SPORTS VIDEO CATEGORIZING METHOD USING CAMERA MOTION PARAMETERS

Shin’ichi Takagi, Shinobu Hattori, Kazumasa Yokoyama, Akihisa Kodate and Hideyoshi Tominaga

GITI, Waseda University
1-3-10 Nishi-Waseda Sinjuku-ku, Tokyo 169-0051, Japan
{takagi,sally,yokoyama,akikod,tominaga}@tom.comm.waseda.ac.jp

ABSTRACT

In this paper, we propose a content based video categorizing method focusing broadcasted sports videos using camera motion parameters. We define two new features in the proposed method; "Camera motion extraction ratio" and "Camera motion transition". Camera motion parameters in the video sequence contain very significant information for categorization of broadcasted sports video, because in most of sports video, camera motions are closely related to the actions taken in the sports, which are mostly based on a certain rule depending on types of sports. Based on the characteristics, we design a sports video categorization algorithm for identifying 6 major different sports types. In our algorithm, the features automatically extracted from videos are analyzed in a statistical manner. The experimental results show a clear tendency the applicability of the proposed method for sports genre identification.

1. INTRODUCTION

There have been many research efforts that strive for analyzing structures of video sequence, for indexing information[1], for making summaries [2], for detecting highlight scene [3,4] and so on. But, a method for extracting and representing a overall feature of a video sequence has not been discussed enough. In this paper, low level metadata which can be extracted from a video sequence directly such as color, shape, motion information can be analyzed as temporal features, our main purpose is to extract and represent an overall feature of a video sequence by analyzing low level metadata. As first step, we discuss a method for categorizing broadcasted sports video by analyzing camera motion parameters statistically.

In most cases, sports video are shot by limited numbers of mostly fixed cameras tracing objects and edited by the directors choosing the best scene shot by those cameras. Therefore, shots taken by one fixed camera show a certain similarity in images which is closely related to the camera's position in the stadium. One of good examples is the shots taken from behind the goals which is closely related to the camera's position in the stadium.

But, the important features for these characters, color information and structure information of the typical images are easily influenced by the factors such as the place (stadium) and time (day or night) of each game. For example, the parameter sets for U.K Open tennis (Wimbledon) on green grass court and the French Open tennis on red clay court must be very different.

On the other hand, there is a certain relationship between camera motion parameters, and types of sports. Through our subjective investigations, camera motion parameters in the broadcasted sports video have some typical patterns which depend on type of sports and these patterns are clearly different from videos other videos such as dramas and other variety programs.

Camera motion parameters can be classified into some status such as FIX, PAN, ZOOM and so on. And the status is changing in time. So we introduce a new feature defining the status transition of camera motion parameters as "Camera motion transition". Camera motion parameters can be extracted from video sequence by analyzing motion information. As the computational cost of camera motion extraction is not very high because of recent development of PC performance. So camera motion parameter has many advantages for categorization of broadcasted sports video.

This paper consists of 5 parts. In the following part, after giving the definition of two new features for analyzing video sequence, we discuss statistic analysis of camera motion parameters. Then the 4th part discusses statistical analysis method for broadcasted sports video and the experimental results for sports genre identification. Although we use only Camera motion extraction ratio in the first analysis, more detailed analysis can be performed by using Camera motion transition. The final part gives the conclusions.

2. CAMERA MOTION PARAMETER EXTRACTION METHOD

In this paper, we use MPEG formatted video sequences as experimental data, because of its fast camera motion parameters extraction capability, because coded motion vector information can be clue. We adopt our camera motion extraction method, which extracts 4 types of camera motion parameters (FIX, PAN, ZOOM, SHAKE) for each GoP (Group Of Pictures) of MPEG video[8].

3. CAMERA MOTION EXTRACTION RATIO AND CAMERA MOTION TRANSITION

In this section, we define two new features; "Camera motion extraction ratio" and "Camera motion transition", both are based on camera motion parameters automatically extracted from MPEG video sequence.
3.1. Camera Motion Extraction Ratio

We define "Camera Motion Extraction Ratio". There is a certain characteristic in camera motion parameters of their appearance probability in the video sequence. Here “Camera Motion Extraction Ratio” \( W \) is defined as

\[
W_x = \frac{\text{Num}_{\text{appear}}}{\text{Num}_{\text{total}}} \times 100(\%)
\]

\( x \in \{ \text{FIX, PAN, ZOOM, SHAKE} \} \)

where \( \text{Num}_{\text{total}} \) is total numbers of GoP included in the video sequence and \( \text{Num}_{\text{appear}} \) is the number of times of an appearance for camera work \( x \). For example, when the camera motion parameter "FIX" appears 10 times in the video sequence which includes 100 GoP, Camera motion extraction ratio for \( W_{\text{FIX}} \) can be computed as 10%. So, we can get 4 camera motion extraction ratios for each video sequence.

3.2. Camera Motion Transition

Camera motion transition is a new concept which represents a state transition of camera motion parameters. In our definition, camera motion parameters are changing the state among 4 types (FIX, PAN, ZOOM, SHAKE) along with time scale in video sequence as shown in Figure 1. For example, camera motion parameter "FIX" appears 10 times in the video sequence and described as a FIX-PAN camera motion transition. As we have 4 types of parameters, total 12 transitions can be defined. Arrows represented in Figure 1 are the camera transition. As we have 4 types of parameters, total 12 transitions of camera motion parameters are changing the state among 4 types (\( s \)). In our definition, Camera motion transition is a new concept which represents a transition appearance probability. Here “Transition ratio” \( T_{ab} \) is defined as

\[
T_{ab} = \frac{\text{Num}_{ab}}{\text{Num}_{\text{total-transition}}} \times 100 \text{ (%)}
\]

where \( \text{Num}_{ab} \) represents appearance times of camera motion transitions between camera motion parameter \( A \) and \( B \), and \( \text{Num}_{\text{total-transition}} \) represents total appearance times of camera motion transitions in the video sequence. For example, when the transitions happen 10 times between the parameter \( A \) and \( B \), and all of transition happened in the video sequence is 100 times, the transition ratio for \( A-B \) can be computed as 10%.

![Figure 1: Concept of Camera Motion Transition](image)

4. STATISTIC ANALYSIS FOR CATEGORIZING SPORTS VIDEO

In this section, we discuss statistical analysis of camera motion parameters which are extracted from broadcasted sports video using 2 features defined above. We use 6 game types of major broadcasted sports videos include 9 leagues as follows; Sumo(NAGOYA-BASHO league (SUMO-A) and AKI-BASHO league (SUMO-B)), Tennis (U.K. Open (Wimbledon)) (Tennis-A) and French Open (Tennis-B), Baseball (NPB: Japan professional baseball league (Baseball-A) and MLB (Baseball-B)), Soccer(90 Worldcup), Football(NFL), Basketball(NBA). All sports video sequences are exactly 15 minutes long from the game starting time, recorded from TV broadcast in Japan, 2002 and don’t include Advertisement films in between. We use 15 video sequences for each league.

4.1. Analysis using Camera Motion Extraction Ratio

Table 1 shows average camera motion extraction ratio of FIX (\( \bar{X} \)) for each sports type video, also shows variance (\( s^2 \)), skewness (\( a_1 \)) and kurtosis (\( a_4 \)) of average extraction ratio.

Each value (\( \bar{X}, s^2, a_1, a_4 \)) can be written as follows.

\[
\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}
\]

\[
s^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{X})^2}{N-1}
\]

\[
a_1 = \frac{\sum_{i=1}^{N} (x_i - \bar{X})^3}{Ns^3}
\]

\[
a_4 = \frac{\sum_{i=1}^{N} (x_i - \bar{X})^4}{Ns^4}
\]

\[
s = \sqrt{s^2}
\]

From skewness and kurtosis values of all types of sports, average extraction ratio can be approximated to be based on a normal distribution, because each skewness value is close to 1.00 and each kurtosis value is close to 3.00. This is a basement fact for following analysis. The average camera motion extraction ratio is dependent on the difference of types of sports.

Next we give an analysis on statistical difference for the average camera motion extraction ratio by using Tukey’s test method. As Tukey’s test needs homoscedasticity of each data, the homoscedasticity of average camera motion extraction ratios should be tested using Hartley’s test. Testing statistic value (\( F_{\text{max}} \)) can be written as follows.

\[
F_{\text{max}} = \frac{\max s^2}{\min s^2} = \frac{51.158}{11.569} = 4.422
\]

When level of significance is set as 5%, and the number of types of sports is 9, and the number of each type of sport is 15, \( F_{\text{max}} \) is smaller than \( F_{0.05}(9,15) = 5.38 \). This means that average ratio in table 1 has homoscedasticity.

Table 3 shows difference values of average camera motion extraction ratios between two types of sports. Here population mean test is done by one-way layout analysis of variance using
multiple comparison. Using Tukey’s test method, the testing statistic value can be written as follow.

\[ q(a,a(n-1);\alpha) \frac{V_n}{\alpha} \approx 5.587 \]

where \( q(a,a(n-1);\alpha) \) ( \( \alpha = 0.05, n = 9, a = 15 \) ) is Studentized range, \( V_n \) is mean square value of variation of internal level, which can be written as

\[ V_n = \frac{S_y}{a(n-1)} \]

\[ S_y = S_y - S_y \]

where \( S_y \) is a variation of internal level, \( S_y \) is a total variation and \( S_y \) is a variation of inter level.

From the testing statistic value above, in this experiment, when the absolute value is larger than 5.587, it shows that the difference of the camera motion extraction ratios between the 2 types of sports has statistical significant difference. The most of values for the different type of sports are larger than 5.587, and the values for the same type of sports, for example U.K Open and French Open are less than 5.587, these mean that the camera motion extraction ratios depend on the difference of the type of sports and don’t depend on the difference of the league of the same sports type.

### Table 1: Probability of Camera Motion “FIX”

<table>
<thead>
<tr>
<th></th>
<th>( \bar{x} )</th>
<th>( s^2 )</th>
<th>( a_3 )</th>
<th>( a_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sumo-A</td>
<td>83.991</td>
<td>12.401</td>
<td>-0.732</td>
<td>3.775</td>
</tr>
<tr>
<td>Sumo-B</td>
<td>83.640</td>
<td>14.133</td>
<td>-0.818</td>
<td>2.799</td>
</tr>
<tr>
<td>Tennis-A</td>
<td>69.712</td>
<td>51.158</td>
<td>-0.143</td>
<td>2.097</td>
</tr>
<tr>
<td>Tennis-B</td>
<td>65.980</td>
<td>23.449</td>
<td>-0.021</td>
<td>2.146</td>
</tr>
<tr>
<td>Baseball-A</td>
<td>70.092</td>
<td>21.630</td>
<td>-0.480</td>
<td>2.214</td>
</tr>
<tr>
<td>Baseball-B</td>
<td>71.478</td>
<td>21.664</td>
<td>0.316</td>
<td>2.037</td>
</tr>
<tr>
<td>Soccer</td>
<td>30.377</td>
<td>11.569</td>
<td>-0.634</td>
<td>2.782</td>
</tr>
<tr>
<td>Football</td>
<td>43.996</td>
<td>24.635</td>
<td>0.364</td>
<td>2.158</td>
</tr>
<tr>
<td>Basketball</td>
<td>49.442</td>
<td>27.454</td>
<td>0.330</td>
<td>2.274</td>
</tr>
</tbody>
</table>

### 4.2. Analysis using Transition ratio

The value for the Tennis and Baseball, and the value for the Football (NFL) and Basketball (NBA) in the table 3, show that there are not statistical differences between these types of sports respectively. This means that analysis using only camera motion extraction ratio is not enough for categorizing sports videos. Here we discuss a more detailed analysis for the camera motion parameters of NFL and NBA using Transition ratio defined in the section 3.2. Table 2 shows average transition ratios for FIX-PAN for NFL and NBA ( \( \bar{x} \) ), also shows variance ( \( s^2 \) ), skewness ( \( a_3 \) ) and kurtosis ( \( a_4 \) ) of average transition ratio. From skewness and kurtosis values of both types of sports, average transition ratio can be approximated to be based on a normal distribution. We can find that the average values of transition ratios are very different. After proof of homoscedasticity, each average value of transition ratio was analyzed by the test of population mean ( \( \mu \) ) based on the statistical hypothesis as follows

\[ H_0 : \mu_{NFL} = \mu_{NBA} \]

When a level of significance is set as 0.05, the hypothesis can be rejected. This shows that the transition ratio for FIX-PAN of NFL and NBA has statistical significant difference. This means that more detailed statistical analysis can be done by using transition ratio. Although camera motion parameters of NFL and NBA can’t be analyzed enough using only camera motion extraction ratio, NFL and NBA can be categorized clearly using transition ratio analysis.

### Table 2: Transition Ratio for "FIX-PAN"

<table>
<thead>
<tr>
<th></th>
<th>( \bar{x} )</th>
<th>( s^2 )</th>
<th>( a_3 )</th>
<th>( a_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Football</td>
<td>14.802</td>
<td>27.674</td>
<td>0.682</td>
<td>2.432</td>
</tr>
<tr>
<td>Basketball</td>
<td>44.760</td>
<td>51.636</td>
<td>0.607</td>
<td>3.407</td>
</tr>
</tbody>
</table>

### 4.3. Analysis on relationship between camera motion parameters and types of sports

Here we discuss a relationship between camera motion parameters and types of sports. In this section, we take Tennis videos as an example where dominant color information of images is greatly dependent on holding stadium. As you know, U.K Open tennis (Wimbledon) is held on green grass court and French Open tennis is held on red clay court as shown in Fig.2. Although these 2 leagues are the same type of sports, color information is very different. On the other hand, there is a certain relationship between camera motion parameters, and types of sports. Here the tests of population variance ( \( \sigma^2 \) ) and population mean ( \( \mu \) ) are done based on the statistical hypothesis as follows

\[ H_0 : \sigma_{UK}^2 = \sigma_{French}^2 \]

\[ H_0 : \mu_{UK} = \mu_{French} \]

When a level of significance is set as 0.05, both of hypothesis can not be rejected. This means that there is not statistical significant difference between U.K Open tennis and French Open tennis. In other words, there is no relationship between color information in video sequence and camera motion parameters. This means that camera motion parameters depend on types of sports, not depend on place of each game.

In case of Soccer video, the same test is done for videos of daytime game and videos of night-time game such as shown Fig. 3. The test result is that both of hypothesis can not be rejected where a level of significance is set as 0.05. This means that camera motion parameters in Soccer videos also depend on types of sports, not depend on holding time and weather conditions.
5. CONCLUSIONS

A method for categorizing broadcasted sports video using camera motion parameters is presented. After giving the definition of two new features “Camera Motion Extraction Ratio” and “Camera Motion Transition” for analyzing video sequence, we discuss statistical analysis of camera motion parameters. Then the 4th part discusses statistical analysis method for broadcasted sports video and the experimental results for sports genre identification. Although we use only Camera motion extraction ratio in the first analysis, more detailed analysis can be performed by using Camera motion transition (Transition ratio). The most important contribution from this work is that broadcasted sports video can be categorized by statistical analysis of only camera motion parameters. 6 major types of sports can be characterized statistically and these statistical characteristic depend on the types of sports and don’t depend on color information such as dominant color of typical images. This means that camera motion parameters can be very effective for representing higher level features of the video sequences. In addition, this can imply that analyzing camera motion parameter may reveal another aspect of features of video sequences such as camera staff’s shooting customs. The focus of this paper has been the study of statistical analysis of camera motion parameters $\text{FIX}$ and $\text{PAN}$ only. We have not evaluated more sophisticated analysis such as analysis on time information of each camera motion parameters and analysis on other camera motion parameters such as $\text{SHAKE}$.

REFERENCES


Table 3: Multiple Comparison

<table>
<thead>
<tr>
<th></th>
<th>Sumo-B</th>
<th>Tennis-A</th>
<th>Tennis-B</th>
<th>Baseball-A</th>
<th>Baseball-B</th>
<th>Soccer</th>
<th>Football</th>
<th>Basketball</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sumo-A</td>
<td>0.351</td>
<td>14.279</td>
<td>18.011</td>
<td>13.899</td>
<td>12.513</td>
<td>53.615</td>
<td>39.995</td>
<td>34.549</td>
</tr>
<tr>
<td>Tennis-A</td>
<td>---</td>
<td>---</td>
<td>3.732</td>
<td>0.380</td>
<td>1.766</td>
<td>39.335</td>
<td>25.715</td>
<td>20.27</td>
</tr>
<tr>
<td>Tennis-B</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-4.112</td>
<td>-5.498</td>
<td>35.603</td>
<td>21.983</td>
<td>16.538</td>
</tr>
<tr>
<td>Baseball-B</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>41.101</td>
<td>27.481</td>
<td>22.036</td>
<td>19.065</td>
</tr>
<tr>
<td>Football</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

Figure 2: U.K Open Tennis on green grass (left) and French Open Tennis on red clay (right)

Figure 3: Daytime game (left) and Nighttime game(right)