Intelligent Motion Picture Making Software System

Jinhong Shen† **Seiya Miyazaki**†† **Aoki Terumasa**† **Yasuda Hiroshi**†

† The University of Tokyo, Gwangju, 153-8904 Tokyo, Japan

{j-shen , aoki, yasuda}@mpeg.rcast.u-tokyo.ac.jp

†† Matsushita Electric Industrial Co., Ltd, 140-8632 Tokyo, Japan

miyazaki.seiya@jp.panasonic.com

Abstract

This paper describes an easy-to-use tool that provides an integrated environment for dealing with the complete process of creating digital film. To make the creating process automated, we propose to employ a verbal screenplay as input form. A software system EMM (Electronic Moviemaker) is developed to visualize user's input screenplay words by sound motion picture with effects of real image, three-dimension animation, or their composition. A virtual director achieves user's intentions by rule-based approach through setting a scene, determining the correspondent shot types and shot sequence, and planning virtual camerawork dependent on the cinematic expertise stored in a domain knowledge base, where real images are extracted from digital video by applying advanced content-based retrieval techniques, "video" stands for all kinds of clips of these visual presentations. The resultant movie is showed on TVML player and all visual information is encoded in XML.

Key words: Filmmaking Knowledge Base, Rule-based Reasoning, 3D Animation, Video Retrieval

1. Introduction

Motion picture is the sort of multimedia full of vast information, enabling us to absorb information the most interestingly, conveniently and effectively compared with pure text and image. Broadband superhighway already allows for the transmission of vast amounts of digitized information at very high speed and makes the transmission of high-resolution video images as quick and easy as voice transmission. In present age Internet and TV are the best ways of entertainment and education, and particularly, Internet is more popular for younger and will be used by more people in the future. One of the most important research directions on Web communication lies in giving such an environment that enables any people to make his presentation and deliver it uncomplicatedly. However, a variety of limits exist in current approaches to make digital motion picture: the time, the labor, the material, etc, these make the moviemaking on computer very expensive. A low-cost easy-to-use moviemaker system has presented a challenge in front of us. The research introduced in this paper aims at developing an *easy-to-learn and easy-to-use* desktop software tool by which a nonprofessional general person can make his own visual contents and deliver it uncomplicatedly. It has good entertainment and education markets.

The software system *EMM* (*Electronic Moviemaker*) in Fig. 1 we are implementing for such automated digital moviemaking may generate shot sequence from screenplay [1]. 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

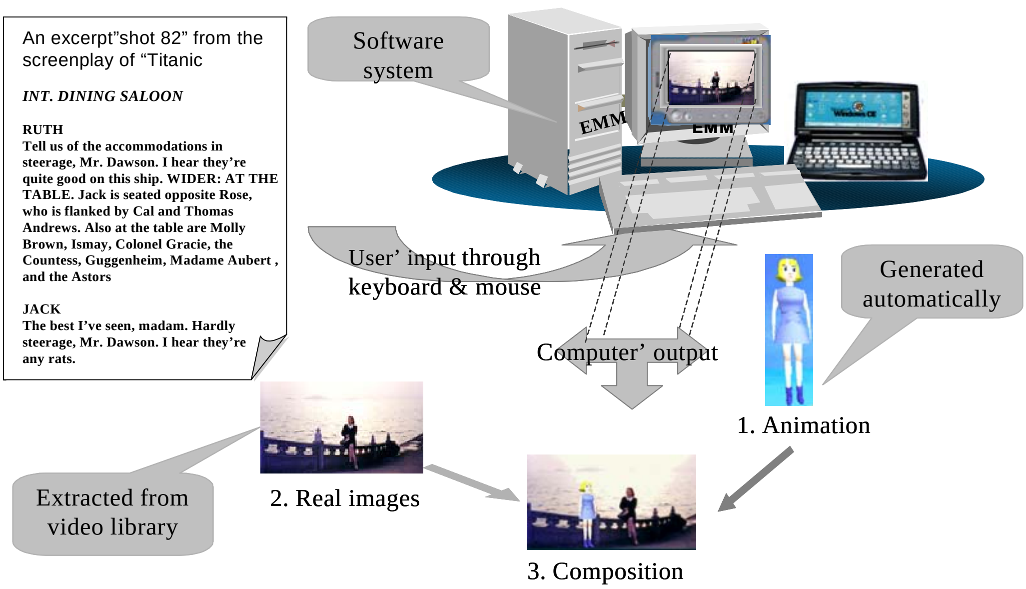


Figure 1. EMM (Electronic-Moviemaker) Platform

animation will be created based on cinematic knowledge.

The next section will have an overview of the related works of other researchers in the field of Artificial Intelligence (AI)-based Animation and automatic video retrieval. Section 3 first analyzes the procedures of animation, film/video generation, and video retrieval, at last gives the approaches to realize the automation in these technical implementations using rule-based method. In section 4 the architecture of EMM system is outlined beginning from screenplay design, then an event example is expounded to show how to automate e-moviemaking by AI approach from a film director's point of view that concerns filmmaking knowledge base building and virtual director's work organizing in detail. The rule-based reasoning is programmed by using language C++ and CLIPS. Finally, I will summary the contributions of my present research and discuss the future work.

2. The State of the Arts and

What are the features of EMM? EMM is used to express user's feelings, meanings and emotion with minimal user intervention corresponding his imaginary in mind written in words so that our software system EMM is different from the most of the virtual world design platforms. Their major focus is about developing interactive 3D models as virtual worlds, which leads to a strong emphasis on the visual aspect of the virtual world in which interactions are attached to 3D models in order to support predefined actions such as animations and teleporting. Our focus is on automation validated by AI technology.

2.1 AI-based Animation Researchers in the field of computer graphics are moving into Artificial Intelligence to find suitable tools for AI animation generation. The **initial computer animation** (e.g. Disney's *CAPS* beginning from 1989) had the process of individual layer creation and composition similar to traditional animation where human animators draw several layers of drawings on *cels* (translucent sheets). The difference is that the successive film frames of computer animation are made by scanning hand-drawn images on 'digital cels', coloring these digital images on computer and compositing them into final image. This layered design permits the reuse of background drawings. But today' computer generated animation technologies are much more complex and heterogeneous, far beyond just creating traditional animation using computer.

Insofar as the widely used **commercial animator packages** like *Maya* and *3D Studio Max* generally consist of three principal components: *modeler*, *animation creator* and *rende*r. This type of computer-generated way extends the tradition animation into the relating three stages of process - *modeling*, *animation* and *rendering*: first building 3D object/character models with respective geometric structures and surface features; then designing how objects/characters move in their environment (defining keyframes in Maya and 3D Studio Max); last, after lighting and camera placing, translating all the data about 3D objects/characters models and environments into 2D images and generating the final animation sequence.

The whole process typically relies on human intervention for the design of resulted animation sequence. Many research groups have worked on higher leveled *motion generation* which controls animation through task-level commands such as "grasp a cup" dependent the above low-level generation of animation systems.

Some groups are working on translating verbal presentations into visualized presentations. AT & T is making a system named *WordsEye* [2] for converting text into representative 3D static scenes automatically. 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. System *Virtual Director* [3] aims at visualizing simple scenario with virtual scene and animation while *Mario* [4] focuses on automatic camera control to create 3D animation from annotated screenplay. Both of them were designed by KB approaches but not for home moviemaking usage. Other AI-based methodologies for computer animation have been put forward. In [5] and [6], domain knowledge base was applied in automatically generating animation which focused on camera shot design while in [7] animation creation mainly concerned human gesture. Cognitive modeling for intelligent agent was employed by John Funge et al to solve the same cinematic problem [8].

2.2 Practical Video and Animation Retrieval

Video retrieval (VR) technology is developed to find interesting video shots/segments from a movie, TV, video database. There are many classes of video applications: *interactive video*, *video-on-demand*, *stock-shot*, *video edition*, etc. For indexing and retrieval of video data, a variety of methods have been proposed such as *text-based*/*metadata-based*, *content-based,* and *integrated* approaches.

There are two main categories of VR methods:

(1) *Text-based Video Retrieval*. This traditional method uses keyword, attribute or free-text to present high-level concepts of video contents that are usually annotated manually. Most frequently used image/video retrieval systems are oriented around text searches, for example, www.google.com where textual annotation was already performed. But the procedure of annotation is so tedious and consuming, and there is no standard for video depiction. These drawbacks made researchers to

explore content-based way for video indexing and retrieval. Besides video media data themselves, metadata also give important video information. Typically metadata contains the video specific data such as video name, date of production, length of video, etc.

(2) *Content-based video retrieval (CBVR) approach.* It concerns the techniques that capture the spatial-temporal distribution of pixels. In this case *content* refers to the properties of image/video data, rather than the meanings viewers percept directly so that it is possible to annotate automatically by computed way. However it is not always successful because there is gap between low-level feature and high-level concepts. CBVR is a database perspective method depending on the understanding of the content of multimedia documents and of their components.

Several researches (*Photobook*, *VisualSEEK*) and commercial (*QBIC*, *Virage*) systems provide automatic indexing and querying based on visual features such as color and texture. While low-level visual content can be extracted automatically, extracting semantic video features automatically such as event is still difficult, and it is usually domain (e.g. sports, dance) dependent [9], [10].

3. Concept, Principle, and Approach 3.1 Feasibility Analysis of Automatic Animation

Computer animation concerns 2D and 3D techniques. **2D animation** includes *sprite-based* animation and *image morphing*. The former is referred to embedding a *sprite* (graphic object) into an existing digital image or removing a sprite from an exiting image. The latter is referred to morphing from one source image to destination one. Whereas 2D techniques tend to concentrate on image manipulation, 3D techniques typically create virtual world filled with modeled objects to be animated.

Professionals conduct **3D computer animation** production through modeling, animation and rendering. Stages of modeling and animation require user's detailed expert knowledge on animation making and are quite time-consuming. 3D computer animation or motion generation techniques include *kinematical* (*geometric*), *procedural* (rule-based), *behavioral*, and *motion capture* approaches (table 1). The former three approaches are classified to *model-based* approaches among of which *behavioral approach* heavily relies on the techniques of AI and is built on other motion generation techniques such as *physically-based simulation* and *inverse kinematics*, so that it is a feasible way to realize automatic animation by behavioral approach at task level. Table 2 shows the AI sub-field to which behavioral approach belongs and others main applications of related AI technologies.

Table 1. 3D Computer Animation Techniques

Stages	Related Technologies	Use AI?
1.Modeling	***Surface modeling** **Polygonal surface **Curved and patched surface ***Solid modeling** ***Particle system modeling** for fire, cloud, mist, spray smoke, and so on	
2.Animation/ Motion generation	***Modeling-based** **Geometric (Kinematical)/ Keyframe-based -Hierarchical animation --Forward kinematics --Inverse kinematics -Shape deformation **Procedural/rule-based -Particle system -Physically-based simulation --Passive system --Active system ****Behavioral** ------------------ ***Motion capture-based** **Magnetics-based **Optical-based	Yes
3.Rendering	***Volume rendering** ***2D image manipulation** and so on	

Table 2. AI approaches Utilized in Animation.

Aspects of Intelligence	Related Technologies
Thinking humanly	Cognitive modeling *World knowledge representation
Thinking rationally	
Acting humanly	**Behavioral approach** *Domain knowledge representation
Acting rationally	Agent Multi-agents

3.2 Feasibility Analysis of Automatic VR

From the data analysis perspective, video surrogates can be classed under the headings *raw video features* (e.g. file size), *physical features* (spatial-temporal distribution of pixels: e.g. color) and *semantic features* (high-level concept: e.g. object). These visual contents are grouped in hierarchical layers as showed in Fig. 2. Query like "find red ball moving from left of the frame to right" relates to primitive level of video contents (color, texture, shape, motion); query like "a plane taking off" relates to high-level content (named types of action), query like "an video depicting suffering" relates to higher abstract level (emotion). Building semantics from raw video data becomes the main problem of content-based video retrieval.

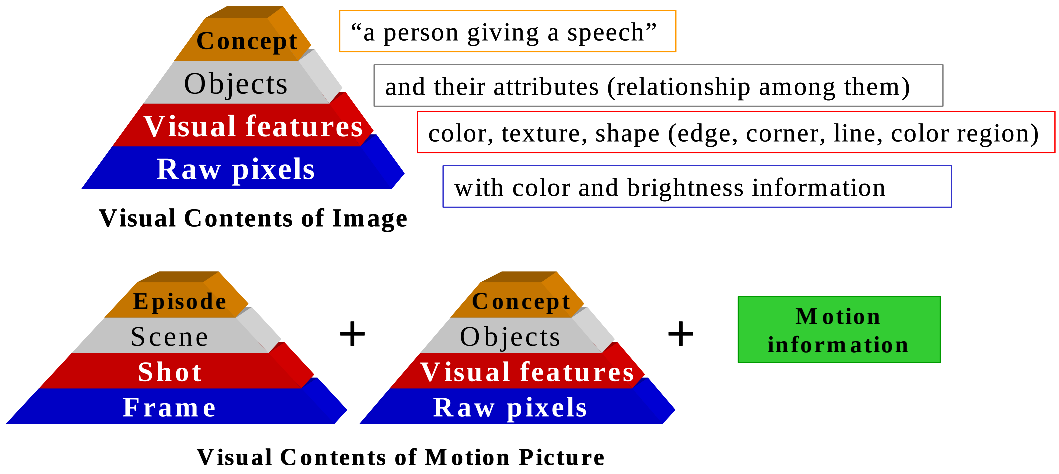


Figure 2. Visual Contents of Image/Video

Raw video naturally has a hierarchy of units from base level of individual frames to higher levels of *segments* such as *shots*, *scenes*, and *episodes*. *Shot* is defined as the single uninterrupted operation of the camera that results in a continuous action. A film/video is made up of shots arranged in sequence. A *scene* contains a group of shots that depict an event in the story and occur in one place. Concept *event* is an important primitive action unit in camera planning procedure such as "a private conversation between two characters" (two-talk). A series of related scenes form an *episode*. An important task in analyzing video content is to detect segment boundaries.

In our considerations, whereas the user has inherent information need expressed in semantics of query, or high-level concepts, the system operates according to the low-level features. That is to say either the user has to make the semantic-content translation or has to find a suitable video clip (or keyframe) to represent the query. We proposed annotation-metadata-based and content-based retrieval system EMMVR (Electrical MovieMaker Video Retrieval) for reusing video clips and computer-generated animation extracted from video repository.

4. EMM System Design

4.1 Screenplay GUI

When designing the form of screenplay, the first feature we consider is common user access. The other important issue concerns the possibility of automatic generation of visual effects made of various media (e.g. animation, video) and modalities (e.g. music, talk) decided by synthetic techniques involving the fields of Linguistics, Artificial Intelligence, Computation, and Computer Vision so that the screenplay design is also built on current technologies in these areas. We chose screenplay as input because it is a formal language for filmmaking that implies the lots of rules of film almost invisible by audience.

EventSP (Event ScreenPlay)
Symbols are used to retain events and ideas in our memory. One kind of screenplay we proposed is called EventSP that can describe abstract relationship between objects (such as *two talk* event). An event description should at least contain information like time/when, place/where, people/who, prop/which as well (Tab. 3). These information need to be input by user. Adopting filmmaking techniques from film theory, we utilize intelligent rule-based approach to formalize the event into a sequence of shots. The filmmaking techniques involve the following contents: 1. *mise-en-scène* (what to shoot, i.e. setting, lighting, figures, 2. *cinematograph* (how to shoot, i.e. camerawork – 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).

Table 3. Foundational Elements in EventSP

Terms	Example	Note
Time	Daytime	When
Place	Sea park	Where
Character	A: girl, B: boy,	Who
Prop	Trees	What
Event	Two talk	What happened

MarkupSP (Markup ScreenPlay)
The mind's picture is a combine of the perceptual elements of color, form, depth and movement along with the verbal thoughts. To be able to describe their imagery concretely, users 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 the above mentioned four aspects of filmmaking techniques (See example showed in Fig. 3).

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

Figure 3. Example MarkupSP Input

Film creates a five-dimensional world of touch, hearing, sight, smell and taste in the two-dimensional screen of sight and sound modes composed of different modalities. For the modality such as smell or taste its realization of sound motion picture may be expressed by speaking words (‘rotted apple’) or image (rotted apple). 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.

4.2 EMM System Architecture

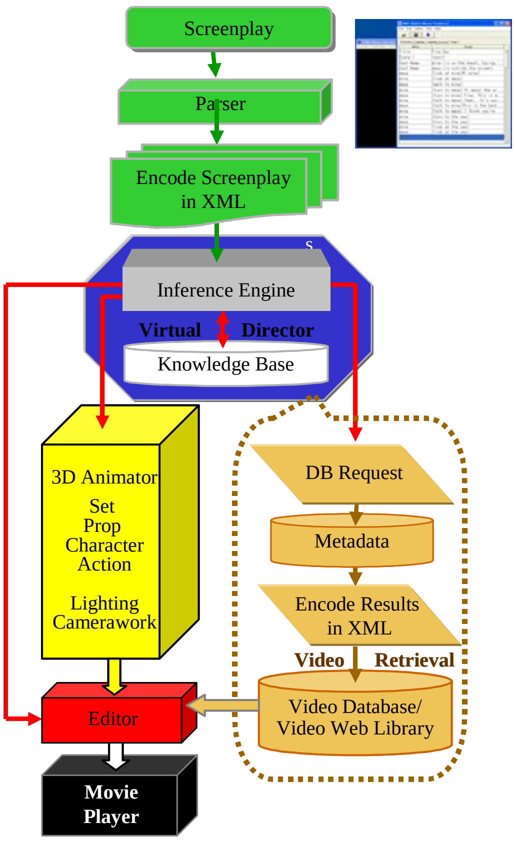


Figure 4. System Architecture Diagram

Virtual Director in Fig. 4 is embedded 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. 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 [11-13]. It has seven *Degrees of Freedom (DOFs)* – three for Cartesian position, three for orientation, and field of view (FOV).Cinematography comprises camera angles, mobile framing and camera movements. Various definitions of shot are based on camera manipulation. We defined shot as the single uninterrupted operation of the camera that results in a continuous action. *Shot* such as *full shot*, *pan*, or *track* is the smallest unit of dramatic action in the movie. All of shots are grouped into three categories derived from the four elements of movement – subject, camera angular, lens, and camera position (Table 4). A complex shot and developing shot are showed in Fig. 5 and Fig. 6 respectively.

Table 4. Categories of Shots

Categories of shots	**Elements of Movement** [ ] = may have, may have not
simple	*subject* (character, prop)
complex	*camera angular* (panning, tilt, rolling), *lens*, [subject]
developing	*Camera position*, i.e., mounting (tracking, dolly, crane), [lens], [camera angular], [subject]

Figure 5. Lens (zoom in) + Pan + Tilt

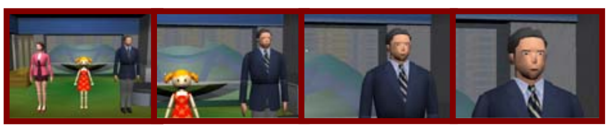


Figure 6. Track Shot – Tracking One’s Walk

A shot *sequence* is a group of shots depicting one action or which seems to belong with or depend upon each other. It is generated step by step to expound how to use cinematic 'rules of thumb' to make a scene. An example shot sequence is showed in Fig. 7.

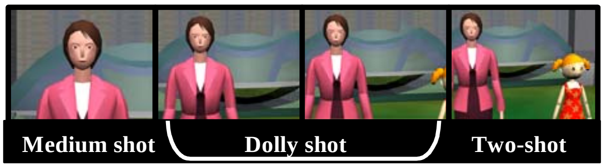


Figure 7. A Shot Sequence

EMMVR Giving a description of EMMVR (EMM Virtual Retrieval, see figure 8) in one sentence, it is a subsystem of EMM with a suitable multi-category video modeling and multi-modal query mechanism with multi-modal video indexing from film director's viewpoint. EMMVR focuses on design multi-modal video (and animation) indexing.

Video partitioning can operate at four levels of granularity: video-level, scene-level, shot-level frame level: A *scene* is a set of contiguous shots having a common semantic significance. The partitioning of the video into shots uses the temporal information, but generally does not refer to any semantic analysis. Types of shot boundaries like *cut*, *wipe*, and *dissolve* can be recognized. At frame level, there is little or no temporal analysis.

Feature extraction is distinguished into generic feature extraction and description feature recognition supervised by heuristics or training. These operations rely on the analysis of the Human Visual System (HVS), ranging from simple statistics to elaborated model-based filtering techniques. The event information can be extracted directly from audio-visual features (coming from visual contents, sound, integral and external text) in some domains by knowledge-based approach.

When building filmmaking semantics, ontology in this domain is necessary. **Ontology** can be seen as a conceptual map where the links between individual pieces of filmmaking knowledge are delineated. In the ontology tree of figure 9, those dark squares indicate the main contents should be extracted from video in order to reuse the video for digital moviemaking.

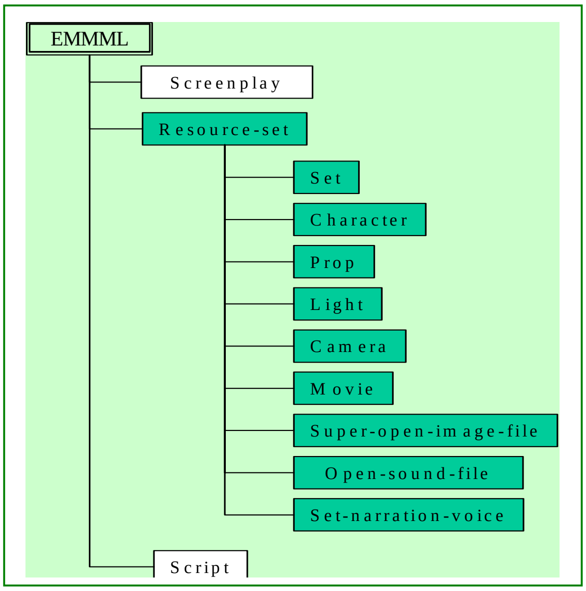


Figure 9. Visual Contents Tree in EMM System

Under the type of <Resource-set>, those subtypes of background <Set>, character <Character>, object <Prop>, lighting information <Light>, camerawork <Camera>, moving picture <Movie>, static image <Super-open-image-file>, music and effects <Open-sound-file>, dialogue or talk <Set-narration-voice> will be utilized to describe the features and contents of videos in library.

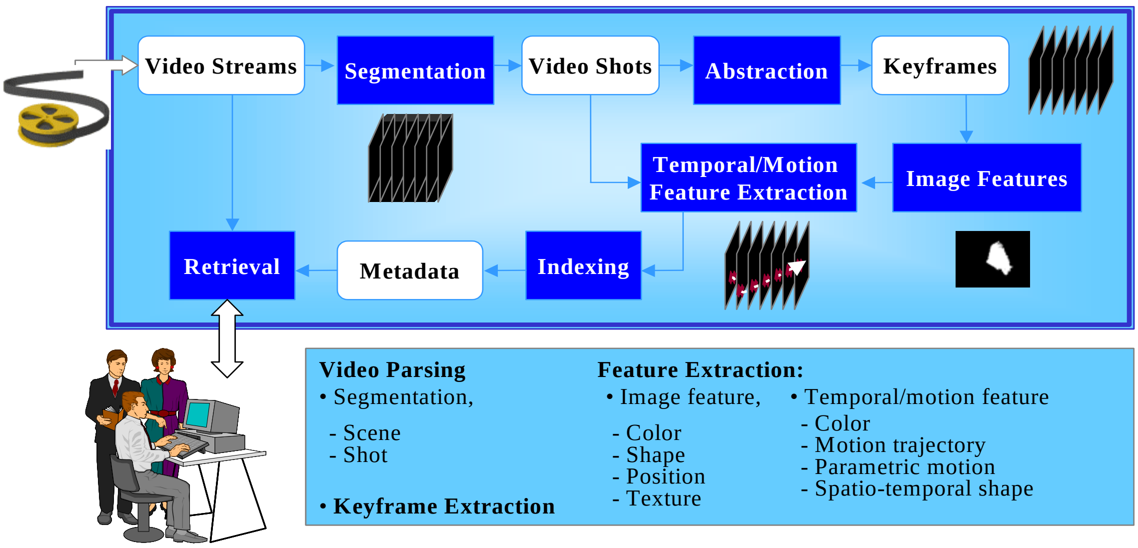


Figure 8. EMMVR (EMM Virtual Retrieval) Architecture

MPEG-7 standard has been used to encode video data because it is mainly intended for content identification purposes while other coding formats such as MPEG-2, 4 are mainly intended for content reproduction purposes. For MPEG-7 (DSs, Ds, DDL based on XML) standardizes the information exchange of descriptive information, we use its low-level and high-level descriptive metadata for video data modeling and retrieval. But only MPEG-7 is not completely suitable enough to serve as a multimedia data model, for its aim was not taking into different purposes. XML tags are supported by our EMMML (XML of EMM). An example EMM XML description of character element and its features like age hair clothes is showed in Fig 10.


```
<RESOURCE-SET>
 <Character name="Father" cid="F011/123456789ABD"
         type="urn:u-tokyo:dmp:cs:v0.5:Object:3DModel"
         href="http://foo.tv/Father.jar">
  <Feature type="Format" value="TVML Character" />
  <Feature type="Type" value="Human" />
  <Feature type="Gender" value="Male" />
  <Feature type="Age" value="Middle" />
  <Feature type="Voice:Style" value="Deep" />
  <Feature type="Voice:Language" value="English" />
  <Feature type="Hair:Style" value="Casual" />
  <Feature type="Hair:Color" value="Black" />
  <Feature type="Hair:Length" value="Short" />
  <Feature type="Skin:Color" value="Yellow" />
  <Feature type="Eye:Color" value="Brown" />
  <Feature type="Glasses:Style" value="Two Point" />
  <Feature type="Clothes:Shirts:Style" value="Open Neck" />
  <Feature type="Clothes:Shirts:Sleeve" value="Short" />
  <Feature type="Clothes:Shirts:Color" value="Striped Blue" />
  <Feature type="Clothes:Trouser:Style" value="Jeans" />
  <Feature type="Clothes:Trouser:Color" value="Blue" />
  <Feature type="Clothes:Trouser:Length" value="Long" />
  <Feature type="Action" value="Walk" />
  <Feature type="Action" value="Talk" />
 </Character>
</RESOURCE-SET>
```


Figure 10. Contents Description in XML Language

4.4 Visualization of Two-talk Event

Heuristics about making a sequence of shots involves the techniques of montage and sound related to image, and unit event. In EMM shot sequence of *two-talk* is decided by 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: *full 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.

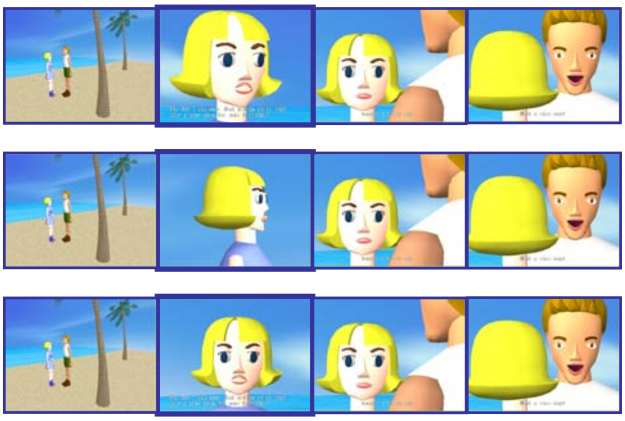
(e) If character talks, OTS (Over the Shoulder) short should be selected first.

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

To stage face to face two-talk, the virtual director determines five basic shots from nine camera positions (two positions for profile, angular, OTS, and point-of-view shots respectively, and one position for two-shot), selects shots from the set and arranges them in order dependent on dialogues. For the excerpt of EventSP in Figure 11, results like in figure 12 may be generated by the above rules (a)-(f).

Place: sea 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 11. The EventSp Input of Two-talk



1. Two-shot by rule (b), firstly decided
   Angular by rule (c) (f), <u>thirdly decided</u>
   OTS (facing A) by rule (c) (e), <u>secondly decided</u>
   OTS (facing B) by rule (d) (e). fourthly decided
2. Two-shot, Profile, OTS (facing A), OTS (facing B)
3. Two-shot, POV, OTS (facing A), OTS (facing B)

Figure 12. Example Scenes of Two Talk

From this example we can see:

- For the same event, virtual director can “image” many scenes. That means EMM can produce plentiful and interesting effects.
- The virtual director or user can control numerous cameras (in this example: 9). The camera that

follows director/user's control can be set more than one and at different positions and varying speeds.

- Some shots in set may be used not only once (e.g. OTS shot) while some may be never used (e.g. profile and POV shots do not appear in Fig. 12. 1).

We complete the practical works of virtual world authoring by TVML, 3D animation programming by C++, and rule-based reasoning by CLIPS. Only personal computer already meets the needs.

5. Conclusion

The rule-based engine in our system EMM can select the contents of presentation from video database and decide the temporal order of video clips (of various spatial-temporal media: video, animation, film, sound, speech, music), or create motion picture of animation, where the rule-based module is embedded as a subsystem in the integrated system environment to realize the automation. A huge selection of 3D models may by many human, human-like and animal.

Cinematic knowledge base extracted from the craftsmen gives great aid for the process of creating high quality motion pictures, be used to automate the digital filmmaking procedure through knowledge-based approach, or help for film theory teaching with easy-to-use user interface. GUI EventSP and MarkupSP are compensative each other. Using screenplay as input, it is feasible for software system to understand it well since by today's technique, it is still impossible to understand full natural language yet.

In our opinion the creator of computer animation will eventually be developed up to this module: the virtual high-leveled commands like a film director would use will direct synthetic actors, lights, cameras. For motion synthesis, we will take advantage of fundamental principles of tradition animation such as anticipation, appeal, arcs, exaggeration, follow through and overlapping action, secondary action, slow in/out, squash and stretch, staging, timing, and so forth.

References

1. Jinhong SHEN, Seiya MIYAZAKI, Terumasa AOKI, Hiroshi YASUDA, "Intelligent Computer Moviemaker with the Capabilities of Cinematic Rule-based Reasoning" (II), The Journal of the Institute of Image Information and Television Engineers (ITE), Vol. 7, pp.974-981, Tokyo, Japan, July 2004

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

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, pp. 163-168, Germany, Sept. 2002

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, pp.1-8, June 2002

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, pp. 487-494 Adelaide, Dec 10-14, 2001

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, pp. 120-133, Amsterdam, 2000

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

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

12. 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, pp. 217-224, New Orleans, Louisiana, United States, August 1996

13. 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