Soccer Video Segmentation: referee and player detection

Juan Rafael Nuñez and Jacques Facon and Alceu de Souza Brito Junior
PPGIA-PUCPR Pontifícia Universidade Católica do Paraná
Rua Imaculada Conceição 1155, 80215-901Curitiba-Pr, Brazil
Phone: (55) 41-32711791 Fax: (55) 41-32712121 E-mail: juanrafael@pop.com.br,{facon,alceu}@ppgia.pucpr.br

Keywords: video segmentation, soccer game, referee and player identification.

Abstract - This paper presents a player segmentation methodology for videos of soccer games. This methodology includes some low and high level video processing algorithms, such as dominant color region detection, referee and player identification. Experimental results have shown the efficiency of the methodology in locating and segmenting referee and players.

1. INTRODUCTION

Many approaches for sport video segmentation and interpretation have been proposed in recent years. By video segmenting and interpreting, we mean dividing sequences of frames into smaller meaningful units which represent information at the scene level. To carry out any performing content video analysis, an effective and accurate segmentation is crucial.

Related works in the literature of sport video analysis have addressed different sport games. For soccer video, Haiping *et al.* [2] have developed a semantic labeling of soccer video based on three steps: a short training video to automatically compute the field color; a video segmentation into relatively static and active parts; and a post-processing, where adjacent active sub-parts are merged with static parts to form semantic segments. As depicted by the authors, the best results are obtained to label "AD" (Far view of audience) with an accuracy of 84.6%.

Ahmet *et al.* [1] have proposed a computationally framework for soccer video analysis and summarization including low-level algorithms, such as: dominant color region segmentation, shot detection and classification, as well as high-level algorithms related to goal, referee, and penalty-box segmentation. By using a database composed of 17 MPEG-1 clips, the proposed algorithm, on the average, achieves a 90.0% recall and 45.8% precision rates.

Another interesting proposal has been presented by Vandenbroucke *et al.* [8]. They have defined a hybrid color space to better discriminate pixel classes and have applied this classification scheme to segment the soccer players and identify their teams. From the two examples used by authors, the classification error rates run around 11%.

Peng Xu *et al.* [4] have proposed a system for Play-Break soccer video segmentation, based on a Learning Color-Based Grass Detector and an unsupervised Grass color segmentation. The best result in segmentation is 86.5%.

A new methodology for detection of referee and players in soccer videos is presented in this paper. For this purpose, we have defined an unsupervised clustering to identify the players and the referee locations. Moreover, a robust evaluation protocol is introduced, in which the importance of the different stages of the proposed process can be weighted.

The paper is organized as follows. Section 2 describes the dominant color segmentation robust to lighting variations. Section 2 also describes the dominant color stage based on grass field color and lighting variation histogram and a zero-crossing step. Section 3 presents the player and referee detection. Section 4 explains the player team and referee identification. Section 5 presents the novel quantitative proposed evaluation. Experimental results over soccer video from different regions of the world and different format and quality are discussed in Section 6.

2. DOMINANT COLOR DETECTION

Even if the assumption that a soccer field has one distinct dominant color (green) seems valid, one cannot forget that weather, lighting and color variations do not permit to assume a specific field color value. To handle possible variations of lighting, we adaptively determine the grass color of the video frame by applying the followed methodology:

A. Convert RGB Frames into the HSV space: To segment the grass color as a homogeneous region the HSV space has shown to be more suitable [5];
B. Compute the histogram of the hue component and locate the highest peak P_H of the hue histogram: this step permits to detect the dominant hue for the grass;
C. Define the interval [H_{min}, H_{max}] around PH: this step is important to better define all the hue variations for the grass. The following method is applied:
   C.1: The hue component histogram is divided into 2 sub-histograms H_i (i=1, 2). The first one, H1, represents the histogram between 0 degree until P_H whereas the second one, H2, represents the histogram between P_H until 360 degrees.
   C.2: From each sub-histogram Hi (i=1,2), cumulative sub-histograms S_i(h) (i=1,2) are computed.

$$S_1(h) = \sum_{j=h}^{PH} P_1(j), \quad 0 \le h \le P_H$$

$$S_2(h) = \sum_{j=PH}^{PH+h} P_2(j), \; P_H \leq h \leq 360\text{-}P_H$$

C.3: From each cumulative sub-histogram Si(h) (i=1,2), the first derivative Di(h) (i=1,2) are then computed by using D_i(h)= D_i(h+1) – D_i(h), i= 1,2.
C.4: The Highest peaks M_i (i=1,2) for each derivative D_i(h) are computed.
C.5: The Nearest Zero-crossings M_{oi} (i=1,2) to each highest peak M_i (i=1,2) are then detected.
C.6 Then H_{min} and H_{max} in the Hue histogram are directly defined from M_1 and M_2 respectively.

3. PLAYER AND REFEREE DETECTION

The player and referee detection begins with Hue component threshold: the threshold process consists in binarizing Hue component from the [H_{min}, H_{max}] interval. Hue pixel values between H_{min} and H_{max} are set to black (the grass field) and the rest is set to white (the other regions). The threshold process permits to create a binary image where players and referee (white label) are surrounded by the background (black label).

Due to the presence of other patterns (like the public, publicity spots, etc…), the binary image presents other white regions than players and referee. However, these additional white regions are isolated since they are out of the background representing the field. The first threshold process is followed by a filtering, which consists in deleting these additional white regions.

A sequence of morphological tools (erosion, reconstruction, dilation) succeeds in preserving players and referee regions without changing their own geometry [6] [7]. Even if this filtering process preserves the regions of interest, it may remain some noisy regions. Due to weather, lighting and color variations, the complete elimination is not always guaranteed in all cases. This situation will be solved in the player and referee identification.

4. PLAYER AND REFEREE IDENTIFICATION

The preserved regions in the last process are now labeled to be classified into players and referee by using an unsupervised clustering. Some questions must be answered:

- How many classes are present in the image?
- Which region belongs to which class?

In the processed images only remain players, referee and some artifacts (regions wrongly segmented). But a goalkeeper has not the same color than the other players of the same team. Thus, an image may contain 6 classes:

- Players of team 1
- Players of team 2
- Referee
- Goalkeeper of team 1
- Goalkeeper of team 2
- Other regions wrongly segmented.

In order to automatically define the number of classes *CN* in each frame, a specific procedure to recognize the referee, players and teams has been developed. Let us consider:

- N the number of segmented classes;
- R_j one of these N classes, (i=1 to N);
- HR_j the histogram of R_j, (i=1 to N);
- $G_j = \{h_{1j}, h_{2j}, h_{3j}, h_{4j}, h_{5j}\}$, the most significant hue set of R_j computed as the maximum sum ($h_{1j}+h_{2j}+h_{3j}+h_{4j}+h_{5j}$) of 5 successive hues of HR_j;
- CV_j the vector of the N most significant central values h_{3j} of HR_j, (i=1 to N). Then $CV_j = \{h_{31}, h_{32}, \ldots, h_{3j}, \ldots, h_{3N}\}$;
- OCV_j the ordered CV_j vector;
- $DOCV_j$ the first derivative of OCV_j;
- AD_j the average of $DOCV_j$;
- *GL* the threshold value vector of $DOCV_j$;
- *CN* the number of classes.

The procedure is:

- Label the thresholded classes R_j, (i=1 to N);
- Assign to each class R_j its histogram HR_j, (computed by using the H component of the pixels in the original frame at the same position than in R_j);
- Analyze the N histograms HR_j in order to locate the most significant hue set G_j of HR_j;
- Define, from each G_j, the most significant central value h_{3j};
- Define CV_j the vector of N significant central values h_{3j} of HR_j;
- Define OCV_j VC(j) by ordering CV_j;
- Compute the first derivative $DOCV_j$ of OCV_j;
- Compute AD_j, the average of $DOCV_j$;
- Use AD_j as a threshold value to define the threshold value vector *GL* of $DOCV_j$;
- Deduce the number of classes *CN*;
- Analyze back CV_j using *GL* threshold values to define the bounds of each class

5. QUANTITATIVE PERFORMANCE EVALUATION

We conjecture that evaluating the performance of player segmentation methodologies in soccer games could be different depending on results to be obtained. A novel quantitative evaluation has been proposed in this context aiming to measure each level and its influence in the methodology. Each methodology level is weighted and the fact of including or excluding partial players and referee located at the physical edges of the image was considered.

The novel quantitative evaluation is described as follows (Table 1):

- *GRAS* weighted by α: Predominant region in the image. Measured by the number of correct grass segmented pixels divided by the ground-truth number of grass pixels;
- *ADJA* weighted by β: Adjacent regions in the image (regions other than grass, players, referee ones, like the

public, publicity spots etc…). Measured by the number of correct Adjacent region segmented Pixels divided by the ground-truth number of Adjacent region pixels;

- $CPRS_i$ weighted by δ: Correct player and referee segmentation: Measured by the number of correct segmented players and referee divided by the ground-truth number of players and referee; Including partial players and referee located at the physical edges of the image;
- $CPRS_e$ weighted by δ: Correct player and referee segmentation: Measured by the number of correct segmented players and referee divided by the ground-truth number of players and referee; Excluding partial players and referee located at the physical edges of the image;
- $CPRI_i$ weighted by ε: Correct player and referee identification: Measured by the number of correct identified players and referee divided by the ground-truth number of players and referee; Including partial players and referee located at the physical edges of the image;
- $CPRI_e$ weighted by ε: Correct player and referee identification: Measured by the number of correct identified players and referee divided by the ground-truth number of players and referee; Excluding partial players and referee located at the physical edges of the image;
- *CSIR* weighted by ϕ: Correct segmented and identified regions: Measured by the number of correct identified clusters divided by the ground-truth number of clusters;

The two first level scores were computed by comparing region pixels at the same location in ground-truth images and segmented ones while other level scores are computed by visual inspection. Then a weighting strategy has been suggested to attain a more robust evaluation, where each level score is weighted and a final methodology evaluation (Quality Rates QR_i.and QR_e) is then computed.

The Quality rates QR_i.and QR_e have been computed as follows:

QR_i including partial players and referee located at the physical edges of the image

$$QR_i = \alpha GRAS + \beta\, ADJA + \delta\, CPRS_i + \varepsilon\, CPRI_i + \phi\, CSIR$$

QR_e excluding partial players and referee located at the physical edges of the image

$$QR_e = \alpha GRAS + \beta\, ADJA + \delta\, CPRS_e + \varepsilon\, CPRI_e + \phi\, CSIR$$

6. EXPERIMENTS

Fourteen international soccer games with different quality (from standard signal to digital cable), *codecs (XVid and DivX)*, *bitrate (from 782 to 2704)*, *FPS* (29 and 25) and resolution (from 352x288 to 960x544) were used in our experiments. From each video, a sequence of minimum 20 frames was randomly chosen.

The determinant grass region is adaptively segmented by dominant hue interval $[H_{min}, H_{max}]$ detection (Section 2). The ground-truth values are handily computed and the GRAS score was computed by comparing region pixels at the same location in ground-truth images and in segmented ones, where the ground-truth values are handly computed by two experts.

The segmentation of players and referee is obtained by thresholding the $[H_{min}, H_{max}]$ interval as explained in Section 3. The referee and player identification was carried out using an unsupervised clustering (Section 4) where no assumption about the existing classes is formulated. Any of 6 classes can be potentially present. The segmentation and identification accuracies are computed by visual inspection.

The results described in Table 2 show the efficiency of the determinant grass region with only two rates lower than 95% and with an average rate of 94.20% and a median rate of 98.21%

The player and referee segmentation show consistent performance. By including partial players and referee located at the physical edges of the image, the mean and median $CPRS_i$ rates are 76.17 and 77.86% (Table 2), respectively. By excluding partial players and referee located at the physical edges of the image, the average and median $CPRS_i$ rates grow up and show accurate segmentation with 86.01 and 85.85%, respectively.

Even if the efficiency of the determinant grass region detection was proved, the results also show that only using hue interval $[H_{min}, H_{max}]$ does not guarantee a segmentation of players and referee as expected and another technique to improve the results must be coupled.

By including partial players and referee located at the physical edges of the image, the mean and median $CPRI_i$ rates are 57.95 and 58.79%, respectively (Table 2). While, by excluding partial players and referee, the average and median $CPRI_i$ rates show a little increasing with 66.48 and 64.79 %, respectively.

The accuracy of the methodology is computed through QR_i.and QR_e. Two sets of weights were applied (Table 1). In the first one, greater weights have been assigned to α (*GRAS)* and β (*ADJA)* while, in the second one, greater weights have been assigned to δ *(*$CPRS_i$ and $CPRS_e$) and ε *(*$CPRI_i$ and $CPRI_e$).

By using higher weights to α (*GRAS)* and β (*ADJA)* in the first evaluation, QR_i. and QR_e show the real efficiency of the methodology (higher than 84% in Table 2). The inclusion or exclusion of partial players and referee located at the physical edges of the image did not influence QR_i and QR_e rates. The QR_i average and median are higher than 84%, while QR_e ones are higher than 85%, respectively.

By weighting more δ *(*$CPRS_i$ and $CPRS_e$) and ε *(*$CPRI_i$ and $CPRI_e$) in the second evaluation, QR_i. and QR_e show interesting results (higher than 71%). The inclusion or exclusion of partial players and referee located at the physical edges of the image plays a higher influence onto QR_i. and QR_e rates than in the first evaluation. The QR_i average and median are higher than 71%, while QR_e ones are higher than 77%, respectively.

Unfortunately, due to the lack of ground-truth databases, experimental protocols and evaluation processes available in the literature, numerical comparisons with related works were not possible. Despite of this fact, we can say that our approach is promising. The objective was achieved since we have shown that our method can achieved about 85% of accurate segmentation on fourteen soccer videos with different configurations and under lightening variations.

Weights	Weights of 1st Evaluation	Weights of 2nd Evaluation
α (*GRAS*)	35%	10%
β (*ADJA*)	35%	10%
δ ($CPRS_i$ and $CPRS_e$)	10%	35%
ε ($CPRI_i$ and $CPRI_e$)	10%	35%
ϕ (*CSIR*)	10%	10%

Table 1 – Weights used in the Quantitative Performance Evaluation

4. CONCLUSIONS

A player segmentation methodology in soccer games was proposed. We developed effective tools for detecting dominant color region, for segmenting and identifying referee and players. The dominant color region detection process has shown to be efficient for great variety of weather, lighting, color variations and quality of soccer games. The results of the player and referee segmentation show that only the dominant hue to extract players and referee is limited. The results of the new referee and player identification, based on an unsupervised clustering and some assumptions show the methodology to be promising in recognizing the two team players and the referee. We design to make available ground-truth databases, experimental protocols and evaluation processes to perform objective evaluations.

REFERENCES

[1]Ahmet Ekin, A. Murat Tekalp, and Rajiv Mehrotra "Automatic Soccer Video Analysis and Summarization", IEEE Transactions On Image Processing, Vol. 12, No. 7, July 2003

[2] Haiping Sun, Joo-Hwee Lim, Qi Tian, Mohan S. Kankanhalli, "Semantic Labeling Of Soccer Video" -Icics-2003

[3] Ohlander R., K. Price, D.R. Reddy, Picture segmentation using a recursive region splitting method, Computer Graphics Image Processing 8 (1978) 313–333.

[4] Peng Xu, Lexing Xie, Shih-Fu Chang, "Algorithms And System For Segmentation And Structure Analysis In Soccer Video"- Proc Icme 2001

[5] Sangwine, S., Horne, R., The Colour Image Processing Handbook. London: Chapman & Hall, 1998.

[6] Serra, J., Mathematical Morphology: Theoretical Advances. New York: Academic Press, 1988, Vol. 2.

[7] Soille Pierre, Morphological Image Analysis: Principles and Applications. Berlin: Springer-Verlag, 1999.

[8] Vandenbroucke Nicolas, Ludovic Macaire, and Jack-Gerard Postaire: "Color image segmentation by pixel classification in an adapted hybrid color space. Application to soccer image analysis", Computer Vision and Image Understanding 90 (2003) 190–216

	Global Measures(%)							1st Evaluation (%)		2nd Evaluation (%)	
Video	GRAS	ADJA	$CPRS_i$	$CPRS_e$	$CPRI_i$	$CPRI_e$	CSIR	QR_i	QR_e	QR_i	QR_e
1	98,32	80,75	76,92	82,47	58,62	63,75	64,06	**82,64**	**83,70**	**71,75**	**75,49**
2	78,69	87,43	58,20	68,75	66,91	84,55	82,26	**78,88**	**81,70**	**68,63**	**78,49**
3	97,85	98,56	78,69	98,46	51,04	64,06	61,79	**87,90**	**91,18**	**71,22**	**82,70**
4	98,87	69,34	90,67	98,87	46,60	50,86	71,13	**79,71**	**80,96**	**71,98**	**76,34**
5	99,42	95,53	65,00	69,40	57,48	63,08	64,23	**86,90**	**87,90**	**68,78**	**72,28**
6	98,67	92,10	91,79	100,00	58,97	64,25	66,12	**88,46**	**89,81**	**78,46**	**83,18**
7	98,10	98,93	77,86	100,00	58,57	75,23	80,65	**90,67**	**94,55**	**75,52**	**89,10**
8	98,76	95,27	85,65	96,95	66,36	75,39	71,91	**90,30**	**92,34**	**79,80**	**86,92**
9	97,62	87,62	51,67	59,24	44,83	55,91	73,54	**81,84**	**83,70**	**59,65**	**66,18**
10	95,93	78,48	80,73	84,62	61,83	65,34	65,89	**81,89**	**82,63**	**73,93**	**76,52**
11	98,72	92,63	82,05	87,07	72,63	77,73	74,82	**89,92**	**90,94**	**80,75**	**84,30**
12	98,83	77,13	71,49	83,58	38,12	45,83	68,04	**79,35**	**81,33**	**62,76**	**69,70**
13	95,53	94,73	85,36	95,33	62,88	70,59	83,75	**89,79**	**91,56**	**79,28**	**85,47**
14	63,48	87,50	72,80	79,50	66,51	74,21	77,10	**74,48**	**75,92**	**71,57**	**76,61**
Mean	94.20	88.28	76.17	86.01	57.95	66.48	71.80	**84.48**	**86.30**	**72.43**	**78.80**
Median	98.21	89.86	77.86	85.85	58.79	64.79	71.52	**84.77**	**85.80**	**71.86**	**77.55**
Standard Deviation	10.29	08.91	12.14	13.22	09.70	10.73	07.18	**05.26**	**05.56**	**06.24**	**06.75**

Table 2 – Quantitative evaluation