Player Number Localization and Recognition in Soccer Video using HSV Color Space and Internal Contours

Matko Šarić, Hrvoje Dujmić, Vladan Papić and Nikola Rožić

Abstract—Detection of player identity is challenging task in sport video content analysis. In case of soccer video player number recognition is effective and precise solution. Jersey numbers can be considered as scene text and difficulties in localization and recognition appear due to variations in orientation, size, illumination, motion etc. This paper proposed new method for player number localization and recognition. By observing hue, saturation and value for 50 different jersey examples we noticed that most often combination of low and high saturated pixels is used to separate number and jersey region. Image segmentation method based on this observation is introduced. Then, novel method for player number localization based on internal contours is proposed. False number candidates are filtered using area and aspect ratio. Before OCR processing extracted numbers are enhanced using image smoothing and rotation normalization.

Keywords—player number, soccer video, HSV color space

I. INTRODUCTION

CONTENT-BASED video analysis targeted at sports events is interesting research area with high commercial potential. Semantic annotation of video stream allows efficient browsing, searching and retrieval of video material. Many efforts in sports video annotation have been made on important events (highlights) detection like goals or shots on goal for soccer video. The comprehensive survey of trends in sports-related indexing and retrieval can be found in [1].

Automatic player identification would be of great importance in soccer video retrieval or in providing augmented information about player. Presence of jersey's number in soccer video is exploited to perform player recognition.

Text in video images is classified into caption text, which is artificially overlaid on the image, or scene text, which exists naturally in the image. Player numbers can be considered as scene text. Detection of this kind of text is generally more difficult because it can have different orientations and it is often distorted by variations in illumination, focus, motion etc. A survey of image and video text information extraction can be found in [2].

There are few methods for player number detection that can be found in literature. Ye et al. [3] proposed jersey number detection method where image is segmented using generalized learning vector quantization algorithm that clusters pixels into limited color-homogeneous regions in order to separate the jersey number from its background. Size and pipe-like attributes of digital characters are used to filter the candidates. K-NN classifier and Zernike moment features are employed for number recognition. In [4] player numbers are recognized using algorithm proposed by Viola and Jones ([5]). This algorithm is originally intended for face detection and it is based on cascade of classifiers. Number detection in [6] is achieved with two-stage approach: first step is extraction and selection of image corners using the Harris detector and second step is extraction of Maximally Stable Extremal Regions. These steps segment image in order to process it with OCR software. Andrade et al. ([7]) use region adjacency graph and picture trees to perform object search using prior knowledge. Region analysis is further applied to the candidate regions to isolate player number which are then processed with OCR.

By observing 50 jersey examples we noticed that in most cases pixels in number and jersey regions significantly differ in their saturation, i.e. number or jersey region pixels lie in achromatic range of HSV color space. Using results from our observation we propose novel image segmentation method combined with internal contour extraction in order to locate number regions. Firstly, image is segmented in chromatic and achromatic pixel regions to separate number and jersey. Next step is separation of jersey region from its background using intensity (if jersey has achromatic color) or hue (if jersey has chromatic color). Result of this procedure is jersey represented as white object with black "hole" (number). Number candidates are extracted as internal contours of white objects. These candidates are then filtered using area and bounding rectangle aspect ratio. Image smoothing and rotation normalization are employed before OCR processing. In cases when jersey and number region pixels lie in chromatic range segmentation is based on their hues.

Rest of the paper is organized as follows. Section 2 discusses image segmentation. Section 3. describes number candidates extraction using internal contours in segmented image. Results are presented in section 4. Conclusions are made in section 5.

Authors are with the Faculty of electrical engineering, University of Split, R. Boskovica bb, Split, Croatia.
e-mail: msaric@fesb.hr, hdujmic@fesb.hr, rozic@fesb.hr

II. IMAGE SEGMENTATION

In soccer video number region must have large color contrast with jersey region to make number clearly visible in different conditions. We investigated how this contrast is manifested in HSV color space. By observing 50 different jerseys of famous European football clubs we noticed that in 43 examples contrast is achieved using combination of low saturated (achromatic) and high saturated (chromatic) pixels. In 7 examples jersey and number pixels have high saturation and noticeable hue difference, i.e. hue is used to separate these two regions.

Standard segmentation method based on hue, saturation and intensity thresholding would extract many non-number objects and numbers wouldn't be separated clearly. Here we propose new number region segmentation method based on conclusions derived from our observation. The procedure is described as follows:

1. *Using sample image define mean colors for jersey and number (denoted by the HSV vectors* $\mathbf{j} = \begin{bmatrix} j_h & j_s & j_v \end{bmatrix}$ *and* $\mathbf{n} = \begin{bmatrix} n_h & n_s & n_v \end{bmatrix}$*). Pixel color is denoted by* $\mathbf{pixel} = \begin{bmatrix} pixel_h & pixel_s & pixel_v \end{bmatrix}$

2. *Create binary image*:

a) *Number has achromatic color and jersey has chromatic color.*

$$if\ \left(\left(n_s < \lambda_s \right) or \left(n_v < \lambda_v \right) \right) and \left(\left(j_s > \lambda_s \right) and \left(j_v > \lambda_v \right) \right)$$

   for each **pixel**

      if $\left(pixel_s < \lambda_s \right) and \left(pixel_v < \lambda_v \right)$

         set **pixel** *as black*

      else

         if $\left(h\_diff \left(j_h, pixel_h \right) < hue\_thresh \right)$

            set **pixel** *as white*

         else

            set **pixel** *as black*

$$h\_diff \left(j_h, pixel_h \right) = \begin{cases} \Delta \left(j_h, pixel_h \right) & if\ \Delta \left(j_h, pixel_h \right) < 180° \\ 360° - \Delta \left(j_h, pixel_h \right) & otherwise \end{cases}$$

$$\Delta \left(j_h, pixel_h \right) = \left| j_h - pixel_h \right|$$

b) *Jersey has achromatic color and number has chromatic color*

$$if\ \left(\left(j_s < \lambda_s \right) or \left(j_v < \lambda_v \right) \right) and \left(n_s > \lambda_s \right) and \left(n_v > \lambda_v \right)$$

$$if\ \left(j_v > V\_thresh \right)$$

   for each **pixel**

      if $\left(pixel_s < \lambda_s \right) and \left(pixel_v < \lambda_v \right)$

         if $\left(pixel_v < V\_thresh \right)$

            set **pixel** *as black*

         else

            set **pixel** *as white*

      else

         set **pixel** *as black*

else

   for each **pixel**

      if $\left(pixel_s < \lambda_s \right) and \left(pixel_v < \lambda_v \right)$

         if $\left(pixel_v > V\_thresh \right)$

            set **pixel** *as black*

         else

            set **pixel** *as white*

      else

         set **pixel** *as black*

c) *Number and jersey has chromatic color*

$$if\ \left(\left(n_s > \lambda_s \right) and \left(n_v > \lambda_v \right) \right) and \left(\left(j_s > \lambda_s \right) and \left(j_v > \lambda_v \right) \right)$$

   for each **pixel**

      if $\left(h\_diff \left(n_h, pixel_h \right) < hue\_thresh \right)$

         set **pixel** *as black*

      else

         set **pixel** *as white*

d) *Number and jersey has achromatic color*

$$if\ \left(\left(n_s < \lambda_s \right) or \left(n_v < \lambda_v \right) \right) and \left(\left(j_s < \lambda_s \right) or \left(j_v < \lambda_v \right) \right)$$

$$if\ \left(j_v > V\_thresh \right)$$

   for each **pixel**

      if $\left(pixel_v < V\_thresh \right)$

         set **pixel** *as black*

      else

         set **pixel** *as white*

else

   for each **pixel**

      if $\left(pixel_v < V\_thresh \right)$

         set **pixel** *as white*

      else

         set **pixel** *as black*

Number and jersey color for each team are known before match. Hue, saturation and value can vary in different matches and because of this we used system initialization that can be done after observing few seconds of video. Operator selects number and jersey regions from which mean color is calculated. HSV color space is applied to a wide range of applications since it clearly separates chromatic and intensity information, but use of this color space has three drawbacks: 1) hue is meaningless when the intensity is very low 2) hue is unstable when the saturation is very low 3) saturation is meaningless when the intensity is very low. Thus, our segmentation algorithm depends on mean saturations $\left(n_s \text{ and } j_s\right)$ .and intensities $\left(j_v \text{ and } n_v\right)$ of jersey and number region.

In case a) number has achromatic color and jersey has chromatic color. In order to separate these two regions image is segmented to chromatic and achromatic area. Pixel is defined as achromatic if its saturation lies below saturation threshold λ_s or if its intensity lies below intensity threshold λ_v .

Similar to method presented in [2], these thresholds are set to value 0.2. This value proved to be good tradeoff between color loss and matching of similar achromatic colors. However, it is noticed that for some tested video clips threshold value 0.3 gives better segmentation results. Jersey region belongs to chromatic area and it is further separated from its background using hue difference. Difference threshold for hue is empirically set to $20°$. Result of this procedure is jersey represented as white object with "hole" representing player number (figure 1b). Such representation of jersey and number is possible because of large color contrast between these regions. Now it is possible to extract number regions by internal contours detection.

In case b) jersey has achromatic color and number has chromatic color. As in case a), first step is separating chromatic and achromatic image region . After that jersey is separated from background using intensity difference because hue is unstable for low saturation. Intensity threshold $\left(V\_thresh\right)$ is empirically set to 0.5
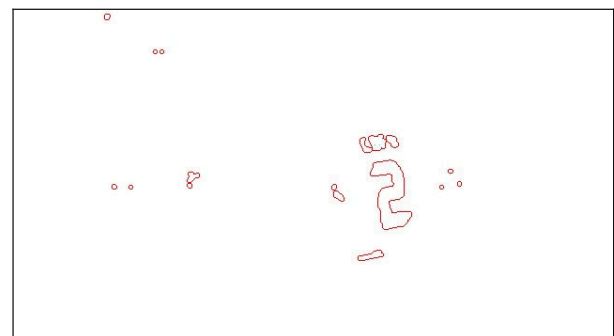
Case c) covers situation where pixels in jersey and number region lie in chromatic range. Most important cue for their separation is their hue and image is segmented using hue difference.
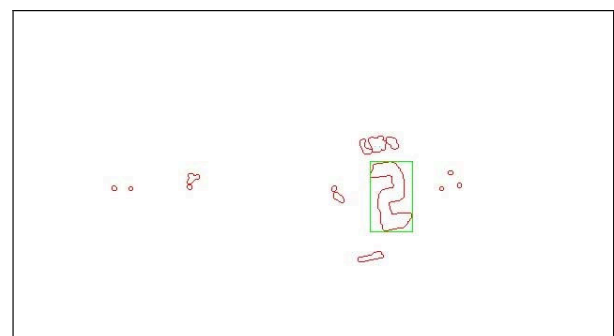

a)
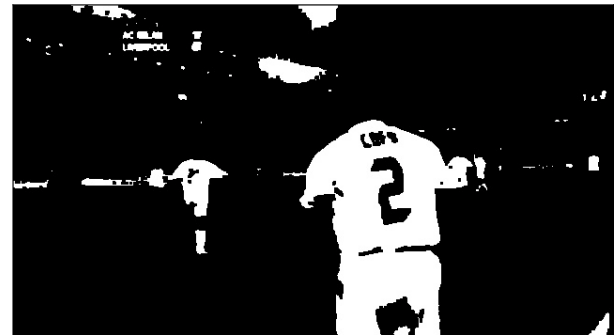
b)

c)

d)

Fig. 1 a) Original image b) Segmented image c) Internal contours d) Number candidates after area and aspect ratio filtering

As we mentioned before, in our observation set only 5 of 50 jersey examples use chromatic colors for both jersey and number

In case d) jersey and number has achromatic colors (typical example is black number on white jersey). This combination appears in 5 of 50 jersey examples used in our observation set. Cue used for their separation is intensity. These 4 cases (a,b,c,d) cover all combinations of number and jersey region saturation.

III. NUMBER REGION LOCALIZATION AND ENHANCEMENT

Image segmentation algorithm described in previous section results with binary image in which jersey is extracted as white object containing "hole", i.e. number. Although segmentation algorithm attempts to clearly separate background as black area, white objects that don't represent jerseys can also be extracted. In accordance to segmentation output we proposed simple solution for number region localization by internal contours detection (figure 1c). Internal contours in fact represent "holes" in objects. Our segmentation algorithm attempts to represent numbers as "holes" in white objects (jerseys). Hence, simple way to localize numbers is internal contour detection.

Internal contours are found using algorithm implemented in OpenCv library [9]. This algorithm firstly breaks binary image into several non-overlapped 1 (0) 4-connected (8-connected) components. Each set consists of equal pixels with equal values, that is all pixels in set are equal to 1 (white) or 0 (black). Since 0-components are complementary to 1-components, algorithm considers 1-components only and their topological structure is processed. A 0-component surrounded by 1-component is called the hole of the 1-component. The border point of 1-component could be any component pixel that has a 4-connected 0-pixel. Connected set of border points forms border (contour). Algorithm makes single pass through the image using line-by-line scanning. Whenever it finds a point belonging to new border the border following procedure is started to retrieve and store the border in chain format.

In order to filter out non-number candidates bounding rectangles are framed around each of them. Contours that don't represent numbers are rejected using area and aspect ratio of bounding frame. Area is useful in eliminating small holes produced by segmentation process.
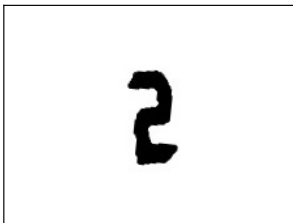

Fig. 2 Number region cut from segmented image

Since it is assumed that player number often occurs in close-up shots we set area threshold to 0.02% of image area. Number bounding rectangle has characteristic aspect ratio where height is larger than width. According to this fact height/width ratio threshold is set to 1.1. Finally remaining bounding rectangles (figure 1d) are used to extract number regions from segmented image (figure 2).

Enhancement of extracted number regions is required before OCR processing. To solve rotation problem we used direction property of number region. This kind of region descriptor makes sense in elongated regions only. Elongation is defined as ratio between the length and width of the region minimum bounding rectangle and most often number region can be considered as elongated region. If the shape moments are known, direction (theta) can be computed as defined in [10]:

$$\theta = \frac{1}{2} * \tan^{-1}\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right)$$

where μ_{11}, μ_{20} and μ_{02} are central contour moments. Then number region is rotated θ angle. Except rotation normalization number regions are also enhanced using median filtering.

Two digit numbers need additional processing. They are represented as two numbers with spatially adjacent bounding rectangles. In order to solve such cases we check distance between rectangles. If this distance is low enough, two rectangles are framed together as new rectangle representing two digit number.

IV. RESULTS

We prepared dataset containing 75 frames captured from 4 different soccer clips in 640x480 and 640x352 resolution. Each frame shows at least one jersey number that can be in various conditions including player moving, illumination change, complex background etc. Two measures are used for method evaluation: localization rate and recognition rate. Number localization is process of determining the location of number in image and generating bounding boxes around number. Localization rate is defined as:

$$LR = \frac{Number\ of\ correctly\ localized\ numbers}{Number\ of\ player\ numbers}$$

Number localization output is input for OCR software. Recognition rate is defined as:

$$RR = \frac{Number\ of\ correctly\ recognized\ numbers}{Number\ of\ player\ numbers}$$

Results for 4 different teams are shown in table 1.

TABLE I
PLAYER NUMBER LOCALIZATION AND RECOGNITION RATES

Team	Frames	Localized	Recognized	LR	RR
A	14	12	7	85%	50%
B	15	10	8	66%	53%
C	22	20	14	90%	63%
D	24	20	10	83%	42%
All	75	62	39	83%	52%

It can be seen that localization rate (83%) is significantly higher than recognition rate (52%). This difference is caused by sensitivity of OCR software to non-rigid deformation, noisy character borders etc. Possible solution is training own number recognizer as in [4]. Errors in number localization occur due to skewed numbers, folded jerseys and especially blur which is caused by player or camera motion. Video resolution and quality also affect segmentation result.

We found 4 papers dealing with player number detection and recognition ([3],[4],[6],[7]). Regarding performance measure direct comparison is possible with [4] and [6]. Our approach achieves higher localization rate and similar recognition rate like method described in [4] (LR=67% and RR=55%). In comparison with [6] (LR=80% and RR=67%) our method has weaker recognition performance and similar localization rate, but we must emphasize that we didn't use temporal consistency across the sequence of video frames as it is used in [4] and [6]. Namely, in [4] and [6] it is required that recognition is stable for a minimum number of frames. We believe that integration of temporal information would improve recognition rate.

V. CONCLUSION

A method for localization and recognition of player numbers in soccer video is proposed in this paper. We introduce new image segmentation method based on observation showing that most often combination of low and high saturated pixels is used to separate number and jersey region. Because of large color contrast between these two regions jersey is represented as white object with "hole" (number). This allows us to propose novel method for number region localization based on internal contours. Non-number regions are filtered out using area and aspect ratio of their bounding rectangles. Before OCR processing numbers are enhanced using rotation normalization and median filtering.

Our method achieves localization rate of 83% and recognition rate of 52%. These rates are better or comparable with other results in available references. In future work we will try to improve recognition performance using a temporal redundancy.

ACKNOWLEDGMENTS

REFERENCES

[1] A. Kokaram et al., "Browsing sports video", *Signal Processing Magazine,* IEEE, Vol 23, issue 2, pp. 47-58, March 2006.

[2] K. Jung, K. I. Kim and A. K. Jain, "Text information extraction in images and video: a survey", *Pattern Recognition*,Volume 37, number 5, pp. 977–997, May 2004.

[3] Q. Ye, Q. Huang and S. Jang. "Jersey number detection in sports video for athlete identification.", *Proc. of VCIP*, July 2005.

[4] M. Bertini, A. D. Bimbo and W. Nunziati, "Player identification in soccer videos", MIR '05: *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*, pp. 25-32, 2005.

[5] P. Viola and M. Jones. "Rapid object detection using a boosted cascade of simple features", *Proceedings of CVPR*, pp. 511-518, 2001.

[6] M. Bertini, A. D. Bimbo and W. Nunziati, "Matching faces with textual cues in soccer videos", Multimedia and Expo, 2006 *IEEE International Conference on Volume* , Issue , 9-12, pp. 537 – 540, July 2006.

[7] E. L. Andrade Neto, E. Khan, J. C. Woods and M. Ghanbari, "Player classification in interactive sport scenes using prior information region space analysis and number recognition", *IEEE International Conference on Image Processing (ICIP) 2003*, Barcelona, Spain, vol. III, pp. 129-132, September 2003.

[8] M. Bertini et al., "Sports video annotation using enhanced hsv histograms in multimedia ontologies", *Image Analysis and Processing Workshops, 2007, ICIAPW 2007, 14th International Conference on*, pp. 160-170, Sept. 2007.

[9] Open Source Computer Vision Library [Online]. Available: www.intel.com/technology/computing/opencv/

[10] M. Sonka, V. Hlavac and M. Boyle, *Image Processing, Analysis, and Machine Vision*, Thomson-Engineering; 3 edition, 2007.