphics, June 2002, p.p.1-8.

[6] Szarowicz, A., Amiguet-Vercher, J., Forte, P., Briggs, J., Gelepithis, P.A.M., Remagnino, P., "The Application of AI to Automatically Generated Animation, Australian Joint Conference on Artificial Intelligence", AI'01, AI 2001:Advances in Artificial Intelligence, Adelaide, Dec 10-14, 2001, p.p. 487-494

[7] Stefan Kopp, Ipke Wachsmuth, "A Knowledge-based Approach for Lifelike Gesture Animation". In W. Horn, editor, ECAI 2000 Proceedings of the 14th European Conference on Artificial Intelligence, IOS Press, Amsterdam, 2000, p.p. 120-133

[8] John Funge, Xiaoyuan Tu, Demetri Terzopoulos, Cognitive Modeling: Knowledge, Reasoning and Planning for Intelligent Characters. *Computer Graphics Proceedings, Siggraph,* 1999, p.p. 29-38

[9] H.J. Zhang, John Y. A. Wang, and Yucel Altunbasak. "Content-based video retrieval and compression: A unified solution", In Proc. IEEE Int. Conf. on Image Proc., 1997.

[10] Salwa, "Video Annotation: the role of specialist text". PhD Dissertation, Dept. of Computing, University of Surrey, 1999

[11] Bob Coyne, Richard Sproat, "WordsEye: an automatic text-to-scene conversion system, Proceedings of the 28th annual conference on Computer graphics and interactive techniques", Aug. 2001, p.p. 487-496

[12] Christianson, Anderson, Wei-he, Salesin, Weld, and Cohen, "Declarative Camera Control for Automatic Cinematography". AAAI/IAAI, Vol. 1, Portland, Oregon, 1996, p.p. 148-155

[13] Li-wei He, Michael F. Cohen, David H. Salesin, "The virtual cinematographer: a paradigm for automatic real-time camera control and directing". Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, New Orleans, Louisiana, United States, August 1996, p.p. 217-224

[14] Amerson, D. and Kime, S., "Real Time Cinematic Camera Control for Interactive Narratives". In the Working Notes of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment, Stanford, CA, 2001

[15] Jinhong Shen, Seiya Miyazaki, Terumasa Aoki, Hiroshi Yasuda "Using Rule-Based Language CLIPS to Design An Intelligent Motion Picture", Forum on Information Technology (FIT2003), Sapporo, Hokkaido, Japan, Sept 10-12, 2003