Automatic Animation Composition System for Event Representation

— Towards an Automatic Story Animation Generation —

Yusuke Manabe[1], Takanori Ohsugi[2], and Kenji Sugawara[1]

[1] Chiba Institute of Technology
{ymanabe,suga}@net.it-chiba.ac.jp
[2] Kyorindo Co. Ltd.

Abstract. This paper proposes an automatic animation composition system based on six databases, which are defined through analyzing a Japanese folktale animation movie. The system can automatically translate a text-based event representation (simple case frame) into an animation script (TVML). We show that our proposed system can compose several TVML scripts, which can represent animations of events as parts of a story.

1 Introduction

The concept of narrative includes several aspects such as story, discourse and representation. This study focuses on story and its visual representation [1–6], especially animation movie. Our final goal is to develop a system that can automatically generate an animation from a text-based story representation such as case frame [7]. In order to convert a text-based story into a visual image such as an animation, it is necessary to build many and various databases or knowledge bases for bridging between a text-based story representation and a visual representation including image resources, cinematic techniques, actions of characters, screen layouts and so on. However no such knowledge base or database exists. Actually some contributions, which try to generate animation movies automatically, face the same drawback that inhibits a realization of automatic animation generation [1–4, 6].

Generally we can illustrate the content of text-based story as visual image through complementing visual scene, character appearance and time background even if the sufficient visual information about a story is not given. This implies that we make up for the lack of visual information through knowledge of many stories that we have already read. From this viewpoint, conventional animation movies include informative animation resources, such as screen layouts, cinematic techniques, actions etc. If such resources can be used as knowledge for generating a new animation, we can automatically generate various animations by a computer program and knowledge bases (databases).

Therefore, in this study we define six databases through analyzing a Japanese folktale animation and developed an automatic animation composition system

T. Yoshida et al. (Eds.): AMT 2013, LNCS 8210, pp. 314–323, 2013.

for event representations. Our developed system can convert a text-based event representation into an animation script, TVML[1].

2 Analysis of a Japanese Folktale Animation and Database Definition

In order to make up for the lack of information for bridging a text-based and a visual-based story representation, we analyze a Japanese folktale animation and define some databases for composing animation.

We extract many image resources from the original Japanese folktale animation and make a TVML script, which realizes as same animation with the original animation as possible. The TVML script can be regarded as a template of animation composition. These elements are classified and stored in each database to be used as knowledge bridging a text-based and a visual-based story representation.

2.1 Animation Material and Analysis Steps

Animation material is a DVD animation, "Fleas are medicine (Nomi Wa Kusuri)", in a series of *Japanese Folktales For Reading to Children* [9]. The reason why we choose this title is that it is composed by minimal animation techniques; the number of characters is two, all of actions are expressed by simple 2D operations such as switching, rotation, transition and scaling of images.

Analysis steps are shown as follows;

1. To make a list of composition elements for all of scenes of the folktale animation.
2. To capture image resources such as characters and backgrounds from the animation.
3. To make a TVML script using captured image resources.

Step 1: Making a List of Composition Elements. We made an analysis list with respect to 14 elements, which are shown as Figure 1. We extracted 49 cuts from whole of a movie (time is 5m45s).

Step 2: Capturing Image Resources. We captured some scene images from all of cuts and clipped many character images and background images. The number of character images is 96, the number of background images is 37. The former are used as templates of character and action representation, the latter are used as templates of location and layout representation. In actual operational phase, all of image resources are replaced by other resources which are free from copyrights, although they are used for making a TVML script in the next step.

[1] TVML stands for TV program Making Language and it is a text-based language developed by NHK Science and Technical Research Laboratories [8]. TVML can be interpreted as a 2D/3D CG animation (OpenGL) in real time by TVML Player.

Table 1. An Example of the Analysis List (Translated in English)

Cut #	1	2	5
Time (start-end)	0:00 - 0:15	0:15 - 0:20	0:29 - 0:34
Narrations, Scripts		とんと昔、ある村に	ある年の夏、仕事で街へ出かけ、
Camera Movements	Move(Top --> Bottom)	Fix	Move(Bottom --> Top)
Characters	Man	Man	Man
# of Characters	1	1	1
Postures of Characters	[Man] left side view, whole body, Kimono, lean tree, get lying down, cross arms claspe hands behind head, cross legs (right leg over left leg)	[Man] left side view, whole body, Kimono, lean tree, get lying down, cross arms claspe hands behind head, cross legs (right leg over left leg)	[Man] back view, whole body, kimono, put a woven hat on, carry something to right shoulder
Activities/Transitions of Characters		[Man] Activity(Nod, Shake right leg)	[Man] Transition(Bottom --> Top)
Objects	[Man] chew a grass	[Man] chew a grass, loop a rope around left foot	
Object Movements		[a grass] rotate reef	
Semantics of Scenes	Show the title	peaceful noon?	[Man] man goes to village
Backgrounds	Sky, Cloud, Sun, Tree, Mountain, Hill?	Sky, Cloud, Tree, Road(Hill)	Tree, Cloud, Mountain,, Sky, Road, Homes
Screen Layouts	[Sun]TopCenter [Cloud]TopLeft, MiddleLeft [Hill]BottomCenter,BottomRight.MiddleRight [Tree]TopRight, MiddleRight [Mountain]BottomCenter [Man]MiddleRight	[Cloud]TopLeft [Tree]TopRight,MiddleRight [Road(Hill)]Bottom [Sky]TopCenter,MiddleLeft,MiddleCenter	[Road]MiddleCenter, BottomCenter [Homes]:MiddleCenter, BottomCenter [Mountain]MiddleLeft, MiddleCenter, MiddleRight [Cloud]TopRight, TopLeft [Tree]BottomCenter [Man]BottomCenter --> MiddleCenter
BGM/Effects	BGM: Music	Nothing	Effect: Cicadas are shrilling

Step 3: Making a TVML Script. Using the clipped image resources, we made a TVML script, which realizes as same animation with the original animation as possible. That is, this is a reconstruction of the original animation by TVML. We employ `prop` commands in TVML for image composition as the following list;

`prop: position()` it can be used for positioning,
`prop: visible()` it can be used for image switching,
`prop: transform()` it can be used for rotation, transition and scaling.

The total number of line of the TVML script we made is 1841 excepting comments and blank lines. This reconstruction does not include narrations, speech conversations and sounds but include actions (character's performances), screen layouts and cinematic techniques.

Here is an example of a TVML script, whose content is a cut that two characters talk. Figure 1 shows six image resources are assigned by `prop: assign()` because each character is configured by the three image resources. Figure 2 shows each character is properly configured by `prop: position()`. Figure 3 shows a 'talk' action can be expressed by `prop: transform()` and `prop: visible()`.

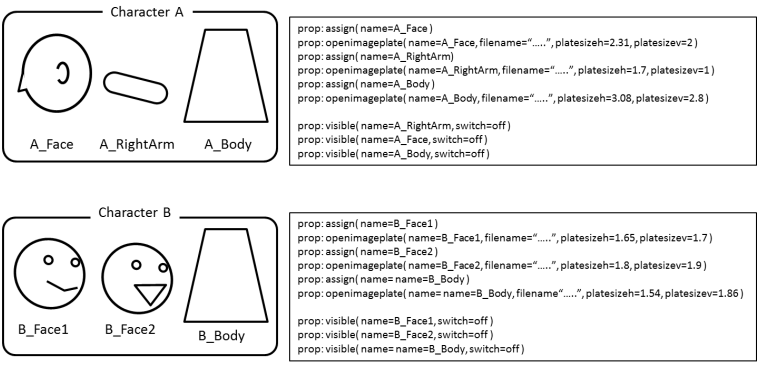

Fig. 1. Assigning Character Image Resources

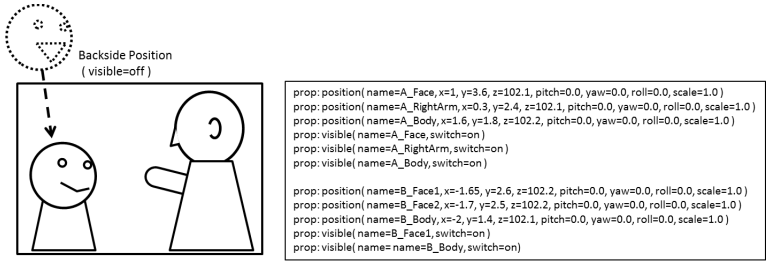

Fig. 2. Positioning Two Characters

As the result, some animations (i.e., fade-in/fade-out, concurrent actions, synchronization between a camera and a character) could not be realized due to the limitations of TVML and TVML Player although we could reconstruct most of the original animation.

2.2 Database Definition

After the above analysis steps, we classified the TVML script, the extracted image resources and the analysis list. As the result, they are stored in six databases; character image, location image, layout, action, transition and activity. In this subsection, we describe about main columns in each database.

Character Image DB. This database consists of 16 columns. `RPosX, RPosY, RPosZ` are relative coordinate values for composing a character by some image resources. `RotX, RotY, RotZ` are rotation degrees with respect to each axis. `Posture, Direction1&2, Embodiment` are properties of character image. `ActivityID` includes several IDs of activity database and it denotes that a relevant character image is used for expressing some activities tied with ID.

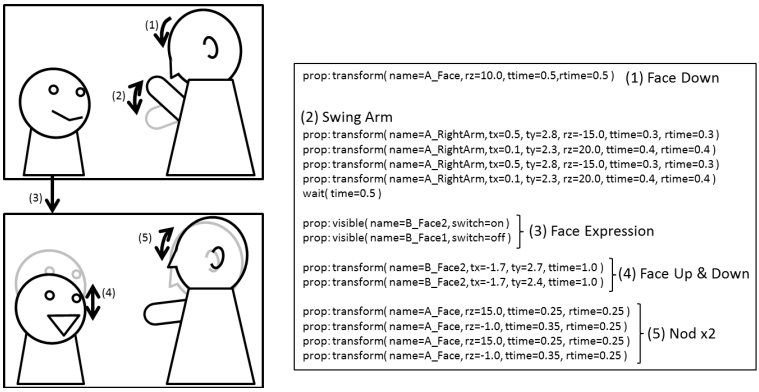

Fig. 3. Action Script

Location Image DB. This database consists of 11 columns. `initCamTVML` is a TVML script to capture properly a background image by a camera. `setLocTVML` is a TVML script to position a location image. `Posture, Direction1&2, Embodiment` are constraint properties for positioning character image. `LayoutID` includes several IDs of layout database and it denotes desirable positioning for character image. `TransitionID` includes several IDs of transition database and it denotes that a relevant location image allows to express some transitions tied with ID.

Layout DB. This database consists of 4 columns and it stores positioning information of character image.

Action DB. This database consists of 5 columns. The most important column is `TransitionID`, which is time sequence of activities and actions. Figure 4 shows an example of composition column in an action, 'come'. This figure shows 'come' action is composed of 2 transitions (MOVE1 and MOVE2) and 3 activities (ACT2CACT3CACT4). The meanings of another descriptor are shown as Table 2.

Transition DB. This database consists of 3 columns. The most important column is `setTransTVML`, which is TVML templates for transition of character image.

Activity DB. This database consists of 7 columns. `CharaID` is a character image ID for composing the activity. Thus an activity should be composed of four character images at a maximum. `setActTVML` is TVML templates for activity of character images.

X@MOVE1 #1,X@ACT2,X@MOVE2 #2,X@ACT3
X@MOVE1 #3,X@ACT2,X@MOVE2 #4,X@ACT3
X@MOVE1 #5,X@ACT2,X@MOVE2 #6,X@ACT4

Fig. 4. An Example of Composition Column in 'come'

Table 2. Descriptors in Composition Column

Symbols	Interpretation
X[Y]	Character Identifier (X and Y in case # of Actor is 2.)
ACT(Number)	ID in Activity Database
MOVE(Number)	ID in Transition Database
(Number) after #	Transition Times
@C#	Segmentation Symbols

3 Automatic Animation Composition System

3.1 Processing Steps

Based on the previous results, we developed an automatic animation composition system, which can generate a TVML script based on text-based event representation. In this study we define an event as a part of a story. That is, a story is composed by a time sequence of many event representations.

Our proposed system is composed of 3 processes; input process, decision process and generation process.

Input process is to set a text-based event representation through GUI operation. The event representation is composed of four elements; Agent(`AGT`), Counter-Agent(`C-AGT`), Location(`LOC`) and Action(`ACT`). Namely our proposed system postulates that a text-based event representation is not natural language but a formal data like simple case frame.

Decision process is to assess whether animation composition can be realized by the databases or not. The system examines the following six conditions.

1. Whether the location DB has location images corresponding to a text-based input.
2. Whether the character DB has character images corresponding to a text-based input.
3. Whether the action DB has action representations corresponding to a text-based input.
4. Whether the action representation refers to the transition DB, the activity DB or the both.
5. Whether the location image DB has location images to allow some transition representations because an action representation depends on location images when it includes the transition representations.
6. Whether the character image DB has character images to allow some activity representations because an action representation depends on character images when it includes the activity representations.

When it is decided that an animation composition can be realized, this system can automatically generate a TVML script.

Generation process is to determine all of image resources that the system needs and to calculate each positions of resources, on-off control of image display, image operations such as transition, rotation and scaling. Figure 5 shows the workflow in generation process.

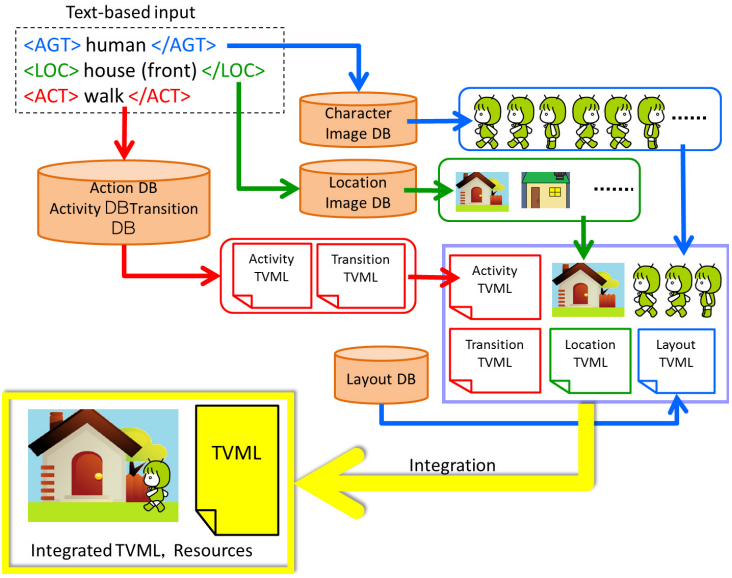

Fig. 5. Workflow in Generation Process

Mainly the system executes the following processing.

1. To get candidates of character images and location images based on a text-based input.
2. To get candidates of transition TVML and activity TVML for composing action representations.
3. To determine transition TVML and activity TVML from their candidates.
4. To determine location images from its candidates and get a location TVML.
5. To determine character images from its candidates.
6. To get a layout TVML for locating character images.
7. To integrate all of TVML and image resources.

Finally a composed TVML script is interpreted as a CG animation by the TVML Player.

3.2 Implementation

We developed the system by C# language under Microsoft Visual Studio 2010. Figure 6 shows a screen shot of the developed system. The menu items in the left

side are used for input of an event. Setting an event and pressing a generation button, a TVML script is automatically written in the text area. The output button is used for output of a TVML script, which can automatically interpreted as an animation by TVML player.

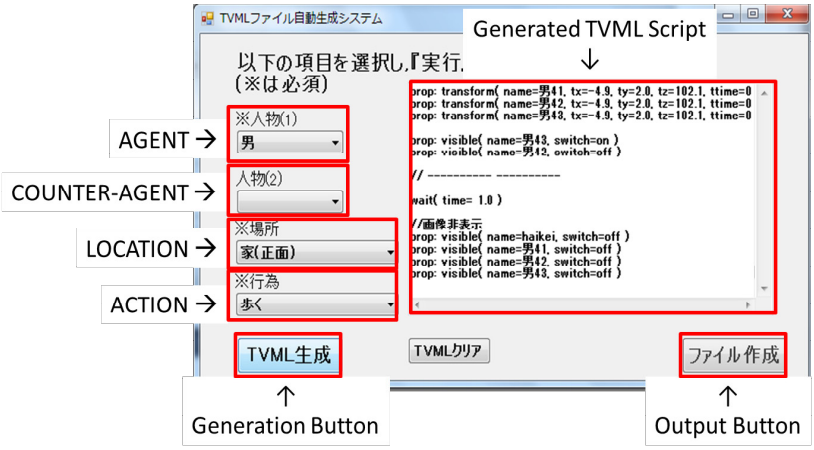

Fig. 6. Screen Shot of Developed System

To design six databases, we employ a database management system, MySQL. The number of data samples is 70 rows in the Character DB, 30 rows in the Location DB, 59 rows in the Layout DB, 9 rows in the Action DB, 4 rows in the Transition DB and 16 rows in the Activity DB.

4 Feasibility of Automatic Animation Movie Generation by Proposed System and Databases

In order to discuss a feasibility of automatic animation generation, we automatically composed some TVML scripts by our proposed system and databases. Then we evaluated animations based on the automatic composed TVML scripts.

In this paper, we show results of 3 animations based on 3 events as follows;

1. <AGT>man</AGT> <C-AGT>old woman</C-AGT> <LOC>room</LOC> <ACT>talk</ACT>
2. <AGT>man</AGT> <LOC>house(front)</LOC> <ACT>pass</ACT>
3. <AGT>man</AGT> <LOC>house(front)</LOC> <ACT>come</ACT>

The first event is equivalent to an event in the original animation movie but the image resources are changed as simple cartoon pictures. The second event and the third event do not exist in the original animation movie. The 'pass' action (from right side to left side by a walk) and the 'come' action (from left

side by a walk) are newly defined based on 'come' action (from right side by a walk) in the original movie. The former corresponds to a distance change and the latter corresponds to a direction change. Image resources are changed as simple cartoon pictures or clip-art pictures[10, 11].

Figure 7, 8 and 9 show some screen shots of the animations based on the generated TVML scripts. Although it is hard to judge each animation by static screen shots, we could confirm that each animation has a sufficient quality of an event representation.

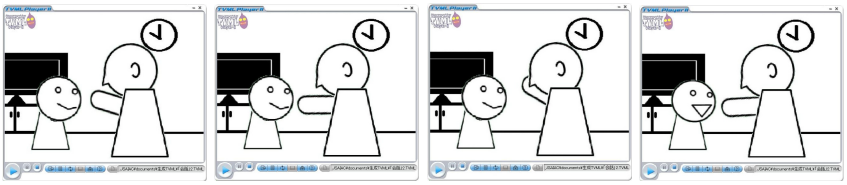

Fig. 7. Generated Animation by Simple Cartoon Pictures ('talk')

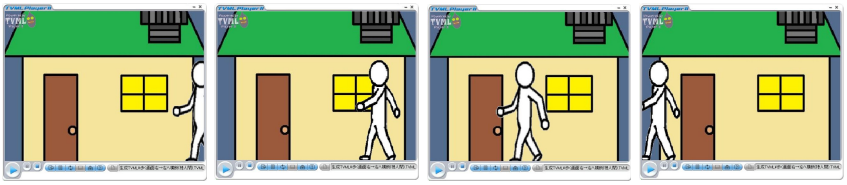

Fig. 8. Generated Animation by Simple Cartoon Pictures ('pass')

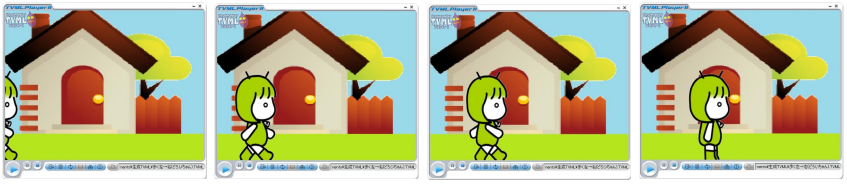

Fig. 9. Generated Animation by Clip-art Pictures ('come')

5 Conclusion

In this paper we analyzed a Japanese folktale animation and defined six databases for composing animation in order to make up for the lack of information for bridging a text-based and a visual-based story representation. Based on the six databases, we developed an automatic animation composition system, which can automatically translate text-based event representation (simple case frame) into animation script (TVML). Finally we found that our proposed system can compose several TVML scripts, which can represent an event animation as parts of a story.

As future works, we would like to deal with the following items;

1. Making a uniform approach to prepare image resources.
2. Making a camera database for controlling cinematic techniques.
3. Refining each database, especially a design of a hierarchical action database.

References

1. Pérez, R., Sosa, R., Lemaitre, C.: A computer model for visual-daydreaming. In: AAAI Fall Symposium on Intelligent Narrative Technologies, pp. 102–109. AAAI Press, Washington D.C. (2007)
2. Miyazaki, K., Nagai, Y., Nakatsu, R.: Architecture of an Authoring System to Support Interactive Contents Creation for Games/E-Learning. In: Pan, Z., Aylett, R.S., Diener, H., Jin, X., Göbel, S., Li, L. (eds.) Edutainment 2006. LNCS, vol. 3942, pp. 70–79. Springer, Heidelberg (2006)
3. Miyazaki, K., Nagai, Y., Wama, T., Nakatsu, R.: Concept and construction of an interactive folktale system. In: Ma, L., Rauterberg, M., Nakatsu, R. (eds.) ICEC 2007. LNCS, vol. 4740, pp. 162–170. Springer, Heidelberg (2007)
4. Hoshina, K., Noguchi, T., Sugimoto, T., Enokidzu, H.: A trial of story comprehension simulation: Through automatic generation of animation from story text. In: Proc. of the 29th Annual Meeting of the Japanese Cognitive Science Society, vol. P1–7, pp. 204–210 (2012) (in Japanese)
5. whiteonwhite: algorithmicnoir, `http://www.rufuscorporation.com/wowpr.htm`
6. Lu, R., Zhang, S.: Automatic Generation of Computer Animation. LNCS (LNAI), vol. 2160. Springer, Heidelberg (2002)
7. Manabe, Y.: How to Automatically Generate Animation Movie of Story — Towards Action Representation Ontology—. In: Proc. of Literature and Cognition/Computer II, vol. 15G-6, 10 pages (2008) (in Japanese)
8. TV program Making Language. NHK Science and Technical Research Laboratories (2013), `http://www.nhk.or.jp/strl/tvml/index.html`
9. Fleas are medicine (Nomi Wa Kusuri). In: DVD Japanese Folktales For Reading to Children (Yomikikase Nihon Mukashi Banashi), vol. 1. Columbia Music Entertainment, Inc. (2006)
10. doroid-chan.net (Doroid Chan Matome Site), `http://droid-chan.net/` (in Japanese)
11. materialand, `http://www.material-land.com/` (in Japanese)