Real-time sports video analysis for video content viewing with haptic information

Masaki Takahashi
Japan Broadcasting Corporation
Shibuya-ku, Tokyo, Japan
takahashi.m-iu@nhk.or.jp

Makiko Azuma
Japan Broadcasting Corporation
Shibuya-ku, Tokyo, Japan
azuma.m-@nhk.or.jp

Takuya Handa
Japan Broadcasting Corporation
Shibuya-ku, Tokyo, Japan
handa.t-@nhk.or.jp

Taichi Ishiwatari
Japan Broadcasting Corporation
Shibuya-ku, Tokyo, Japan
ishiwatari.t-@nhk.or.jp

Masanori Sano
Japan Broadcasting Corporation
Shibuya-ku, Tokyo, Japan
sano.m-@nhk.or.jp

Yuko Yamanouchi
Japan Broadcasting Corporation
Shibuya-ku, Tokyo, Japan
yamanouchi.y-@nhk.or.jp

CCS CONCEPTS

• **Computer methodologies** → **Computer vision problems**.

KEYWORDS

Image recognition, machine learning, haptic devices

ACM Reference Format:
Masaki Takahashi, Makiko Azuma, Takuya Handa, Taichi Ishiwatari, Masanori Sano, and Yuko Yamanouchi. 2021. Real-time sports video analysis for video content viewing with haptic information . In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters (SIGGRAPH '21 Posters), August 09-13, 2021.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3450618.3469135

1 INTRODUCTION

Video content mainly conveys visual and audio information, but it would be more immersive by adding haptic information. The sense of presence when watching video could be enhanced through tactile stimuli from haptic devices, enabling video content to provide immersive experiences.

We developed a haptic perception system that can be used at public sports viewing events. The live sports broadcast video is displayed on a large screen in front of audiences, and haptic stimuli that are synchronized with the video are supplied to the audiences in real time. We targeted Judo content because it is a physical sport, and haptic stimuli related to athletes' bodily motions would improve audiences' sense of immersiveness.

Figure 1 shows an overview of our system. First, broadcast video is analyzed and type and timing of haptic stimuli, which are related to athletes' motions are measured using the video-analysis module. This module is key to enabling public viewing with haptic devices because the devices have to be automatically actuated in real time to synchronize with live broadcast content. The haptic devices, which audiences have and sit in, are then automatically controlled via wireless WiFi. Video analysis requires processing time, so the original video is delayed by about one second, and the haptic stimuli and audio/visual information are synchronized.

We also developed two types of haptic devices, i.e., handheld and chair. The handheld device mainly conveys the intensity of athletes' motions, and the chair device conveys the timing and impact of a specific action event, such as throwing trick and ground fighting. These devices contribute to a viewer understanding the game situation and improving the sense of immersiveness during a public viewing event.

We evaluated the performance of the video-analysis module of our system, and the results indicate that it has sufficient accuracy of action recognition. We demonstrated our system at an exhibition event in June 2021, and plan to put it to practical use at a live sports public viewing event in July 2021. To the best of our knowledge, this live public viewing event will be the first attempt at using such a system.

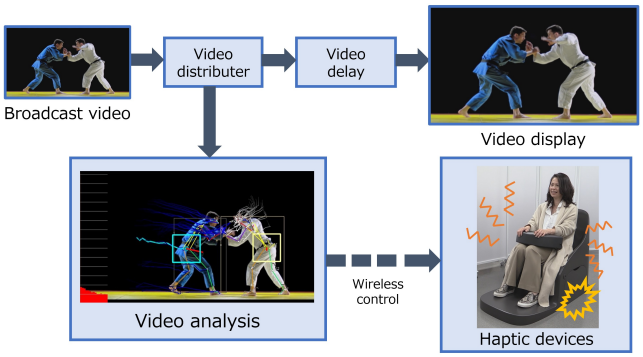

Figure 1: Overview of haptic perception system at public viewing event.

2 DEVELOPED SYSTEM

2.1 Video-analysis module

The video-analysis module of our system measures the type and timing of haptic stimuli from live Judo broadcast video in real time. Athletes' skeletons are inferred every frame from the video, and 30 points of athlete's joints are consequently tracked [Kristan et al. 2020]. The intensity values of two athletes' motions are then

SIGGRAPH '21 Posters, August 09-13, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8371-4/21/08.
https://doi.org/10.1145/3450618.3469135

calculated from the average magnitude of their optical flows and the distance of the two athletes. The handheld haptic device is continuously actuated in accordance with the intensity values.

Specific actions in a Judo match, such as a throwing trick and ground fighting, are also recognized with this module. Although a convolutional neural network (CNN) is commonly used for image recognition, it is difficult for a common two-dimensional CNN to recognize actions from a single image because of the lack of temporal information [Chen et al. 2020].

We developed an image representation called joint trajectory image (JTI) containing temporal information. The processing flow for creating a JTI is shown in Figure 2. First, the skeleton positions of two athletes are inferred, as shown in the middle image. Next, the skeleton positions are tracked in subsequent frames, and the trajectories of the athlete's 30 points of joints are created by connecting their optical flows. Temporal information is expressed by decreasing the brightness of past optical flows, as shown in the right image.

A CNN identifier is trained for classifying JTIs into four Judo action events, i.e., stand-up fighting, ground fighting, throwing trick, and others. The CNN identifier detects the action event from a JTI, and the chair haptic device is actuated at that time.

Figure 2: Creation of joint trajectory image.

2.2 EXPERIMENT

We evaluated the accuracy and speed of the video-analysis module of our system. The CNN identifier was trained from about 2,000 JTIs, which were created from videos of four different Judo matches at the Rio de Janeiro Olympic Games. The f-measure of action recognition [%] and its average processing speed [frames/second] were calculated over about 1,000 images from other Judo matches from the training set. We used three image representations, i.e., the original image, skeleton image, and skeleton motion history (SMH) image [Phyo et al. 2017], for comparison with a JTI . The results are shown in Figure 3.

The JTI was the most accurate. The other images tended to miss the specific action events because they do not contain temporal information. The total frame rate with our system, which includes the video-capture process, was faster than 15 frames/second, which is fast enough for use at live public viewing events.

2.3 Haptic devices

Figure 4 shows that two haptic devices for a public viewing event. The handheld device conveys the intensity of athletes' motions, and the chair device conveys the timing and impact of specific actions. Both devices are automatically controlled in accordance with the results of the video-analysis module via a wireless WiFi network. The shapes and materials of the devices are designed to satisfy both high transmissibility of haptic stimuli and comfortability during video content watching at public viewing events.

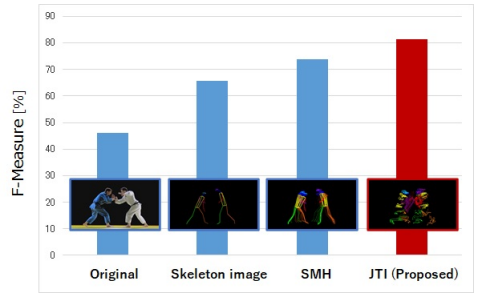

Figure 3: Average accuracy of Judo action detection.

Figure 4: Handheld haptic device (left) and chair haptic device (right).

3 PRACTICAL USE

We demonstrated our haptic perception system in an exhibition event in June 2021, and plan to put it to practical use at a live sports public viewing event in July 2021. We estimate more than 100 people will experience our system during the event.

External participants and researchers have already experienced the system. They enjoyed watching Judo video content with the haptic devices and said that the system was very impressive because it was the first time for them to have haptic experiences synchronized with video content. Many participants mentioned that the system helped improve the feeling of immersiveness when watching.

REFERENCES

H. Chen, Y. Wang, H. Shu, Y. Tang, C. Xu, B. Shi, C. Xu, Q. Tian, and C. Xu. 2020. Frequency Domain Compact 3D Convolutional Neural Networks. In *In Proc. of the International Conference on Computer Vision and Pattern Recognition (CVPR),*. 1641–1650.

M. Kristan, A. Leonardis, and J. Matas et al. 2020. The Eighth Visual Object Tracking VOT2020 Challenge Results. In *Proc of the European Conference on Computer Vision Workshops (ECCV Workshops),*. 547–601.

C. N. Phyo, T. T. Zin, and P. Tin. 2017. Skeleton Motion History based Human Action Recognition Using Deep Learning. In *In Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).*