Text-To-Vision Player

- Converting Text to Vision Based on TVML Technology -

Masaki HAYASHI

Internet Research Institute, Inc.

E-mail: hayasim@iri.co.jp

ABSTRACT

We have been studying the next generation of video creation solution based on TVML (TV program Making Language) technology. TVML is a well-known scripting language for computer animation and a TVML Player interprets the script to create video content using real-time 3DCG and synthesized voices. TVML has a long history proposed back in 1996 by NHK, however, the only available Player has been the one made by NHK for years.

We have developed a new TVML Player from scratch and named it T2V (Text-To-Vision) Player. Due to the development from scratch, the code is compact, light and fast, and extendable and portable. Moreover, the new T2V Player performs not only a playback of TVML script but also a Text-To-Vision conversion from input written in XML format or just a mere plane text to videos by using 'Text-filter' that can be added as a plug-in of the Player.

We plan to make it public as freeware from early 2009 in order to stimulate User-Generated-Content and a various kinds of services running on the Internet and media industry. We think that our T2V Player would be a key technology for upcoming new movement..

Keywords: computer graphics, animation, computer language

1. Introduction

In the recent years, the media industry has matured and different fields like internet-based communication, television broadcasting, and the age-old field of printed books have increasingly overlapped. In particular, the rise of video sharing programs like YouTube has been very remarkable. Nowadays, medias that were focused on text and still images so far have also started handling video content. It has been observed in the case of various medias that the distribution of content through a media first gains popularity among professionals. Then, as the media becomes mature, gradually semi-professionals and finally amateur users start producing and distributing content in that media. This type of content where the average user produces and distributes the content is called UGC (User-Generated-Content), which is also a proof of the media entering the maturity stage. Therefore, it can be easily expected that the handling of video content will become a common practice in the field of UGC.

However, at present, not enough support is provided to enable the average user to produce video content. The problem of video quality in video production has already been resolved because of the availability of high resolution cameras for home use in the market. However, when such video programs are considered, it is still not easy to produce videos with sufficient content acceptable to everyone. It is because of this reason that the most frequently accessed videos on YouTube etc. are actually copied from television programs or movies. On the other hand, in the field of website design, which is around for quite some time now, and where the users created websites using HTML authoring tools so far, a new tool, blog, has appeared, which has completely changed the way of transmitting and disclosing personal information on the Web. In the field of video production as well, the advent of a revolutionary tool corresponding to the blog is awaited.

2. From TVML to T2V

For more than 10 years, we have been proposing and researching TVML (TV program Making Language)[1] as a technology, using which an average user can easily produce video content. TVML was proposed by NHK (Japan Broadcasting Corporation) Science and Technical Research Laboratories in 1996. It is a special text-based language for describing video content. A software called TVML player reads the video script written in this language and instantly generates videos by using technologies such as real time CG and speech synthesis. TVML is not a programming language like C; instead, it is

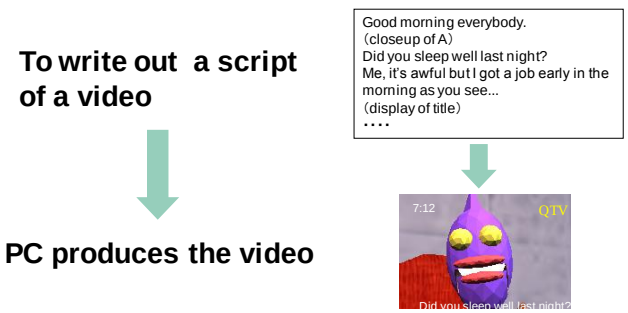

Fig. 1: Text-To-Vision (T2V) Conversion

a type of scripting language. One of its salient features is that the user need not be very highly skilled for using it. However, it is still not very easy for an average user to easily write in it, as it needs some computer knowledge. Therefore, as a next phase, NHK proposed a technology for video production that can be easily used even by a user who is only familiar with the usage of a word processor. This technology is called APE (Automatic Program Production Engine)[2]. APE is used to automatically convert a word processor script written by the user into TVML, which is then played back on a TVML player to produce a video. From the user point of view, this becomes an easy-to-use video production environment as it enables the user to produce a video by simply writing the word processor script.

Thus, since it became possible to "produce video from a word processor script," the basic foundation of our research (the concept called "Text-To-Vision conversion" and technical R&D for realizing it) also became very clear. At present, we are advancing our R&D program based on this concept of Text-To-Vision (Fig. 1), which is very simple yet powerful and has immense potential.

Although T2V was originally designed to enable the general user to easily produce a video, its potential is much more than merely the usage for UGC. Today, the amount of textual information distributed over the internet is enormous, and a lot of textual information is continuously transmitted in the form of e-mails, blogs, and homepages. Using our T2V technology, it is possible to automatically convert such text information into a video[3]. This technology is called "Automatic Video Production." Although the potential of this technology is still unknown, we believe it is very high. In addition, original TVML technology not only converts text into video in real time, it can also be applied to real time communication such as chat, 3D virtual world, and exchanging the text in real time.

If we summarize the direction of research and development described above, it can be broadly divided into three areas: (1) UGC, (2) Automatic video production, and (3) Interactive content. At present, we are progressing with our research in these three directions, and we are trying to realize a complete new world of future content by linking it to a service that can be actually used by the society.

3. Importance of TVML

Although we described the broad direction of our research in Chapter 2, for this research to be of real benefit to the future society, it is needless to say that the underlying technology to be used in this research has to be robust and mature. In our approach, since the basic technology is TVML, the key success factor would be sufficiently raising the perfection level of TVML and developing it as a highly extensible technology. However, although TVML has been highly appreciated in Japan, it is still in the research phase and yet to fully mature.

Here, problems associated with TVML are roughly divided into problems with the scripting language and problems with the player application that plays back this script. Ideally speaking, keeping aside the problems associated with the player, it is possible to expand the specifications of the scripting language without any limitations. However, it is extremely difficult to completely describe a video using the language, and although we are not professionals of video production we can easily imagine that any attempts to do so will result in infinite complications of the language itself. This is one of the problems associated with the standardization of the technology. So far, there have been several cases where tremendous amount of effort was spent to complete the technical standards; however, they became too complex and could not be used practically.

Therefore, we believe that the correct path for making this technology viable is to practically resolve the problems by simultaneously updating the scripting language and the player. An excellent example of this can be found in the history of the development of HTML, which is used to script the web page and web browser, which is used to display HTML script. As everyone knows, the parent organization for the standardization of HTML is W3C. When W3C was established, several browsers were already available, and therefore, the necessity for standardization became the driving force behind it. Here the important point lies in a arriving at a scheme where the joint development of the language is taken up as a non-profit work and browsers for displaying it are developed according to the strategy of each company.

On the other hand, the field of TVML is very small. At present, only one TVML player is available for the playback of TVML, which was developed by NHK[4]. NHK has been stating that TVML is an open language and any one can freely develop the player based on its specifications. However, so far there has been no real example of such development.

To realize the immense potential of TVML, Internet Research Institute, Inc. (IRI) decided to develop a new TVML player independent of the NHK's player. By learning from the above-mentioned example of HTML, we have formed collaborative relationship with NHK for creating and developing the language specifications of TVML. We have also started reflecting the practical requirements of a TVML player shared between both the sides in the language specifications.

Hereafter, the features and prospects of the new TVML player developed by IRI are described.

4. T2V Player

We named the newly-developed player as T2V Player. After one year of development efforts, we have been able to complete the beta-version for Windows. Moreover, this player has been developed from scratch without using any part of the code of NHK's TVML player.
T2V Player's development policy, which was formulated

around one year ago, is as follows:

(1) It should be developed from scratch.
(2) It should be light and compact.
(3) Animation engine should be redesigned and equipped with advanced functionalities.
(4) It should be possible to generate movie film data.
(5) It should be possible to use it as a plug-in of a Web browser
(6) It should be gradually ported to MacOS, Linux, and other embedded operating systems
(7) Automatic video production functionality should be built in.

As of January 2008, we have been able to complete the development of a player that satisfies the above policy requirements from (1) to (4). Regarding (5), testing has been successful and it will be completed soon. Although porting to other platforms as per (6) is yet to be done, since the program code has been designed keeping portability in mind, we expect that the porting is not going to be very difficult. Regarding (7), the functionality itself has already been built-in and from now on, we plan to start the development of an automatic video production engine.

During the development of this player, the language specifications of TVML itself were revised to a certain extent based on our discussions with NHK, and updated language specifications (version 3.0) were prepared. Moreover, the format of CG character data that can be used by TVML was not clear in the earlier specifications. This was also clarified (BM format Version 3.0) and development was done based on that.

5. Technical features of T2V player

T2V player is comprised of (1) TVML core engine that reads TVML script and plays it back, (2) User Interface (U/I) having GUI as user application. This core engine and U/I are designed and developed independently. Furthermore, not only the playback of TVML script, but also that of word processor script mentioned earlier or of complete plain text is also possible by using the automatic video production module, APE. In particular, the concept of an arbitrary text filter known as FIL (Flexible Interpretation Loader) was introduced. By creating a suitable FIL and plugging it in to the player, users can use it very flexibly for converting the input text format designed by them into a video and for converting the HTML code of the website whose format is decided beforehand into a video.

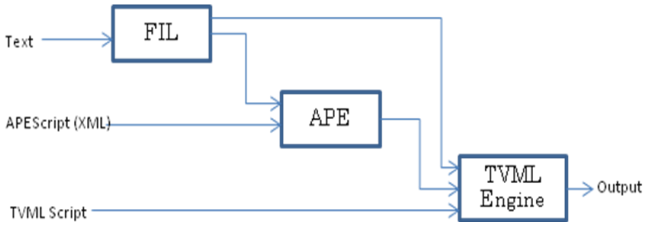

Fig. 2: APE, FIL and TVML

Fig. 3: T2V Player

- The open source Irrlicht Engine is used for the drawing engine because it offers excellent portability.
- It is designed to ensure compatibility with various speech synthesis software.
- It is designed for multilingual use, and it can be used for programs in languages other than Japanese and English.
- Coding is in native C++ and TVML command is implemented as the method of class.
- It is designed for simultaneous execution of TVML script and the abovementioned method. Therefore, during the playback of the script, real time operations are possible such as moving the camera randomly or moving the characters simultaneously.
- Core engine is provided as a dll. Therefore, it is possible to use the engine from any application in addition to the player U/I.
- Multiple core engines can be used simultaneously.

(2) U/I

- It has the movie file output function.
- It is possible to playback three types of scripts: TVML script, APE script, and plain text.
- User can also create the specifications of APE and FIL as they are open source. Plug-ins can be registered in the player as dlls.
- The user interface is very similar to that of any conventional video player. It also has a text editor screen where the user can edit and playback the script.
- It is possible to change the skin.

5. Conclusion and future perspectives

In this paper, we described the recent trends in media, introduced the TVML technology that we have been researching continuously, and discussed the importance of TVML. Then we introduced the indigenously developed T2V player and its salient technical features. Since T2V player has been developed independently of the TVML player used so far, we believe that it will act as a basic infrastructure for our media and content related activities.

In this paper, we briefly described the technical features of the newly developed player. Although we will separately present the applications and future prospects of this player, we would like to describe the broad vision of our research in the end.

As described in Chapter 2, our research is based on the concept of “Text-to-video conversion” in the following three directions: (1) UGC, (2) automatic video production, and (3) interactive content. Among them, although (3) interactive content is quite understandable, actual strength of (1) and (2) will be realized by the presence of T2V player function on the user side. If we look at the current state of services over the Internet, in majority of the cases, a complex function like T2V player resides on the server side and the user accesses this function by communicating with the server. We believe that this is a transitional method, although it has its own merits and demerits. At present, perhaps the most successful and popular Web browser add-on having a function similar to T2V player is Flash player. Though the feature enhancement of Flash player is quite remarkable, our final aim, which is based on the concept of “Text-To-Vision conversion,” is in a different direction altogether.

If T2V player is installed on the client side and connected to the network, instead of direct transmission of the signal data of a video, textual information of the script is transmitted (of course, everything is not converted into textual information). First of all, the physical merit of this is a drastic reduction in the traffic load as the volume of text data is much less compared to that of the video signal data. In addition to this, the field of distributing textual information and converting it into a video on the client side holds enormous potential. For example, when presenting textual information as a video, the users can customize it the way they like (for example, changing the newscaster according to the preference). Moreover, by processing semantic information, it is also possible to customize textual information itself (for example, multilingualization by automatic translation, changing the tone, summarizing), and it is possible to freely re-use the information (for example, the omnibus and parody versions can be easily produced).

Our final aim is the kind of world where semantic information is freely circulated and shared as text. It is often said about the transition of current media and content is that time-shift occurred as the first stage, relieving the user from timing constraints by using a hard disk and a large-scale server. Then, space-shift occurred as the second stage, relieving the user from location constraints due to the emergence of ubiquitous computing from mobile technology. Perhaps semantic-shift will occur as the third stage, where the user will be relieved from semantic restrictions. In the age of semantic-shift, technologies for generating, distributing, and processing semantic information will become very important. Natural language processing, knowledge processing, AI, etc. shall become the relevant technical fields. However, these technologies are insufficient in their current form as they do not necessarily meet the requirements of the development of the next generation media. Therefore, to bring forth the semantic-shift, it is necessary to have a new concept and drive the research forward with the aim of creating a new field.

Text-To-Vision technology introduced in this paper shall become one of the key technologies in the semantic-shift of media and content described earlier. Moreover, we think that the basic concept of “Text-To-Vision conversion technology” that we are advocating will gradually merge into a larger concept of “Semantics-To-Expressions conversion technology.” “Semantics" and “Expressions” are very fundamental concepts like “two wheels of a cart” in the life of a human being. Comprehensive R&D related to their mutual conversion and consolidating the results obtained as a technology is going to be very exciting.

REFERENCES

[1] Hayashi, Ueda, Kurihara, and Yasumura: "TVML (TV program Making Language) - Automatic TV Program Generation from Text-based Script -" Imagina99 proceedings (1999)

[2] Hayashi, Douke, and Hamaguchi: "Automatic TV Program Production with APEs" C5 The Second International Conference on Creating, Connecting and Collaborating through Computing Conference Proceedings pp.18-23 (2004.1).

[3] Douke, Hayashi, and Makino: "A Study of Automatic Program Production Using TVML" Short Papers and Demos, Eurographics '99, pp.42-45 (1999)

[4] http://www.nhk.or.jp/strl/tvml/index.html