Active Contour based Automatic Feedback for Optical Character Recognition

Joanna Isabelle Olszewska
School of Computing and Technology, University of Gloucestershire, The Park, Cheltenham, GL50 2RH, U.K.

Keywords: Active Contours, Multi-feature Vector Flow, Tracking, Optical Character Recognition, Pattern Recognition, Unsupervised Segmentation, Object Detection, Team Sport Video Analysis, Automated Scene Understanding.

Abstract: In this paper, we present a new optical character recognition approach. Our method combines chromaticity-based character detection with active contour segmentation in order to robustly extract optical characters from real-world images and videos. The detected character is recognized using template matching. Our developed approach has shown excellent results when applied to the automatic identification of team players from online datasets and is more efficient than the state-of-the-art methods.

1 INTRODUCTION

Automatic scene understanding of team sports (Olszewska and McCluskey, 2011) is essential for sport events' refereeing and analysis and it involves vision-based technologies such as object detection (Alqaisi et al., 2012), object recognition (Olszewska, 2012a), tracking (Olszewska, 2012b), or spatio-temporal reasoning (Olszewska, 2011).

In particular, automatic identification of team players is of prime importance to support both sport comments production and media archiving (Alsuqayhi and Olszewska, 2013).

For that, face recognition techniques such as (Wood and Olszewska, 2012) have been applied to process soccer games. However, this biometric approach is intrinsically not adapted to identify a player whose back is turned to the camera, in which case his face is poorly or not visible at all.

As a result, optical character recognition (OCR) methods have been developed to recognize numbers on team player's uniform. Most of them exploit the temporal redundancy of a character across several frames and thus are only limited to video analysis (Andrade et al., 2003), (Kokaram et al., 2006), (D'Orazio and Leo, 2010), (Huang et al., 2006), (Ekin et al., 2003), (Niu et al., 2008) and not suited for tasks such as still image dataset retrieval. Other works use both facial and textual cues (Bertini et al., 2006), but their computational speed is low.

Hence, in this work, we focus on the sport scene analysis based on the automatic player identification in images of any type, relying on the detection and recognition of the player's jersey number, and therefore, on the development of a full, efficient optical character recognition (OCR) system for this purpose.

OCR major phases are (i) character extraction and (ii) character recognition. In the first step, the system localizes and extracts the character by detecting its geometrical features like edges or color features, or both (Lin and Huang, 2007). In the second step, character recognition is usually performed by matching (Guanglin and Yali, 2010) or by using classifiers, e.g. AdaBoost (Chen and Yuille, 2004). However, these existing OCR systems are mainly applied to recognize license plate numbers or handwritten characters, whereas player number recognition presents additional challenges. Indeed, the foreground, i.e. the character, could be highly skewed with respect to the camera, or the background, i.e. the jersey, could be folded so that part of the number could be hidden. Moreover, sport images are often blurred, since cameras or players or both are quickly moving.

In this paper, we propose to automatically extract characters from images based on their local properties such as their pixel chromaticity and relying on their global properties processed by the active contours, while we use a digit template to recognize the extracted characters, leading to an OCR system robustly dealing with sport applications, while being computationally effective.

No temporal redundancy assumption is made in our method, which is thus valid not only for video frames, but also for still images such as those con-

Olszewska J..
Active Contour based Automatic Feedback for Optical Character Recognition.
DOI: 10.5220/0004935603180324
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (MPBS-2014), pages 318-324
ISBN: 978-989-758-011-6

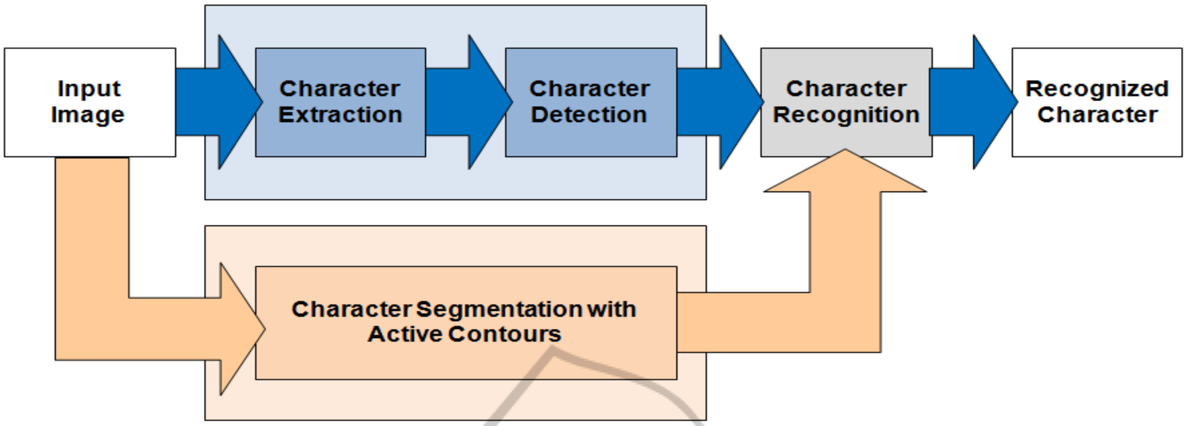

Figure 1: Our Optical Character Recognition system's architecture.

tained in sport datasets or on Internet.

In our approach, players could be identify even in back profile, since our OCR system detects and recognize characters which could be anywhere on the team player's clothes.

Hence, the contribution of this paper is:

- the use of active contours as an automatic feedback for the chromatic/achromatic segmentation approach in order to extract characters robustly;
- the development of a new powerful OCR system based on the association of this automatic feedback for character detection with the template matching-based technique for the fast character recognition, in context of the automated identification of team players in online image and videos.

The paper is structured as follows. In Section 2, we describe our optical character recognition approach for fast and effective number extraction and identification. Our method has been successfully applied to soccer players' real-world image datasets as reported and discussed in Section 3. Conclusions are presented in Section 4.

2 CHARACTER RECOGNITION AND IDENTIFICATION SYSTEM

In this section, we present our optical character recognition approach (Fig. 1) for the reliable identification of soccer player's numbers present in real-world images and videos. Firstly, the studied image is segmented by both chromaticity-based approach and active contour approach, as explained in Section 2.1. Finally, the extracted character is recognized by means of template matching described in Section 2.2.

2.1 Character Extraction

Character extraction consists here in image segmentation and character detection. On one hand, the image is binarized based on chromaticity properties of the foreground and background pixels as described in Section 2.1.1. Next, the characters' inner boundary tracing algorithm is applied in order to extract the numbers as presented in Section 2.1.2. On the other hand, active contours are processed and then, they delineate the boundaries of the character under investigation as explained in Section2.1.3. Hence, this later approach gives a feedback on the first processed extraction, leading to a more robust character detection.

Algorithm 1: Achromatic-color Number & Achromatic-color Jersey

if $((N_S < Y_S$ or $N_V < Y_V)$ and $(J_S < Y_S$ or $J_V < Y_V))$ **then**
  if $J_V > V_{thresh}$ **then**
    for all P do
      if $P_V < V_{thresh}$ **then**
        $I_B(\mathbf{P}) = 0$ ▷ set pixel as black
      else
        $I_B(\mathbf{P}) = 1$ ▷ set pixel as white
      end if
    end for
  else
    for all P do
      if $P_V < V_{thresh}$ **then**
        $I_B(\mathbf{P}) = 1$ ▷ set pixel as white
      else
        $I_B(\mathbf{P}) = 0$ ▷ set pixel as black
      end if
    end for
  end if
end if
return I_B

Algorithm 2 : Achromatic-color Number & Chromatic-color Jersey.

```
if ((N_S < Y_S or N_V < Y_V) and (J_S > Y_S and J_V > Y_V)) then
    for all P do
        if ((P_S < Y_S) and (P_V < Y_V)) then
            I_B(P) = 0                  ▷ set pixel as black
        else
            if (h_diff(J_H, P_H) < H_thresh) then
                I_B(P) = 1              ▷ set pixel as white
            else
                I_B(P) = 0              ▷ set pixel as black
            end if
        end if
    end for
end if
return I_B
```

Algorithm 3 : Chromatic-color Number & Achromatic-color Jersey.

```
if ((J_S < Y_S or J_V < Y_V) and (N_S > Y_S and N_V > Y_V)) then
    if J_V > V_thresh then
        for all P do
            if ((P_S < Y_S) and (P_V < Y_V)) then
                if P_V < V_thresh then
                    I_B(P) = 0          ▷ set pixel as black
                else
                    I_B(P) = 1          ▷ set pixel as white
                end if
            else
                I_B(P) = 0              ▷ set pixel as black
            end if
        end for
    else
        for all P do
            if (P_S < Y_S and P_V < Y_V) then
                if P_V > V_thresh then
                    I_B(P) = 0          ▷ set pixel as black
                else
                    I_B(P) = 1          ▷ set pixel as white
                end if
            else
                I_B(P) = 0              ▷ set pixel as black
            end if
        end for
    end if
end if
return I_B
```

2.1.1 Image Segmentation

Let us consider a color image I, where M and N are its width and height, respectively. The first step to extract numbers or foregrounds of this still image is to separate them from their background. In fact, in football, players' number color is chosen by the football league to be in contrast with players' kit (shirt and sweater), in order to allow visibility of the number in diverse conditions. The study of (Saric et al., 2008) has found that this contrast is the most important in the hue, saturation and value (HSV) color space when looking at the saturation of the number pixels and the jersey pixels. Consequently, the image I could be segmented based on the low and high saturated pixels, i.e. objects' achromatic and chromatic colors, respectively, leading to a binary image I_B. In particular, a color pixel under investigation $\mathbf{P} = [P_H, P_S, P_V]$ is considered as achromatic if its saturation (P_S) is below the saturation threshold (Y_S) or if its intensity (P_V) is below intensity threshold (Y_V). If the pixel saturation and intensity are above these thresholds, then it is considered as chromatic.

Algorithm 4: Chromatic-color Number & Chromatic-color Jersey.

```
if ((N_S > Y_S and N_V > Y_V) and (J_S > Y_S and J_V > Y_V)) then
    for all P do
        if (h_diff(N_H, P_H) < H_thresh) then
            I_B(P) = 0                  ▷ set pixel as black
        else
            I_B(P) = 1                  ▷ set pixel as white
        end if
    end for
end if
return I_B
```

The segmentation is initialized by defining the mean color vector of the jersey $\mathbf{J} = [J_H, J_S, J_V]$ and the mean color vector for the number $\mathbf{N} = [N_H, N_S, N_V]$, based on provided image samples. Next, the image I is processed depending if the number color is chromatic or achromatic and if the jersey color is chromatic or achromatic, leading to four cases, i.e. to four Algorithms 1-4. The segmentation is based on the hue threshold H_{thresh} and the hue difference in the case of a chromatic-color jersey, whereas the intensity difference and the intensity threshold V_{thresh} are used in the case of an achromatic-color jersey (Saric et al., 2008). In the case where the number has an achromatic color and the jersey color is chromatic (Algorithm 2), the hue difference h_{diff} is defined as follows:

$$h_{diff}(J_H, P_H) = \begin{cases} \Delta(J_H, P_H) & \text{if} \Delta(J_H, P_H) < 180^\circ, \\ 360^\circ - \Delta(J_H, P_H) & \text{otherwise,} \end{cases} \quad (1)$$

with

$$\Delta(J_H, P_H) = \mid J_H - P_H \mid . \quad (2)$$

When both the jersey and the number have chromatic colors, the image is segmented as described in

Algorithm 4, using the hue difference h_{diff} defined as follows:

$$h_{diff}(N_H,P_H) = \begin{cases} \Delta(N_H,P_H) & \text{if}\Delta(N_H,P_H) < 180^\circ, \\ 360^\circ - \Delta(N_H,P_H) & \text{otherwise,} \end{cases} \tag{3}$$

with

$$\Delta(N_H,P_H) = \mid N_H - P_H \mid . \tag{4}$$

2.1.2 Character Detection

In the binarized image I_B computed by the process explained in Section 2.1.1, jerseys appear as white objects, while numbers as black ones. Based on that fact, tracing internal boundaries of these objects is an efficient method for number region localization and extraction. For this purpose, we have adapted the Boundary Tracing approach (Ren et al., 2002). Hence, our process presented in Algorithm 5 initiates by tracing all the boundaries B_i within the segmented binary image, and then, in relation to the specific area aspect ratio F characterizing the number region, the boundaries are filtered, in order to select only those containing numbers. Once this process is completed, the binary image I_B is cropped and the cropped image I_C is transferred to the recognition stage which then identifies the numbers as detailed in Section 2.2.

This section has presented the single digit case. The identification of two-digit numbers is as follows. If two cropped images are of the same size and are in adjacent bounding rectangles, they are flagged as forming a two-digit number.

Algorithm 5: Boundary Tracing.

Step 1
Find boundaries $B = \{B_i\}$ of all objects

Step 2
for all B_i **do**
  if B_i of black object **then**
    if B_i dimensions = F dimensions **then**
      $x_1 = \min(B_i[1])$
      $y_1 = \min(B_i[2])$
      $x_2 = \min(B_i[1])$
      $y_2 = \max(B_i[2])$
      $I_C = I_B[x_1 : x_2][y_1 : y_2]$
    else
      Ignore boundary
      Go the next boundary
    end if
  end if
end for
return I_C

2.1.3 Active Contours

In this work, multi-feature vector flow active contours are used to provide another character segmentation in order to have a feedback on the results computed in Sections 2.1.1-2.1.2.

Indeed, multi-feature active contour is a parametric planar deformable curve $\mathcal{C}(s) = [\mathcal{C}_x(s), \mathcal{C}_y(s)]$, with $0 \leq s \leq 1$, which evolves from an initial position to object boundaries with the use of the innovative MFVF field $\Xi(x,y) = [\xi_u(x,y), \xi_v(x,y)]$.

In this framework, the convergence of the curve is guided by internal and external forces, which are involved in a gradient descent process. The internal forces constrain the active contour shape, in the way to ensure regularity and smoothness of the curvature. MFVF external force regroups all the selected features in one original bidirectional force, enabling the active contour to reach its final accurate position, even in complex situations.

Formally, the deformable curve $\mathcal{C}(s,t)$ is modeled itself by a B-spline paradigm in order to be computationally efficient, and must satisfy the following dynamic equations,

$$\mathcal{C}_{xt}(s,t) = \alpha\, \mathcal{C}_x''(s,t) - \beta\, \mathcal{C}_x''''(s,t) + \xi_u(x,y) \tag{5}$$

$$\mathcal{C}_{yt}(s,t) = \alpha\, \mathcal{C}_y''(s,t) - \beta\, \mathcal{C}_y''''(s,t) + \xi_v(x,y), \tag{6}$$

where $\mathcal{C}_x''$, $\mathcal{C}_y''$, $\mathcal{C}_x''''$, $\mathcal{C}_y''''$, respectively, are the second and fourth-order derivatives with respect to the parameter s of the curve, α is the curvature elasticity coefficient, and β is the curvature rigidity coefficient.

The active contour, found by solving (5) and (6), could be, in practice, roughly initialized from a distance of the target, as the MFVF force offers a large capture range. This obtained fast multi-feature active contour owns high-deformation capabilities that are well suited for tracking non-rigid objects whose shapes change markedly. Indeed, tracking with the multi-feature active contour could be performed by minimizing the associated energy functional, for each corresponding feature, in each frame.

Moreover, this computed curve enables precise foreground segmentation, without any kind of assumption about the object appearance (Olszewska, 2012b).

2.2 Character Recognition

For the recognition of the characters extracted either with the chromaticity-based technique or active contour method, we have adopted template matching approach. Indeed, this pattern classification method is well suited in the identification of small regions (Brunelli, 2009), which is the case in our application.

Figure 2: Examples of results obtained with our OCR system. First column: input image. Second column: chromaticity-based segmentation. Third column: active contour-based segmentation. Fourth column: recognized character.

The basis of template matching is that a processed image is compared to each of the images stored within a template. In many instances, the extracted number region has smaller or larger dimensions compared to the template dimensions, or has not the same orientation. Thus, the extracted number image has first to be rotated and rescaled to fit the template orientation and size, respectively. Then, the correlation coefficient r between the two compared images is computed as follows:

$$r = \frac{\sum_m \sum_n (T_{mn} - \bar{T})(S_{mn} - \bar{S})}{\sqrt{\left[\sum_m \sum_n (T_{mn} - \bar{T})^2\right]\left[\sum_m \sum_n (S_{mn} - \bar{S})^2\right]}}, \quad (7)$$

where T_{mn} are the values of the pixels of the template image with an $m \times n$ size and a mean $\bar{T}$; S_{mn} are the values of the pixels of the processed image, i.e. the rescaled cropped binarized image, with a mean $\bar{S}$.

When the structure of the processed image is greatly similar to the structure of one of the template images, then the correlation coefficient value is high and this means the number is identified.

A digit is considered as recognized when at least one segmentation technique has succeed to extract it and when the template matching has processed successfully and coherently. In case when the two types of segmentation followed each by template matching provide different results, the digit is flagged as unrecognized.

To recognize two-digit numbers, single numbers flagged as constituting a two-digit number in Section 2.1 are recognized individually by matching each of them with the template. The two-digit number is then formed based on that information.

We can notice that the use of the template matching technique is well suited for our system of automatic number recognition of soccer players. On one hand, template matching is particularly fast when used in context of our system, because it requires only the recognition of numerical characters, rather than a wider range of alphanumerical characters as in other applications, such as license plate recognition (LPR). Indeed, our template stores in total only 10 images of one-digit numbers (0 to 9). Hence, the matching is performed against a maximum of ten stored images, in order to recognize the extracted character, which is computationally very efficient. Moreover, the scale sensitivity of the template matching technique is used

Table 1: Average rates of the automatic character extraction and the automatic character recognition obtained for all the dataset using approaches of ◇(Bertini et al., 2006), □(Saric et al., 2008), △ (Alsuqayhi and Olszewska, 2013), and our.

Rate	◇	□	△	**our**
character extraction rate	80.0%	83.0%	88.0%	**95.0%**
character recognition rate	67.5%	52.0%	86.0%	**90.0%**

in our work as an advantage, since smaller dimensions of the template dimensions lead to a faster matching. On the other hand, the recognition rate obtained by our implementation of this method in our system is much higher than those presented in the literature as discussed in Section 3.

3 RESULTS EVALUATION AND DISCUSSION

To validate our method, we have carried out experiments which consist in automatically recognizing numbers from the soccer players' jerseys within a database containing data images with soccer-related content, as such illustrated in Fig. 2.

For this purpose, our system has been applied on a dataset containing 4500 football images whose average resolution is of 230x330 pixels and which were captured in outdoor environment. This database owns challenges of quantity, pose and scale variations of the players. Moreover, the colors of the teams' uniforms have various colors and the fonts on the players' jerseys could vary strongly.

All the experiments have been run on a computer with an Intel(R) Core(TM)2 Duo 2.53 GHz processor, 4 Gb RAM, and using our OCR software implemented with MatLab. Our system is able to support different types of image formats such as jpeg, tiff, bmp, and png.

In order to assess the performance of our OCR system, we use the following criteria:

$$extraction\ rate = \frac{CL \times 100}{TT}, \tag{8}$$

$$recognition\ rate = \frac{CR \times 100}{TT}, \tag{9}$$

with CL, the number of correctly localized characters, CR, the number of correctly recognized characters, and TT, the total number of tested characters.

Some examples of the results of our OCR system are presented in Fig. 2. These samples present difficult situations such as variability of the jerseys' colors, i.e. different pixels' chromaticity properties of the foregrounds and the backgrounds; numbers' changing characteristics, i.e. different characters' geometrical and spatial properties; scale effects such as zoom out or close-up.

We can observe that using our approach, characters are correctly extracted and correctly recognized, despite their geometrical and chromatical differences. Hence, our OCR system is robust towards changes in numbers and colors of the foregrounds and the backgrounds as well as towards variations of fonts, size, and orientation of the characters. Moreover, the system is robust even in case the chromatic detection provides a sparse result such as in Fig. 2(f), because of the effect of the feedback provided by the active contours as displayed in Fig. 2(e).

In Table 1, we have reported the extraction and recognition rates of our OCR method against the rates achieved by approaches using chromatic/achromatic segmentation (*C/A Segm.*) or template matching (*TM*) *MSERE + TM* (Bertini et al., 2006), *C/A Segm. + CL* (Saric et al., 2008), and *(C/A Segm. + TM)* (Alsuqayhi and Olszewska, 2013).

We can see in Table 1 that our OCR method relying on the active contour feedback into the OCR process which combines chromatic/achromatic segmentation and matching-based recognition outperforms the state-of-art approaches for soccer player's number identification. In particular, we can notice than the extraction rate is improved when using the active contour as feedback for the chromatic/achromatic segmentation instead of using *C/A segm.* alone. Our OCR method outperforms also other state-of-the-art techniques such as maximally stable extremal region extraction. On the other hand, we can observe the positive effect of our active contour based automatic feedback approach on the recognition rate compared to other classification methods.

From Table 1, we can conclude also that the incorporation of the active contours increases the robustness of the OCR system. Indeed, it helps in improving the detection rate, thus the recognition rate is higher as well.

Moreover for all the dataset, the average computational speed of our combined OCR method is in the range of few seconds, and thus, our developed system could be used in context of online scene analysis.

4 CONCLUSIONS

Reliable team player identification in online data such as images and videos is a challenging topic we have copped with. For this purpose, we have developed a new OCR approach relying on both chromaticity-based segmentation and active contour method which provides a feedback to the system to reinforce the robustness of the character extraction. Template matching is used for the character recognition step. Our OCR system shows greater performance than the ones found in the literature in both extraction and recognition of soccer players' numbers. Moreover, our OCR approach is well suited for the automatic retrieval and analysis of online, visual data about team sports.

REFERENCES

Alqaisi, T., Gledhill, D., and Olszewska, J. I. (2012). Embedded double matching of local descriptors for a fast automatic recognition of real-world objects. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'12)*, pages 2385–2388.

Alsuqayhi, A. and Olszewska, J. I. (2013). Embedded double matching of local descriptors for a fast automatic recognition of real-world objects. In *Proceedings of the IAPR International Conference on Computer Analysis of Images and Patterns Workshop (CAIP'13)*, pages 139–150.

Andrade, E. L., Khan, E., Woods, J. C., and Ghanbari, M. (2003). Player classification in interactive sport scenes using prior information region space analysis and number recognition. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'03)*, pages III.129–III.132.

Bertini, M., Bimbo, A. D., and Nunziati, W. (2006). Matching faces with textual cues in soccer videos. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages 537–540.

Brunelli, R. (2009). *Template Matching Techniques in Computer Vision: Theory and Practice*. John Wiley and Sons.

Chen, X. and Yuille, A. L. (2004). Detecting and reading text in natural scenes. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages II.366–II.373.

D'Orazio, T. and Leo, M. (2010). A review of vision-based systems for soccer video analysis. *Pattern Recognition*, 43(8):2911–2926.

Ekin, A., Tekalp, M., and Mehrotra, R. (2003). Automatic soccer video analysis and summarization. *IEEE Transactions on Image Processing*, 12(7):796–806.

Guanglin, H. and Yali, G. (2010). A simple and fast method of recognizing license plate number. In *Proceedings of the IEEE International Forum on Information Technology and Applications*, pages II.23–II.26.

Huang, C.-L., Shih, H.-C., and Chao, C.-Y. (2006). Semantic analysis of soccer video using dynamic Bayesian network. *IEEE Transactions on Multimedia*, 8(4):749–760.

Kokaram, A., Rea, N., Dahyot, R., Tekalp, A. M., Bouthemy, P., Gros, P., and Sezan, I. (2006). Browsing sports video: Trends in sports-related indexing and retrieval work. *IEEE Signal Processing Magazine*, 23(2):47–58.

Lin, C.-C. and Huang, W.-H. (2007). Locating license plate based on edge features of intensity and saturation. In *Proceedings of the IEEE International Conference on Innovative Computing, Information and Control*, pages 227–230.

Niu, Z., Gao, X., Tao, D., and Li, X. (2008). Semantic video shot segmentation based on color ratio feature and SVM. In *Proceedings of the IEEE International Conference on Cyberworlds*, pages 157–162.

Olszewska, J. I. (2011). Spatio-temporal visual ontology. In *Proceedings of the EPSRC Workshop on Vision and Language*.

Olszewska, J. I. (2012a). A new approach for automatic object labeling. In *Proceedings of the EPSRC Workshop on Vision and Language*.

Olszewska, J. I. (2012b). Multi-target parametric active contours to support ontological domain representation. In *Proceedings of the RFIA Conference*, pages 779–784.

Olszewska, J. I. and McCluskey, T. L. (2011). Ontology-coupled active contours for dynamic video scene understanding. In *Proceedings of the IEEE International Conference on Intelligent Engineering Systems*, pages 369–374.

Ren, M., Yang, J., and Sun, H. (2002). Tracing boundary contours in a binary image. *Image and Vision Computing*, 20(2):125–131.

Saric, M., Dujmic, H., Papic, V., and Rozic, N. (2008). Player number localization and recognition in soccer video using HSV color space and internal contours. In *Proceedings of the World Academy of Science, Engineering and Technology*, pages 531–535.

Wood, R. and Olszewska, J. I. (2012). Lighting-variable AdaBoost based-on system for robust face detection. In *Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, pages 494–497.