A REAL-TIME SCORE DETECTION AND RECOGNITION APPROACH FOR BROADCAST BASKETBALL VIDEO

Guangyi Miao[1], Guangyu Zhu[3], Shuqiang Jiang[2], Qingming Huang[1,2], Changsheng Xu[4], Wen Gao[2]

[1]Graduate School of Chinese Academy of Sciences Beijing, P.R. China, 100080
[2]Institute of Computing Technology Chinese Academy of Sciences Beijing, P.R. China, 100080
[3]School of Computer Science Harbin Institute of Technology Harbin, P.R. China, 150001
[4] Institute for Infocomm Research 21 Heng Mui Keng Terrace Singapore, 11961
{gymiao, gyzhu, sqjiang, qmhuang, wgao}@jdl.ac.cn, xucs@i2r.a-star.edu.sg

ABSTRACT

For broadcast sports video, score information is an effective mid-level representation to facilitate high-level video content analysis. In this paper, we propose a real-time approach to detect score region and recognize scores in broadcast basketball video. First, score region is automatically detected using frame difference and texture information without any prior knowledge. Then, score digit is recognized using a coarse-to-fine scheme based on score spatial structure, temporal correlation and changing rules. Compared with traditional video text recognition method, our approach is computing-insensitive and independent of digit model training. Experimental results show that our approach achieves real-time performance and is robust for the variation of digit font, size, color and noise caused by low frame resolution in broadcast basketball video.

1. INTRODUCTION

Score board appears in many scenes of broadcast basketball sports video since score is critical for a game and provides important information to the audience. Knowing the score can help us to analyze the process of the game and the tactic of each team. For example, using score information, we can obtain the information like "team A makes a highlight with 13:0 in one minute", "the two teams get few scores in five minutes". Therefore, by analyzing score change, we can extract abundant high-level semantic information which is useful for basketball sports video analysis.

Actually, there are two score regions in a score board overlay to indicate the scores of the two teams respectively. Figure 1 shows a few samples of score board in some basketball games including NBA and Olympic Game. Generally, the names of the two teams and game time are also shown on the overlay. In this paper, we care about the two score regions.

To obtain scores of the game, we need to detect and recognize the score characters in the frames. Some similar work has been done. Liu et al [1] used image analysis technology to detect the frames that contain texts or close captions, and used the video OCR technology to recognize the texts. But his approach is time consuming and can not work well on score characters with low resolution. In [2], the author used an algorithm for caption text extraction and recognition. But his approach needs to pre-select font bases as template characters and train the graph model and estimate the transition probabilities before recognition. In [3], the authors employed neural network to recognize digits. Unfortunately, it needs large samples to train the recognition model. In [4], a novel approach was proposed to detect and recognize the game time. The authors utilized the temporal periodicity for time runs periodically. But score does not have the periodically temporal changing pattern.

Figure 1. Variety of video score board overlay.

In our research, we face the following challenges as shown in figure 1: 1) the score images are severely blurred by low region resolution; 2) the score board may disappear sometimes and re-appear later; 3) there is no any prior knowledge about the appearance of the score digit in the video such as its font, size, and color because these characteristic varies among different videos; 4) the background of the score board is usually complex; 5) The algorithm needs to be compute-insensitive to realize fast processing. We propose a real-time approach to solve all the challenges.

This work was supported by National Hi-Tech Development Program (863 Program) of China under Grant: 2006AA01Z117 and National Hi-Tech Development Program (863 Program) of China under Grant: 2006AA01Z315

2. FRAMEWORK OF PROPOSED APPROACH

Figure 2 shows the flow chart of our proposed approach. There are two major modules: score detection and score recognition. In score detection module, first we locate the score board region based on the difference of consecutive frames and the gradient information of every frame. And then we detect two score regions in the score board by searching changed pixels in the score board region. Thus, we find the positions of the two scores. In score recognition module, we employ a coarse-to-fine approach to recognize the scores. First, a score is segmented into single digit characters. Then, we extract features from digit characters and recognize the character by rules which are defined hierarchically. After coarse recognition, we refine the results twice based on temporal correlation and changing rules of the score. Finally we obtain the recognized result.

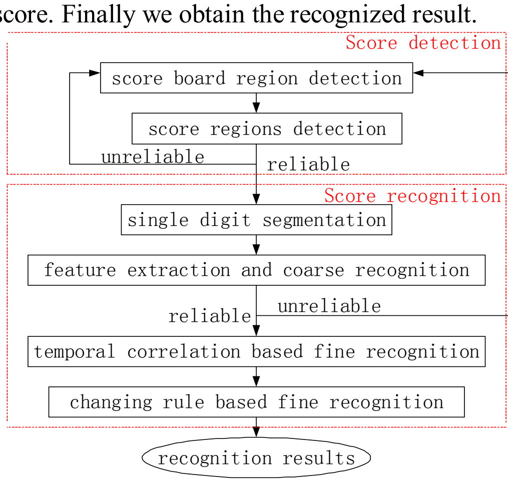

Figure 2. Algorithm flow chart

3. SCORE DETECTION

3.1. Score board region detection

Generally speaking, the position of the score board does not change in the video and the score board region has abundant texture information. Such characteristics will help us to fast detect the score board. We calculate the difference map of two consecutive frames and the gradient map of every frame, incorporate the two maps and get a incorporate map. Then we employ density-based region rowing method [5] to extract candidate score board regions. Some of the candidate regions are discarded based on prior knowledge such as the size, shape and position of the score board region. We select the region with the maximum probability as the score board region.

In our approach, we process 25 frames from 25 consecutive seconds and synthesize the results to locate score board region exactly. Once score board is located, we obtain a score board pixels template and update it every frame. In some video frames, the score board may disappear because of the replay insertion, game break or advertisement broadcasting. Using this template, score board detection is robust for replay and advertisements.

3.2. Score regions detection

For some basketball video such as NBA, the score board is always complex, as shown in figure 3(a). A scoreboard region will often contain multiple components including text, images, and line drawings. It is difficult to locate the score regions in the board. Connected Component Analysis (CCA) often gets false result. In basketball game, the scores are always changing according to the game process. So we randomly select score boards from different frames. Then, we compare these boards and search changed pixels as shown in figure 3(b). Changed pixels are connected into regions. Time region has a bigger width and height ratio. The two score regions have similar height and their positions are parallel with the same height or width. With this knowledge, we can locate time region and score regions.

If time region and score regions cannot be located, it means that we located a wrong score board region. This region is discarded and we relocate a new score board region.

(a) complex score board (b) changed pixels in the board

Figure 3. Score regions detection.

4. SCORE RECOGNITION

We propose a coarse-to-fine approach to recognize score characters. In the coarse recognition, some reliable features are extracted from segmented single character. Digit characters are recognized using hierarchical rules. In the fine recognition, temporal correlation and score changing rule are considered for two-time optimization. Finally we achieve a reliable result.

4.1. Feature extraction from a digit character

We have the following knowledge about the scores in score board: each score has maximum 3 digits from right to left representing ONE, TEN and HUNDRED respectively. A projection profile operation [6] is employed to separate score region into single digit characters. After separation, the pixels in every single digit character region are binarized.

Though score digit characters may be various in terms of font and size, the structure of score digits is standard and consistent. Thus, we can extract reliable features based on digit structure analysis. The features are extracted from two aspects: *CR* and *TP*, which are defined as below and shown in figure 4.

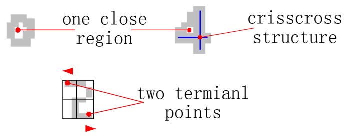

Figure 4. Examples of *CR* and *TP*.

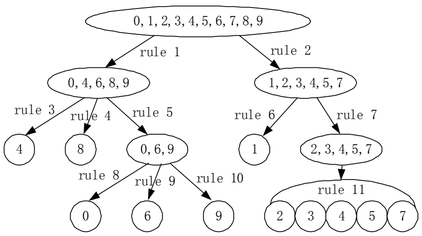

Figure 5. Coarse recognition.

Table 1. Rules for coarse recognition.

rule 1	*CloNum* is not 0 , and *TerNum* is less than 1
rule 2	*CloNum* is 0 , and *TerNum* is more than 2
rule 3	*CloSize* is smaller than a threshold, and it has a *Crisscross*
rule 4	*CloNum* is 2
rule 5	*CloNum* is 1 and it does not have *Crisscross*
rule 6	width is less than a threshold
rule 7	width is less more than a threshold
rule 8	*TerNum* is 0, and *CloSize* is bigger than a threshold, and *CloPos* is at the middle position of the character
rule 9	*CloSize* is smaller than a threshold, and *CloPos* is at the bottom of the character
rule 10	*CloSize* is smaller than a threshold, and *CloPos* is at the top of the character
rule 11	defined in figure 6

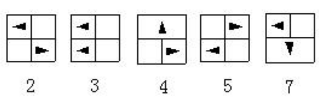

Figure 6. Rule 11 based on *TerPos* and *TerDir*.

CR is defined as the feature of the close regions of a digit character as follows: 1) the number of close regions (*CloNum*), 2) the position of every close region (*CloPos*), 3) the size of every close region (*CloSize*), and 4) the crisscross structure (*Crisscross*). As we can see in figure 4, for character '0', *CloNum* is 0, *CloSize* is 8 pixels and *CloPos* is the middle position of the height. For character '4', it has a *Crisscross* structure.

TP is defined as the feature of terminal points of a digit character. First, the character is thinned without destroying its connectivity. Then, the *TP* based features are extracted, which include: 1) the number of terminal points (*TerNum*), 2) the position of every terminal point (*TerPos*), and (3) the direction of every terminal point (*TerDir*). In figure 4, the character region is divided into 2×2 grids. For character '2', the *TerNum* is 2. For one terminal point, *TerPos* is at top-left grid and *TerDir* is left; for the other terminal point, *TerPos* is at right-bottom grid and *TerDir* is right.

4.2. Coarse recognition

We exploit a rule-based scheme to coarsely recognize digit characters. The digits are hierarchically classified as shown in figure 5. The rules are defined in table 1.

For character satisfying rule 2 and rule 7, it may be '2', '3', '4', '5', or '7'. It is further recognized by rule 11, as shown in figure 6.

If none of the rules mentioned above is satisfied, we discard the result of this character. This will reduce the misrecognition which will bring bad effect to the following fine recognition.

4.3. Temporal correlation based fine recognition

In the fine process, we first use temporal correlation of the score to conduct the recognition. Score digits change slowly. In our experiments, a digit will at least last 2 seconds before it changes and the frame rate of the video is 25. Therefore, a digit will last at least 50 frames in the video before it changes. We compare coarse results of several consecutive frames and obtain the result with maximum probability. The steps of proposed method are illustrated in figure 7.

1) Get the coarse results of N consecutive frames CR_{ij}, where i ranges from 0 to 2 representing the digit ONE, TEN and HUNDRED of a score, and j ranges from 0 to $N-1$ representing the coarse results from current frame to prior $N-1$ frame.
2) Vote digit 0 to 9 according to CR_{ij} and calculate $Vote_{in}$, where n ranges from 0 to 9, representing the ballot of every digit.
3) Arrange $Vote_{in}$ and obtain $Digit_{i,prob}$ which is the digit with the $prob$ th probability to be the result.
4) The fine result $FR1$ is recognized by Eq. (1) as below in which the three $prob$ are set to (1,1,1) for three $Digit_{i,prob}$ respectively.

Figure 7. The first fine recognition

$$FR1 = Digit_{0,prob} + Digit_{1,prob} \times 10 + Digit_{2,prob} \times 100 \qquad (1)$$

After this process, accuracy improves to a high value.

4.4. Changing rule based fine recognition

In basketball game, the score changing rule is: score can only move forward with step no more than 3. If score decreases or adds 4 once, it is wrong with the result $FR1$. We need to improve the result. The steps are as follows.

If $FR1$ does not satisfy score changing rule, we recalculate $FR1$ according to $Digit_{i,prob}$. The formula is also Eq. (1). Subscript $prob$ can be set as 1 or 2. When $prob$ is set as 2, $Digit_{i,prob}$ means the digit with the second probability to be the result. There are three $Digit_{i,prob}$ in Eq. (1), so there are $2^3 = 8$ cases for Eq. (1) with different $prob$ value for the three $Digit_{i,prob}$ respectively and orderly: (1,1,1), (2,1,1), (1,2,1), (2,2,1), (1,1,2), (2,1,2), (1,2,2), (2,2,2). Case (1,1,1) is used in 4.3, so we try other seven cases orderly. If the score changing rule is satisfied, we use it as the final result $FR2$. If all the seven results are wrong, we set $FR2$ as the same as the result of prior frame.

5. EXPERIMENTAL RESULTS

Our experiments were conducted at a personal computer with P4 1.4Ghz processor and 256M main memory. To demonstrate the effectiveness of the proposed approach, we carried out the experiment on 18 broadcast basketball video segments. The videos are recorded from NBA and Olympic Game live broadcast television program. The videos all have replay insertion, game break or advertisement broadcasting during the game. The resolution of our tested video data is 352×288.

Table 2 shows the score detection result. For score board region detection, all the score board regions in the test videos can be detected correctly in a few seconds. For score region location, only one video cannot be located correctly because the score board in the video is semitransparent and the background changes sharply.

Table 3 shows the speed of score recognition algorithm. Excluding the run time of video decoding, our approach can process about 602 frames per second averagely. This means that our approach is real-time.

Table 4 lists the score recognition accuracy after correct locations of score regions are found. 6 different basketball game video segments with different length (the shortest is 25 minutes and the longest is 50 minutes) are tested. For convenience, we check every score changing frame to calculate the accuracy. For the reason of fine recognition, the checked frame will delay about 10 frames compare with real score changing frame. If it is less than 20, we set it as correct. The false recognition is mainly because the score characters in the video are always blurred and faint severely.

6. CONCLUSION

In this paper, we propose a real-time approach to detect score region and recognize scores in broadcast basketball video. With our approach, we can locate score regions reliably without detecting all the texts in the frame, and recognize the scores free of any training. It is simple and very fast. So it can be integrated into a system without affecting its speed. Experimental results show that our approach is fast and robust. It can also be used to other sports games with the pattern of two rivals and high scores, such as volleyball.

Table 2. Score detection accuracy (%)

	right	total	accuracy
Score board detection	18	18	100%
Score regions detection	17	18	94.4%

Table 3. Score recognition speed

	frames per second
recognition only	601.93
decoding only	76.00
decoding + recognition	67.48

Table 4. Score recognition accuracy (%)

gn	1	2	3	4	5	6	average
sct	20	40	24	32	32	48	/
drr	100	90	83.3	84.4	87.5	97.9	91.35

gn: game number. sct: score change times. drr: digit recognition rate = 100 * (times of score recognized correctly) / (total score change times).

1: Olympic 2005 Korea Vs China on August 14. 2: Olympic 2005 China Vs Spain on August 16. 3: Olympic 2005 Argentina Vs Spain on August 17. 4: NBA 2006 MIA Vs CHI on April 28. 5: NBA 2006 PHX Vs LAL on May 5. 6: NBA 2006 DET Vs MIA on May 28.

7. ACKNOWLEDGEMENTS

This work was supported in part by National “242” project (2006A09), “Science100 Plan” of Chinese Academy of Science(99T3002T03), and the Natural Science Foundation of Beijing, China (4063041)

8. REFERENCES

[1] H. Liu, D. Zhou. “Content-based News Video Story Segmentation and Video Retrieval”, Proc. of SPIE Image and Graphics, pp.1038-1044, 2002

[2] D. Zhang, S. Chang, "Event detection in baseball video using superimposed caption recognition", ACM MM, pp. 315-318, 2002.

[3] E. Kim, G. Lee, C. Jung, S. Kim, J. Kim, W. Kim, "A Video Summarization Method for Basketball Game", PCM, pp. 765–775, 2005.

[4] Y. Li, C. Xu, K.W. Wan, X. Yan, X. Yu, "Reliable Video Clock Time Recognition", International Conference on Pattern Recognition (ICPR), pp. 128-131, 2006

[5] Q. Ye, Q. Huang, W. Gao, D. Zhao, "Fast and robust text detection in images and video frames", Image and Vision Computing, pp.565-576, 2005.05

[6] R. Lienhart, A. Wernicke, "Localizing and segmenting text in images and videos", IEEE Transactions on Circuits and Systems for Video Technology, pp.256-268, 2002