Development of a TV System Augmented Outside the TV Screen

Hiroyuki Kawakita[*,**], **Michihiro Uehara**[*], **Toshio Nakagawa**[*], **Makoto Sato**[**]

*NHK, 1-10-11 Kinuta, Setagaya-ku, Tokyo 157-8510, Japan
**Tokyo Inst. of Tech., 4259 Nagatsuta-cho, Yokohama, Kanagawa 226-8503, Japan
Keywords: augmented reality, augmented TV, synchronization method, position & rotation estimation, 3DCG

ABSTRACT

We advocated "Augmented TV" that gives an appearance of connecting the onscreen world, captured by the TV camera, to the real world in front of the TV screen. To realize the goal, we have developed AR system for TV video images, which is characterized by smoothness of the connecting representation.

1. INTRODUCTION

Various advances in TV technology have been made, including full color, high definition, wide screens (with an aspect ratio of 16:9), and binocular stereoscopic 3DTV. One of the main objectives for these advances has been to improve the sense of realism so that the space captured by the TV camera seems to expand before one's very eyes. We have taken it one step further by proposing "Augmented TV", which gives the appearance of connecting the onscreen world, captured by the TV camera, to the real world in front of the TV screen [1,2]. The viewer can enjoy effects such as TV characters leaping out from the screen. To achieve this goal, we developed an AR (augmented reality) system for TV video images, a versatile technology which is characterized by the smoothness of the "leaping out" representation (see Fig.1). The TV screen is viewed with a camera built into a mobile device (smartphone etc.), on which 3DCG animation is superimposed at the calculated position on TV video images captured by the camera. The animation is synchronized with the TV video images, enabling characters to seem to leap out from the screen. This paper shows primarily the synchronization method and estimation method of TV position and rotation for the system as a summary of the studies to date.

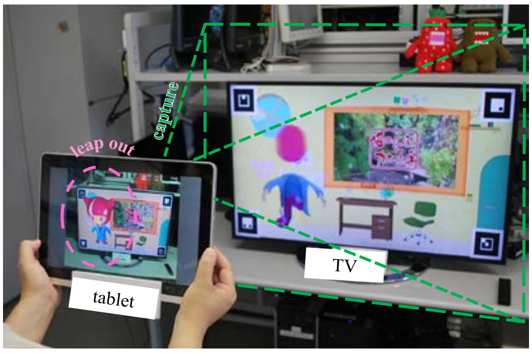

Fig. 1 Example of character leaping out

2. REQUIREMENTS

To realize the "leaping out" representation, the 3DCG is necessary to be superimposed at an appropriate time and position in the mobile device. Owing to this, the software running on the device needs always to know playback time of TV video images and TV position and rotation. We need to develop the following methods:

- synchronization method
- estimation method of TV position and rotation

2.1 Basic Requirements

To consider applicability to various display areas including digital signage as a TV, configuration of the system needs to be simple, and only mobile device needs to have processing function of both methods. We defined the following requirements:

- Both methods must be implemented only with TV video images and functions of the mobile device on the market. (The TV is not need to connect with a network)
- The software on the mobile device implementing both methods must perform real-time processing.

2.2 Synchronization Method Requirements

To implement the smoothness of the "leaping out" representation, 30fps-frame-accurate synchronization on the mobile device screen between the 3DCG animation and the TV video images captured by the mobile device camera and superimposed is required.

2.3 Estimation Method Requirements

To estimate TV position and rotation in a mobile device, some cues to detect the TV are needed. According to requirements in 2.2, viewer's environment such as furniture must not be available as the cue, therefore we displayed a cue on the TV. On the other hand, a highly flexible representation in which the character leaping out can go right around the viewer is desired. We defined the following requirements:

- The design of TV detection cues displayed on TV must be simple so as not to interfere with productions.
- Estimation must be possible even when a camera is not capturing a TV screen.

3. PROPOSED METHOD

3.1 Synchronization Method

This section shows our synchronization method between superimposition of 3DCG and display of TV video images captured by mobile device camera on the mobile device screen. TV video images are assumed to be played without jitter or delay of the frame. Not always during playback, but in advance, the synchronization processing can be executed.

In the synchronization method, as shown in Fig. 2, playback time of TV video images is shown as a figure on TV, and the figure is captured by the mobile device camera. To achieve highly accurate synchronization, we composed the figure of a 2D matrix as an integer part of the playback time and a moving marker, called "time marker", as a fraction part of the playback time (see Fig. 3). In the mobile device, a fraction part of playback time of

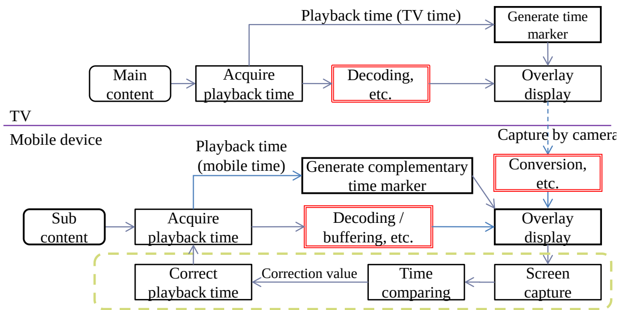

Fig. 2 Block diagram of the synchronization

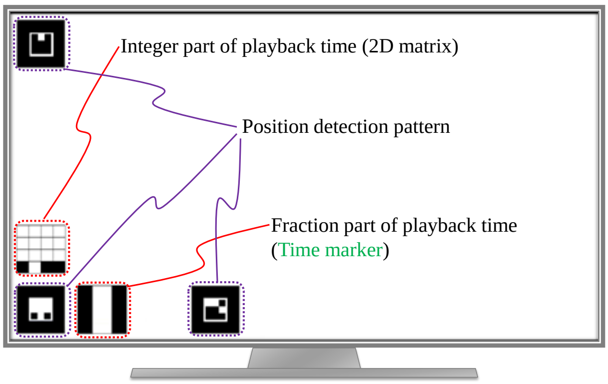

Fig. 3 Time marker on TV

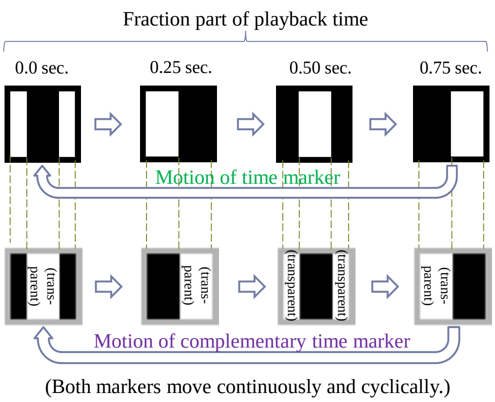

Fig. 4 Motion of (complementary) time marker

3DCG animation is also shown as a moving marker, called "complementary time marker", which is superimposed at position of time marker captured by the mobile device camera. The motions of both markers are shown in Fig. 4. The mobile device screen captured, and the playback time of 3DCG animation is corrected in accordance with the positional deviation between both markers on the mobile device screen.

3.2 Estimation Method of TV Position & Rotation

This section shows our estimation method of the TV position and rotation, which is implemented in a mobile device. Efficiently to estimate TV position and rotation, the method is characterized by using image processing by the mobile device camera and gyro sensors. Thereby, estimating by gyro sensors enables not only to estimate without capturing the TV but also to reduce area of image processing.

Fig. 5 and Table 1 show the overall flow of the proposed estimation method. This flow is executed more than each camera frame input from the mobile device camera. With this method, first the initial TV screen position and rotation necessary for estimation using gyro sensors are obtained. Then, using the gyro sensors' estimation as the origin, image processing with the camera is executed, and "TV frame" (see Fig. 6) as a detect cue on the TV is recognized (see Fig. 7). Thereafter, it repeats the processing of the gyro sensors' estimation and image processing. Only measuring the mobile device rotation by gyro sensors, the position cannot be estimated correctly when the position changed.

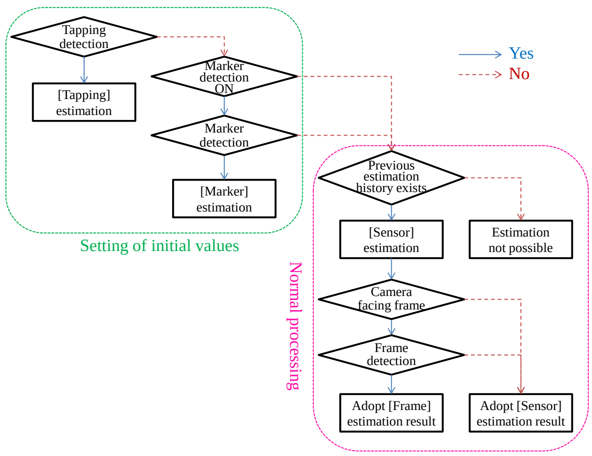

Fig. 5 Overall estimation flow

Table 1 Estimation method

[Tapping]	[Frame] estimation in which search starts in left, right, top and bottom directions from tapping position
[Marker]	Estimation by AR marker detection
[Sensor]	Estimation in which gyro sensor's rotation amount is reflected in the previous estimation result
[Frame]	Estimation by searching frame by image processing (Fig. 7)

Fig. 6 TV frame as a detect cue

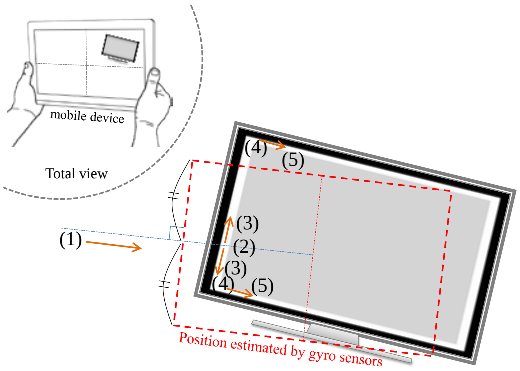

Fig. 7 TV frame search path

3.3 System Architecture

Fig. 8 shows the system architecture in the mobile device. To help produce the content, we incorporated an easy TV program production language called TVML (TV program making language), which NHK has developed [3,4]. A program script written in TVML is read by software called a TVML player to produce the program by real-time rendering. 3DCG character's basic motions can be designated by by action commands in TVML. Since the TVML player can be controlled by external software, interactive content can also be produced using user software. We introduce a TVML player as a presentation environment in the mobile decice, and implemented the synchronization method as shown by the red broken line.

4. EVALUATION

4.1 Synchronization Method

To verify the proposed synchronization method, we implemented and evaluated it with a tablet as a mobile device. Table 2 shows the parameters used in the evaluation. Fig. 9 shows the synchronous error time for the distance between the TV and the tablet. We measured the synchronous error time five times for the same distance, and extended the distance by 0.25 m. Most measurement values are within plus or minus 30 msec. We calculated the average of the absolute values of the measurement values for the same distance. The averages are about 10-20 msec. regardless of the distance. From these results, we confirmed the frame-accurate synchronization at 30 fps.

4.2 Estimation Method Processing Time

We calculated the processing time for each frame (using parameter in Table 2). Since the frame rate for the environment used this time was 30 fps, it was necessary to keep image processing for 1 frame to 33 msec. or less. With the proposed estimation method, the processing time is 1-2 msec., which is no problem whatsoever.

4.3 Hands-on Demo

To verify that the proposed methods are implemented with sufficient accuracy for general viewers to enjoy the content, we held a hands-on demo with our system (see Fig. 10). As shown in Fig. 11, in main part of the content, a giant squid leaps out from the TV screen, turning two

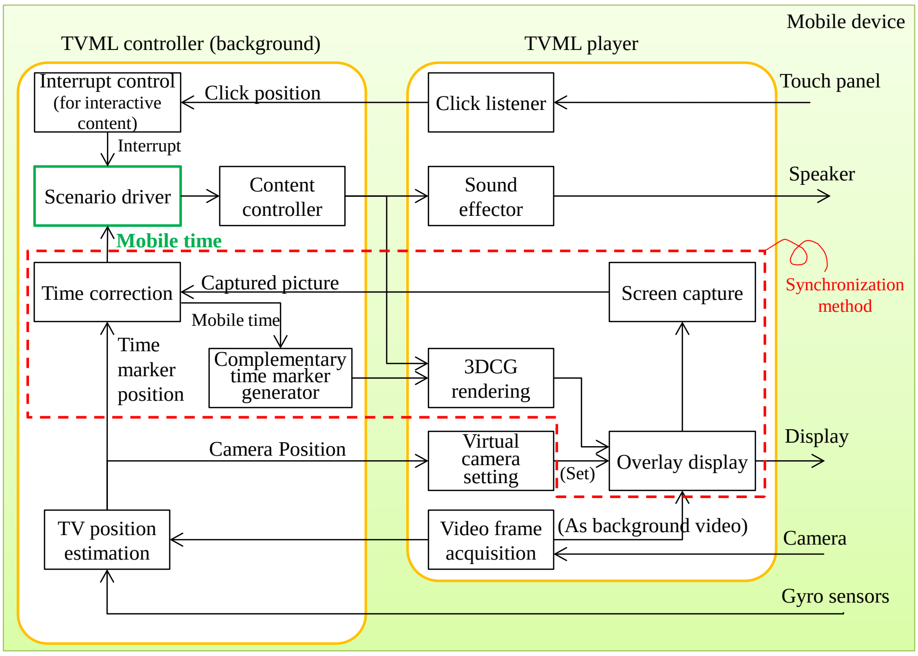

Fig. 8 System architecture

Table 2 Experimental environment

Item			Value
Tablet	OS		Windows8
	CPU		Intel Core i3 1.8GHz (2-core)
	System memory		4 GB
	Image sensor resolution		1920 x 1080 px
	Camera resolution setup		640 x 480 px
	Camera frame rate		30 fps
	Camera horizontal angle of view		50.4 degrees
	Gyro sensor		3-axes
	External dimensions		295 x 191 x 11 mm
	Weight		955 g
TV	Available screen area		886 x 498 mm (40”)
	B&W frame width	Top/bottom side	White:13 mm Black:12 mm
		Left/right side	White:18 mm Black:28 mm
TV – Tablet Distance			Approx. 130 cm
Compiler			Visual C++ 2012

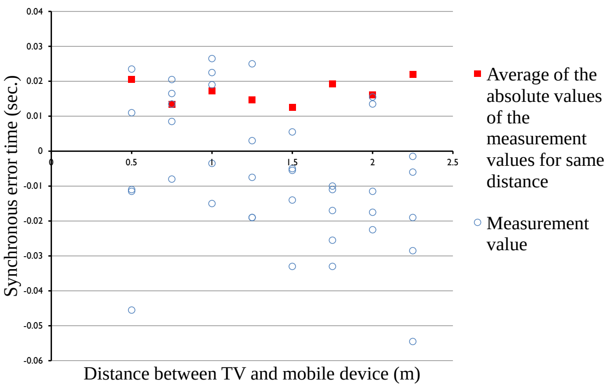

Fig. 9 Synchronous error time against distance

Fig. 10 Hands-on demo

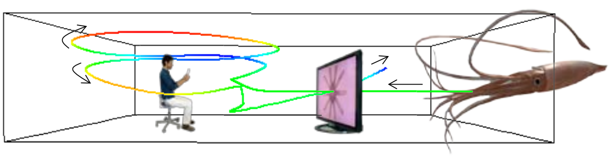

Fig. 11 Schematic diagram of the demo content

laps around the viewer, and go into the screen again. The squid turning, the viewers turned to follow it. About 150 families participated in the demo.

As a result, when the squid turned and came back in front of the TV, the mobile device could recognize the TV frame again in most cases. This result means successful transition from estimation only by gyro sensors to estimation by image processing and gyro sensors. Many of the reaction, especially by children, were “great!”, “amazing!”, and so on.

5. PROSPECT

In our present system, a viewer needs to grip a mobile device and turn the mobile device camera towards a TV screen to experience Augmented TV service. In the future, a small and high-performance head mounted display device like glasses with a camera will become popular, and the device will recognize a TV screen and superimpose related animation on a daily basis. As much as such device is wearable, unconscious views desired, the superimposed animation should be matched with main TV content. Namely, accuracy of synchronization among appearances derived from different display principles and estimation or recognition of position and rotation will become important.

6. SUMMARY

We advocated a media concept, named Augmented TV, which is able to give an appearance of connecting the onscreen world, captured by the TV camera, to the real world in front of the TV screen. We developed an AR system to implement Augmented TV with a mobile device on the market. We also developed the required methods for the system, which are synchronization method and estimation method of TV position and rotation. We confirm the effectiveness of the system with a demo content.

REFERENCES

[1] H. Kawakita, T. Nakagawa, and M. Sato, “Augmented TV: An Augmented Reality System for TV Programs Beyond the TV Screen,” International Conference on Multimedia Computing and Systems, pp. 955-960 (2014).

[2] H. Kawakita, T. Nakagawa, and M. Sato, “TV Augmented Outside the TV Screen Using Mobile Device,” Proc. ASIAGRAPH 2015, Vol.10, No.1 (2015).

[3] NHK TVML Player, http://www.nhk.or.jp/strl/tvml/english/player2/index.html

[4] M. Hayashi, H.Ueda, and T. Kurihara, “TVML (TV program Making Language) – Automatic TV Program Generation from Text-based Script-,” Proc Imagina’99 (1999).