EasyChair Preprint

№ 2604

Motion Recognition of Assistant Referees in Soccer Games via Selective Color Contrast Revelation

Yoonhyung Kim, Chanho Jung and Changick Kim

February 7, 2020

Motion Recognition of Assistant Referees in Soccer Games via Selective Color Contrast Revelation*

Yoonhyung Kim[1][0000−0002−5608−8473], Chanho Jung[2][0000−0003−3145−6732], and Changick Kim[1][0000−0001−9323−8488]

[1] Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea
{yhkim1127,changick}@kaist.ac.kr
[2] Hanbat National University, Daejeon, Republic of Korea
peterjung@hanbat.ac.kr

Abstract. In this paper, we propose a method to recognize motions of assistant referees in soccer games. Since the major motions of assistant referees in soccer games are closely related to the direction of pointing the flag, positional information of the flag and the upper body can be utilized as an important clue for motion recognition. Based on this observation, we propose to generate and utilize a heatmap image which reveals the contrast between colors of the flag and the upper body. For quantitative evaluation, we used K-league 2017 dataset, and our proposed method achieved 97.56% accuracy surpassing various deep learning-based classifiers. We expect that this study would be used as a useful benchmark for researchers developing sports event recognition systems.

Keywords: Motion recognition · Color contrast · Heatmap generation.

1 Introduction

Along with the rapid growth of sports industries in recent years, the amount of sports data (i.e., data which are emerged while playing sports games) has been increased accordingly. However, the procedure of collecting and refining sports data is time-consuming and costly because the task is heavily relied on manual labelling by a limited number of experts. Recently, to overcome this limitation, automatic sports analysis systems, which are equipped with state-of-the-art computer vision technologies, have been suggested as one of the most efficient alternatives. For instance, an automatic sports analysis system can be composed of the following three modules. First, locations and bounding boxes of players and referees are acquired by a visual tracking module[3],[4]. Second,

* This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Reasearch & Development Program 2016 (R2016030044, Development of Context-Based Sport Video Analysis, Summarization, and Retrieval Technologies)

Table 1. Defined motion list of soccer assistant referee

No.	Motion label	Description
1	Flag Up	Raising the flag upwards
2	Flag Left	Raising the flag to the left side
3	Flag Right	Raising the flag to the right side
4	Idle	Motions that are irrelevant to events (e.g., walk, run with letting down the flag)

motions of the objects are identified by a motion recognition module. Finally, using the data obtained by the previous two modules, sports events are extracted via an event recognition module.

This paper is focused on developing a motion recognition module for assistant referees in soccer games. Since many events of soccer games (e.g., offside, goal line out, foul) are directly associated with the motions of assistant referees, recognizing motions of assistant referees is essential for establishing a soccer event analysis system. Our objective of developing the motion recognition module is to figure out the direction of the assistant referee's flag based on the observation that the direction of pointing the flag is closely related to the events. To this end, we transform an input RGB image towards a heatmap image which reveals the contrast between colors of the flag and the upper body. For quantitative evaluation, we used K-league 2017 dataset, and our proposed motion recognition method achieved 97.56% accuracy with 600 FPS computational speed. To validate the strength of our proposed method, we conducted comparative evaluations against several representative deep learning-based images classifiers and confirmed that our approach surpasses those methods by a large margin.

2 Proposed Method

For an input RGB bounding box image, the goal of our motion recognition module is to correctly classify the motion of the assistant referee in the input image as one of the four candidate motions, which are defined in Table 1. To this end, we generate a heatmap image I_H by means of the following linear summation model:

$$I_H = w_R^* I_R + w_G^* I_G + w_B^* I_B, \tag{1}$$

where I_R, I_G, I_B are the red, green, blue channels of the input image, respectively. In (1), w_R^*, w_G^*, w_B^* are the channel coefficients, which are the parameters of the optimization model described in the next paragraph.

To determine the channel coefficients, we first collect RGB pixel intensities from the training images. To be specific, for the n-th input image, we manually extract three kinds of pixel intensities $p_n^f(C), p_n^u(C), p_n^b(C)$, which are picked from the flag, the upper body, and the background regions, respectively. Here, $C \in \{R, G, B\}$ denotes the channel. For N_t training images, our goal is to determine the optimal channel coefficients based on the optimization model which is

established as the following linear system:

$$\begin{bmatrix} p_1^f(R) & p_1^f(G) & p_1^f(B) \\ \vdots & \vdots & \vdots \\ p_{N_t}^f(R) & p_{N_t}^f(G) & p_{N_t}^f(B) \\ p_1^u(R) & p_1^u(G) & p_1^u(B) \\ \vdots & \vdots & \vdots \\ p_{N_t}^u(R) & p_{N_t}^u(G) & p_{N_t}^u(B) \\ p_1^b(R) & p_1^b(G) & p_1^b(B) \\ \vdots & \vdots & \vdots \\ p_{N_t}^b(R) & p_{N_t}^b(G) & p_{N_t}^b(B) \end{bmatrix} \begin{bmatrix} w_R \\ w_G \\ w_B \end{bmatrix} = \begin{bmatrix} c^f \\ \vdots \\ c^f \\ c^u \\ \vdots \\ c^u \\ c^b \\ \vdots \\ c^b \end{bmatrix}. \tag{2}$$

In (2), c_f, c_u, c_b denote the target intensities for pixels corresponding to the flag, the upper body, and the background regions, respectively and their default values are set as $(c_f, c_u, c_b) = (1, -1, 0)$ for all experiments. Since the above optimization model is an overdetermined linear system, the channel coefficients can be optimized by solving the least squares problem. By representing the linear system in (2) as $\mathbf{Ax} = \mathbf{b}$, the solution can be obtained as follows[8]:

$$\mathbf{x}^* = \begin{bmatrix} w_R^* \\ w_G^* \\ w_B^* \end{bmatrix} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} = \mathbf{A}^\dagger\mathbf{b}, \tag{3}$$

where $\mathbf{A}^\dagger$ indicates the pseudo inverse of the matrix $\mathbf{A}$. By means of the above optimization process, the solution is obtained as $w_R^* = 5.29$, $w_G^* = -4.17$, $w_B^* = -0.96$. Several examples of heatmap images I_H are represented in Fig. 1. As can

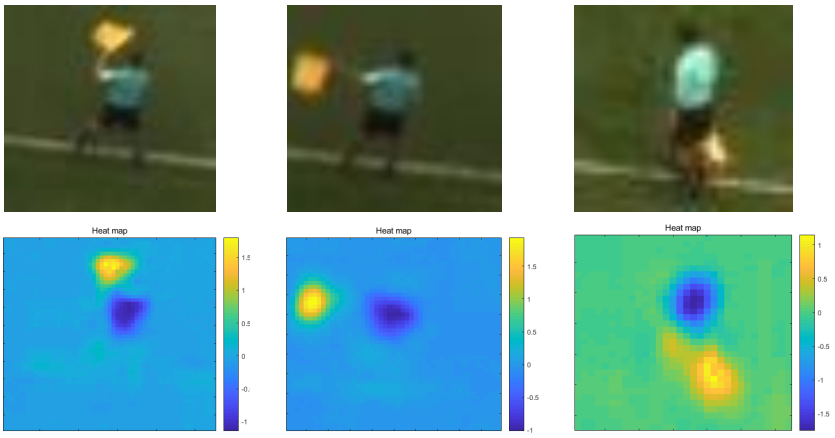

Fig. 1. Input RGB bounding box images (top) and corresponding heatmap images (bottom). Best viewed in color.

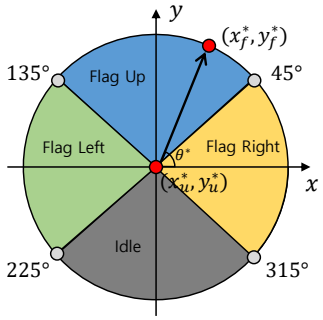

Fig. 2. Illustration of the method determining assistant referee's motion via the angular relationship of coordinates of the flag and the upper body.

be found in the figure, the heatmap intensities are high for the flag regions and low for the upper body regions. From each heatmap, we determine the pixel coordinates whose intensities are the highest and the lowest as the locations of the flag and the upper body, respectively.

By letting the coordinates of the flag and the upper body as (x_f^*, y_f^*), (x_u^*, y_u^*), respectively, the motion is recognized via the angular relationships of the two points. To this end, as illustrated in Fig. 2, the angle θ^* of the vector $v_d^* = (x_f^*, y_f^*) - (x_u^*, y_u^*)$ is calculated. The final stage of our method is to determine the motion as one of the four classes based on the angle θ^*, as depicted in Fig. 2.

Table 2. Comparative evaluation results of our proposed motion recognition method and other deep learning-based methods.

Method	AlexNet[5]	VGGNet-16[7]	ResNet-18[1]	DenseNet-121[2]	Proposed
Accuracy (%)	84.44	90.25	74.59	79.88	97.56

Table 3. Pose recognition results of basketball players (Confusion matrix)

Prediction \ GT	Flag Up	Flag Left	Flag Right	Idle	Precision (%)
Flag Up	**904**	30	0	0	96.79
Flag Left	3	**1,106**	0	43	96.01
Flag Right	38	0	**1,024**	2	96.24
Idle	9	25	3	**3,081**	98.81
Recall (%)	94.76	95.26	99.71	98.56	-

(a) (b) (c)

Fig. 3. Impact of the gamma correction on the illuminance of an input image. (a) $\gamma = 0.5$, (b) $\gamma = 1.0$, (c) $\gamma = 2.0$

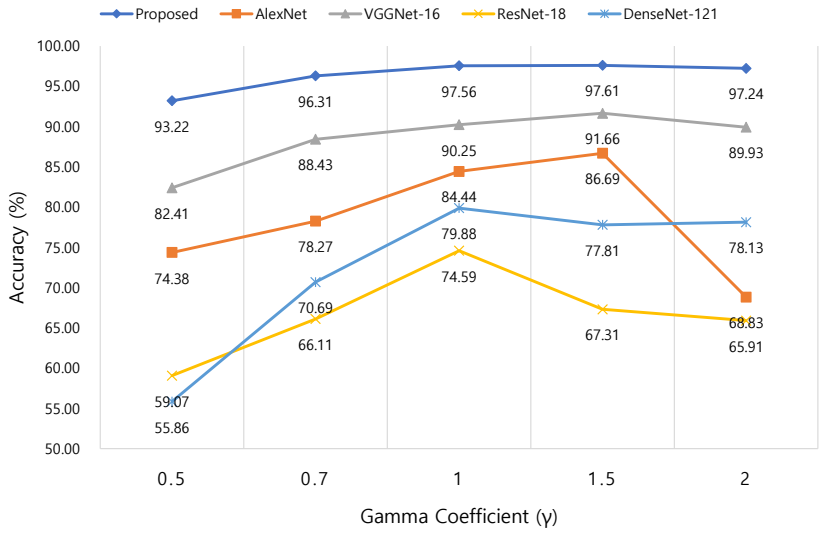

Fig. 4. Impact of the gamma correction on the average accuracies of various methods. Best viewed in color.

3 Experiments

For experiments, we used K-league (Korean-league) classic match videos in 2017. From the videos, we manually obtained 7,835 bounding box images of assistant referees. Among the images, we randomly selected 1,567 images for training and the remaining 6,268 images for validation. It is worth noting that our method does not require a large-scale training set and we found that only a small-scale training set containing approximately 10 images is sufficient for obtaining the optimal solution $\mathbf{x}^*$ in (3). But, in order to conduct a fair comparative evaluation with deep learning-based image classifiers which generally require a large number of training images, we assigned sufficient amount of images for train-

ing. For comparison, we adopted AlexNet[5], VGGNet-16[7], ResNet-18[1], and DenseNet-121[2], which are the representative deep learning-based image classification models. To train the deep models, we used PyTorch[6] implementations and a single NVIDIA Titan-X GPU. The learning rate is 0.001 and is multiplied by the factor of 0.2 for every 20 epochs. The total number of epoch, the batch size, and the momentum rate are 50, 32, and 0.9, respectively. The quantitative evaluation results are given in Table 2. As we can see, our proposed method shows the best accuracy surpassing other deep learning-based models. The confusion matrix of our method is given in Table 3.

Since soccer games are played in the outdoor spaces, motion recognition modules need to be robust to variations of illuminance caused by weather conditions. Based on this observation, we further conducted additional experiments to verify the robustness to illuminance variations. Specifically, as depicted in Fig. 3, we intentionally applied gamma corrections with four different gamma coefficients $\gamma \in \{0.5, 0.7, 1.5, 2.0\}$ to the validation images to simulate various illuminance conditions. As can be seen in Fig. 4, our method is more robust to illuminance variations than other deep learning-based methods. The computational speed of our method is around 600 FPS on an ordinary PC, which is equipped with a 4.0 GHz Intel i7 CPU and a 8GB RAM.

4 Conclusion

In this paper, we have proposed a motion recognition method for assistant referees in soccer games. The key motivation of our method is to generate a heatmap image which reveals the contrast between colors of the flag and the upper body for localizing the coordinates of the two parts. Our proposed method achieved 97.56% accuracy with real-time operation surpassing other deep learning-based methods. We expect that our work would be a useful and practical benchmark for researchers who are interested in developing automatic sports analysis systems.

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition pp. 770–778 (2016)
2. Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., Keutzer, K.: Densenet: Implementing efficient convnet descriptor pyramids. arXiv preprint arXiv:1404.1869 (2014)
3. Kim, W.: Multiple object tracking in soccer videos using topographic surface analysis. Journal of Visual Communication and Image Representation **65**, 102683 (2019)
4. Kim, W., Moon, S.W., Lee, J., Nam, D.W., Jung, C.: Multiple player tracking in soccer videos: an adaptive multiscale sampling approach. Multimedia Systems **24**(6), 611–623 (2018)
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Neural Information Processing Systems (NeurIPS) (2012)

6. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)
7. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
8. Watkins, D.S.: Fundamentals of matrix computations, vol. 64. John Wiley & Sons (2004)