Event Detection in Broadcasting Video for Halfpipe Sports

Hao-Kai Wen, Wei-Che Chang, Chia-Hu Chang, Yin-Tzu Lin, and Ja-Ling Wu
National Taiwan University
{r01922083, r02922031}@csie.ntu.edu.tw,
{chchang, known, wjl}@cmlab.csie.ntu.edu.tw

Figure 1: Flowchart of the proposed system.

ABSTRACT

In this work, a low-cost and efficient system is proposed to automatically analyze the halfpipe (HP) sports videos. In addition to the court color ratio information, we find the player region by using salient object detection mechanisms to face the challenge of motion blurred scenes in HP videos. Besides, a novel and efficient method for detecting the spin event is proposed on the basis of native motion vectors existing in a compressed video. Experimental results show that the proposed system is effective in recognizing the hard-to-be-detected spin events in HP videos.

Categories and Subject Descriptors
H.5.1 [Information Storage and Retrieval]: Content Analysis and Indexing—Indexing methods

Keywords
halfpipe; vert ramp; sports video; action recognition

1. INTRODUCTION

A halfpipe (HP) is one of the most popular terrains in gravity sports like snowboarding, skateboarding, skiing, freestyle BMX, and inline skating. For players of HP sports, a common way to improve their skills is to learn by imitation. In order to learn the best tricks in the world, they imitate tricks performed by elite players through watching recorded broadcast videos, such as Sochi 2014 Olympics and KIA World Extreme Games. As shown in the top-left part of Figure 1, a general HP contest can be hierarchically divided into rounds and runs. In each run, all contestants will take their turns to perform tricks. For each player’s turn, she or he starts from one side of the HP, i.e., the vert, passes through the flat, reaches to another vert, and goes back and forth several times, as shown in Figure 2. Each time the player flies above the vert, she or he performs one type of tricks, such as spin or grab. We defined a trick segment as a period of time during which the player passes through the flat, reaches to one of the verts, and then goes back to the flat. Finding a particular trick in a video is time-consuming and labor-intensive because certain tricks may be sparse in the video and the timestamps of them are unknown to a video watcher. As a result, finding trick segments and recognizing trick types automatically are highly wanted by and beneficial to HP players.

Recognizing the trick type in HP videos is challenging because (i) there are often more than one shots taken at different locations by shoulder-mount camcorders in a single trick segment, and (ii) the scenes in an HP video are motion blurred since players move fast. As a result, it is difficult to estimate global motion by using either pixel-based method (e.g., optical flow) or feature-based method (e.g., SIFT matching). And thus, segmenting a player’s silhouette by background subtraction techniques becomes impractical.

1https://en.wikipedia.org/wiki/Half-pipe
Most of the existing studies [1] recognized the actions of players with the help of wearable sensors. However, it is hard to apply sensor-based methods to all contestants because of privacy issues. Therefore, in the proposed system, HP contest videos are automatically analyzed with low-cost and efficient methods without sensor information. The flowchart of the proposed system is illustrated in Figure 1. We will address how HP sports videos become readable with the aid of video segmentation, player detection and tracking, and event detection mechanisms, presented in the following sections. Please also refer to our demo video\textsuperscript{2} for real illustrations.

2. VIDEO SEGMENTATION

By using the scoreboard information, we segment an input HP broadcasting video into run-level segments. By observation, we found that the percentage of coverage of HP court color in the scene is relatively high when the player is on the flat. So the video could be segmented into trick segments by finding peaks in the ratio curve of the court color coverage along the time line. However, the peaks may happen several times in between the two verts because of shot transitions, which affect the distribution of the peaks in the curve of color coverage ratio. To improve the precision of the segmentation, we apply a moving average filter, of about half a trick long, on the curve of color coverage ratio.

3. PLAYER DETECTION AND TRACKING

When a player moves fast on an HP, the background is blurred but the player is relatively clear because the camera moves along with the player. So we perform multi-scale contrast saliency detection\textsuperscript{2} to find player region candidates. Besides, the scenes in these frames are often composed of a player surrounded by the HP court, i.e., the scenes are of high coverage ratio of the court color. As a result, the connected components in the non-court color parts of the scene can be also chosen as player region candidates. Finally, the player region can be detected by combining the salient object region and the non-court color region in the frames with high court color coverage ratio. For other scenes, we apply the algorithm proposed in [3] to track the player. Since several shot changes may occur in one trick segment, we also re-estimate the player region at each detected shot change.

4. EVENT DETECTION

To recognize the event in each trick segment, i.e., the type of tricks, we detect the moment when the player reaches the highest point because the player’s moves near to the highest point are representative of performed tricks. Once the player flies above the HP, she or he can be treated as a projectile. And the scene will stop changing when the highest point is reached. Meanwhile, the coverage ratio of court color will remain low for a few frames. As a result, we detected the highest points in trick segments by finding the valleys of the color coverage ratio used in Section 2.

Then, we intend to recognize if there is a spin – one major type of HP tricks. The movements of the native motion vectors, of a compressed video, will behave like a circle, as shown in Figure 3, when a player performs a spin trick. As a result, the following mechanism is suggested to recognize such an event. Let \( A \in \mathbb{C}^{M \times N} \) be the matrix representing the motion field of the player region, where \( \mathbb{C} \) denotes the complex field. We accumulate the vertical components of \( A \) along the horizontal axis, \( u_j = \sum_{i=1}^{M} \text{Im}(A_{i,j}) \), and likewise the horizontal components along the vertical axis, i.e., \( v_i = \sum_{j=1}^{N} \text{Re}(A_{i,j}) \). As shown in Figure 3, vectors \( u \) and \( v \) will cross the zero-valued axes, once they reach the center of the circle. However, motion vectors may not form a perfect circle because the rotating movements behind the body of the player cannot be captured by the camera when the rotation axis of the player is nearly parallel to the camera view. In such a situation, some part of the circle may be missed, but fortunately, either \( u \) or \( v \) could still capture the pre-described property. So, we could recognize a spin event by counting the number of zero crossing points in \( u \) and \( v \) of the detected highest point, and its neighboring frames. The more frames with only one zero crossing point in \( u \) or \( v \), the more likely that there is a spin in a trick.

5. RESULTS

To evaluate, we used the video of “KIA World Extreme Games 2013–AIL Vert” from Youtube (in H.264 format). Eito Yasutoko’s 2\textsuperscript{nd} run and Nikolaj Najdenov’s 1\textsuperscript{st} run were chosen as testing video sequences. Each run in the contest is of 45 seconds long. In Eito Yasutoko’s 2\textsuperscript{nd} run, the result of the highest points detection is 9 hits, 0 false alarm, and 1 miss. There is a miss because the player falls in that run. For spin detection, there are 4 hits, 1 false alarm, and 0 miss. In Nikolaj Najdenov’s 1\textsuperscript{st}, the result of the highest points detection are 9 hits, 0 false alarm, and 2 misses. For spin detection, there are 3 hits, 0 false alarm, and 2 misses. The miss detections are due to the relatively small detected player region in the scene.

6. REFERENCES


\textsuperscript{2}http://www.cmlab.csie.ntu.edu.tw/~r01922083/halfpipeEvent/