Vision-based Detection of Humans on the Ground from Actual Aerial Images by Informed Filters using Only Color Features

Takuro Oki[1], Risako Aoki[1], Shingo Kobayashi[1], Ryusuke Miyamoto[2],
Hiroyuki Yomo[3] and Shinsuke Hara[4]

[1]*Department of Computer Science, Graduate School of Science and Technology, Japan*
[2]*Department of Computer Science, School of Science and Technology,*
Meiji University, 1-1-1 Higashimita, Tama-ku, Kawasaki-shi, Japan
[3]*Department of Electrical and Electronic Engineering, Faculty of Engineering Science,*
Kansai University, 3-3-35 Yamate-cho, Suita-shi, Japan
[4]*Graduate School of Engineering, Osaka City University,*
3-3-138 Sugimoto Sumiyoshi-ku, Osaka-shi, Japan

Keywords: Player Detection, Actual Aerial Images, Informed-filters.

Abstract: In this paper, we construct a novel sensor network system that can measure real-time vital signs of the human body during exercise even at high speeds and in crowded regions. The sensor network estimates locations of sensor nodes using image processing to extract locations of humans wearing sensors. This paper evaluates the accuracy of human detection by informed-filters using only color features for actual aerial images. To carry out the evaluation, a novel dataset composed of actual images captured using a camera mounted on a drone was created. Experimental results show that significant accuracy can be achieved by the detector. In addition, the number of weak classifiers in a strong classifier can be reduced to 125 without significant degradation of the detection accuracy.

1 INTRODUCTION

Real-time sensing of the vital signs of the human body during exercise can improve the effectiveness of exercise and prevention of sudden illness. Some researchers have attempted to construct a novel sensor network system (Hara et al., 2017). To construct such a sensor network. the most important challenge is the location estimation of sensor nodes. In this application, widely used routing schemes based on Received signal-strength indicator (RSSI) or global positioning system (GPS) do not work well because sensor nodes attached to exercisers moves quickly and their density often becomes prohibitively high.

To solve this problem, image-assisted routing that estimates locations of sensor nodes using image processing was proposed as a solution (Miyamoto and Oki, 2016). Image-assisted routing uses images captured from a few cameras mounted on drones and fixed tripods as input. Then, vision-based human detection and tracking processes are applied to obtain the location of humans wearing sensor nodes. In this process, human detection using aerial images captured using drones is a challenging problem because of the required high accuracy of operation on embedded systems.

Previous research in (Oki et al., 2019) evaluated the accuracy of player detection from aerial images using a computer graphics (CG) dataset that was created from actual locations of players manually annotated using a video sequence of a soccer game held in an indoor soccer field (Miyamoto et al., 2017). In (Oki et al., 2019), several drone locations were evaluated in terms of their detection accuracy to find good positions of a drone to record sports scenes. Results showed that for best player detection accuracy, the drone should be located where the angle between the view direction and the ground only approaches 90 degrees. The detector used in the evaluation was a classifier constructed with informed-filters using only color features showing better accuracy than state-of-the-art schemes based on deep neural networks (Redmon and Farhadi, 2017; Girshick et al., 2014; Zoph et al., 2018) in the CG dataset (Miyamoto et al., 2019). Considering the detection accuracy and processing speed on GPUs (Oki and Miyamoto, 2017),

Oki, T., Aoki, R., Kobayashi, S., Miyamoto, R., Yomo, H. and Hara, S.
Vision-based Detection of Humans on the Ground from Actual Aerial Images by Informed Filters using Only Color Features.
DOI: 10.5220/0008166000840089
In *Proceedings of the 7th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2019)*, pages 84-89
ISBN: 978-989-758-383-4

this scheme is practical for this application.

Previous research has shown that a classifier based on informed-filters using only color features works well in aerial images, but detection accuracy using actual images recorded by a camera mounted on a drone has not been evaluated. To verify the detection performance of the informed-filters using actual aerial images, this paper constructs an actual dataset using aerial images and evaluates the detection accuracy of the informed-filters. In the evaluation, the relationship between the detection accuracy and the number of weak classifiers in a strong classifier is investigated in order to determine the possibility of reducing the computational cost while maintaining the detection accuracy.

The rest of this paper is organized as follows. Section Ⅱ summarizes the evaluation of detection accuracy at several view points using the CG dataset in (Oki et al., 2019) and Section Ⅲ discusses how to construct a novel dataset using actual aerial images of humans during exercise. Section Ⅳ discusses how to train a detector and reduce the number of weak classifiers. Section Ⅴ evaluates the detection accuracy using the new dataset and the paper is concluded in section Ⅵ.

2 ACCURACY OF PLAYER DETECTION IN AERIAL IMAGES USING THE CG DATASET

This section summarizes the previous research in (Oki et al., 2019) that evaluates the detection accuracy from aerial images using the CG dataset.

2.1 The CG Dataset

The CG dataset was created to evaluate the detection accuracy of humans during exercise at several viewpoints in aerial images (Miyamoto et al., 2019). In the dataset, UnityChan, which is a freely usable 3D model of a character, is used to represent humans on the soccer field. The locations of the 3D characters correspond to the locations of humans determined manually from an actual image sequence during exercise. The dataset is constructed as a virtual 3D space using the Unity 3D engine so that we can easily generate two-dimensional images from arbitrary viewpoints. Fig. 1 shows examples of the CG dataset corresponding to several viewpoints. Fig. 2 and 3 show examples of positive and negative samples of the CG dataset.

2.2 Detection Accuracy for Several Viewpoints

Fig. 4 shows the heat map representing accuracy of detection at many viewpoints. The deep red color regions correspond to a lower miss rate but the miss rate is greater where the red color is lighter. When the angle between the view direction and the soccer field (or ground) approaches 90 degrees, the detection accuracy is maximized. However, when the drone is located right above the center of the soccer field, the detection accuracy is poor. Based on these simulation results, the drone is located at moderate positions as described in the following section.

3 A DATASET COMPOSED OF ACTUAL IMAGES CAPTURED FROM A CAMERA MOUNTED ON A DRONE

This section details a novel dataset created using actual images recorded from a camera mounted on a drone. The dataset is used to evaluate the accuracy of human detection during exercise.

3.1 An Experimental Setup for Capturing Aerial Images using a Drone

Fig. 5 shows the experimental setup used to capture aerial images using a drone. Here, a mini game of soccer was performed by eight humans in the soccer field. The filed is marked by white lines on the ground and it has a dirt surface. DJI Phantom 4, shown in Fig. 6, was used to capture the aerial images. It can record 3840×2160 video sequences.

3.2 How to Create Ground Truth

To create ground truth for the recorded image sequences, we developed a novel software named QtKukeiKakuKun using python and Qt. Fig. 7 shows a screenshot of QtKukeiKakuKun. The software can initialize the ground truth of the current working frame by the ground truth of the previous frame. This function can drastically reduce the human load when ground truth is created from images corresponding to successive frames of a video sequence. In addition, the software was designed to enable easy cooperation because annotations about ground truth were usually created by many people. Using QtKukeiKakuKun,

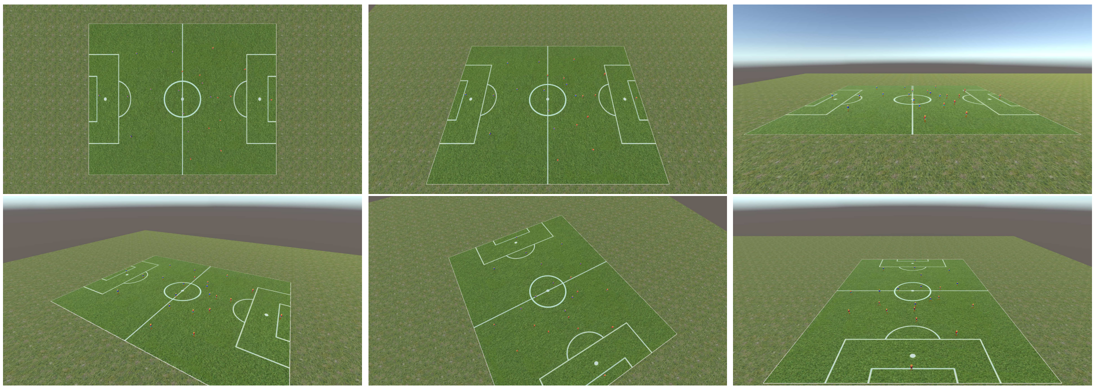

Figure 1: Examples of the CG dataset created in previous research(Oki et al., 2019).

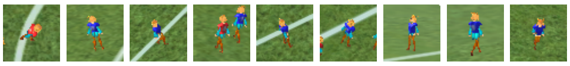

Figure 2: Examples of positive samples in the CG dataset.

Figure 3: Examples of negative samples in the CG dataset.

we created a novel dataset composed of 5000 images. Fig. 8 and 9 show the positive and negative samples included in the novel dataset.

4 CONSTRUCTION OF A CLASSIFIER FOR PLAYER DETECTION IN ACTUAL IMAGES

This section describes template design for informed-filters that is necessary for this scheme and how to construct a strong classifier using the designed templates.

4.1 Template Design for Informed-filters

In this paper, we applied previously designed templates to create a feature pool. We used the template designed for the VSPETS dataset (VS-PETS, 2001). Fig. 10 shows an edge map computed using positive samples of the VSPETS dataset and 11 shows cell division and labels assigned using the edge map. According to the assigned labels, feature templates are shown in Fig. 12.

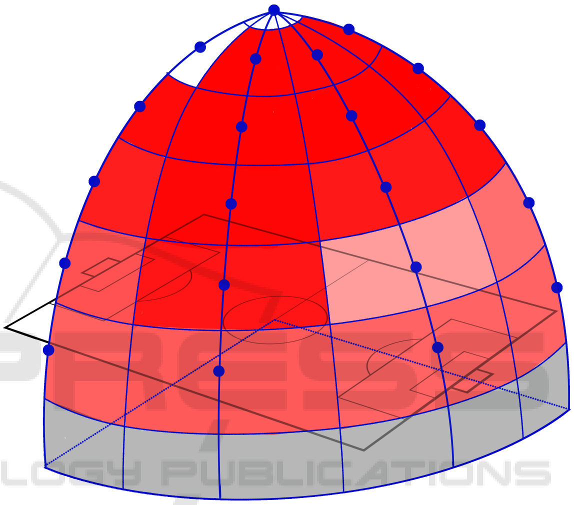

Figure 4: Heatmap.

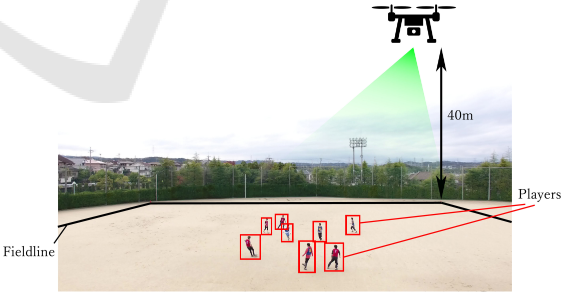

Figure 5: Experimental setup for capturing aerial images using a drone.

4.2 How to Construct a Strong Classifier using the Designed Templates

After the design of feature templates, a strong classifier is trained by the AdaBoost algorithm (Freund and Schapire, 1997). In the boosting process, effective features are selected from a feature pool where

Figure 6: The drone used in the experiment.

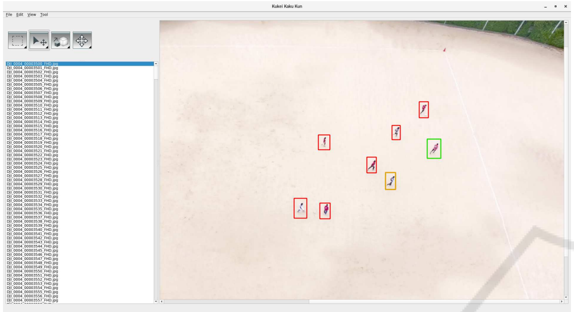

Figure 7: Screenshot of QtKukeiKakuKun developed for this experiment.

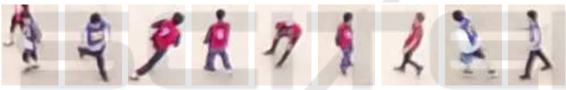

Figure 8: Examples of positive samples in the new dataset.

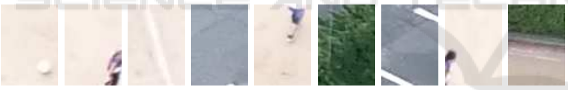

Figure 9: Examples of negative samples in the new dataset.

many features are generated randomly using the designed templates. After the selection of effective features as weak classifiers, a strong classifier composed of cascaded weak classifiers is obtained as a soft cascade classifier constructed by Multiple-Instance Pruning (MIP) (Zhang and Viola, 2007).

In our experiment, 3000 features were generated randomly using the designed templates. To determine the detection accuracy for reduced weak classifiers, the number of weak classifiers in a strong classifier was varied from 5 to 200.

5 EVALUATION

This section discusses the results of player detection in aerial images captured using a camera mounted on a drone.

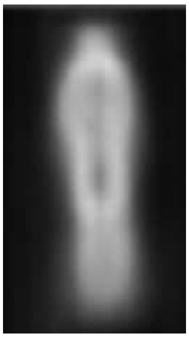

Figure 10: Edge map for PETS samples.

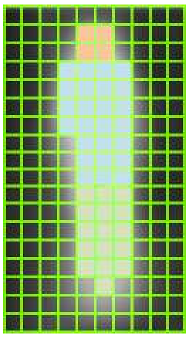

Figure 11: Labeling for PETS samples.

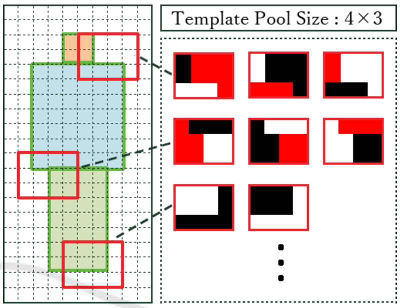

Figure 12: Examples of designed templates.

5.1 Detection Error Tradeoff Curves

Fig. 13 shows the detection error tradeoff curves when the number of weak classifiers used in a strong classifier is varied from 5 to 200. The detection performance improves as the number of weak classifiers increases. The best detection performance is achieved when the number of weak classifiers is 175, though 200 weak classifiers should be most powerful to represent target features. This is because randomness

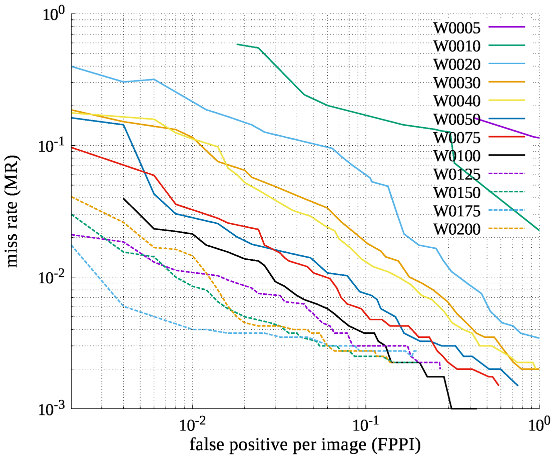

Figure 13: Detection error trade off curves for several numbers of weak classifiers.

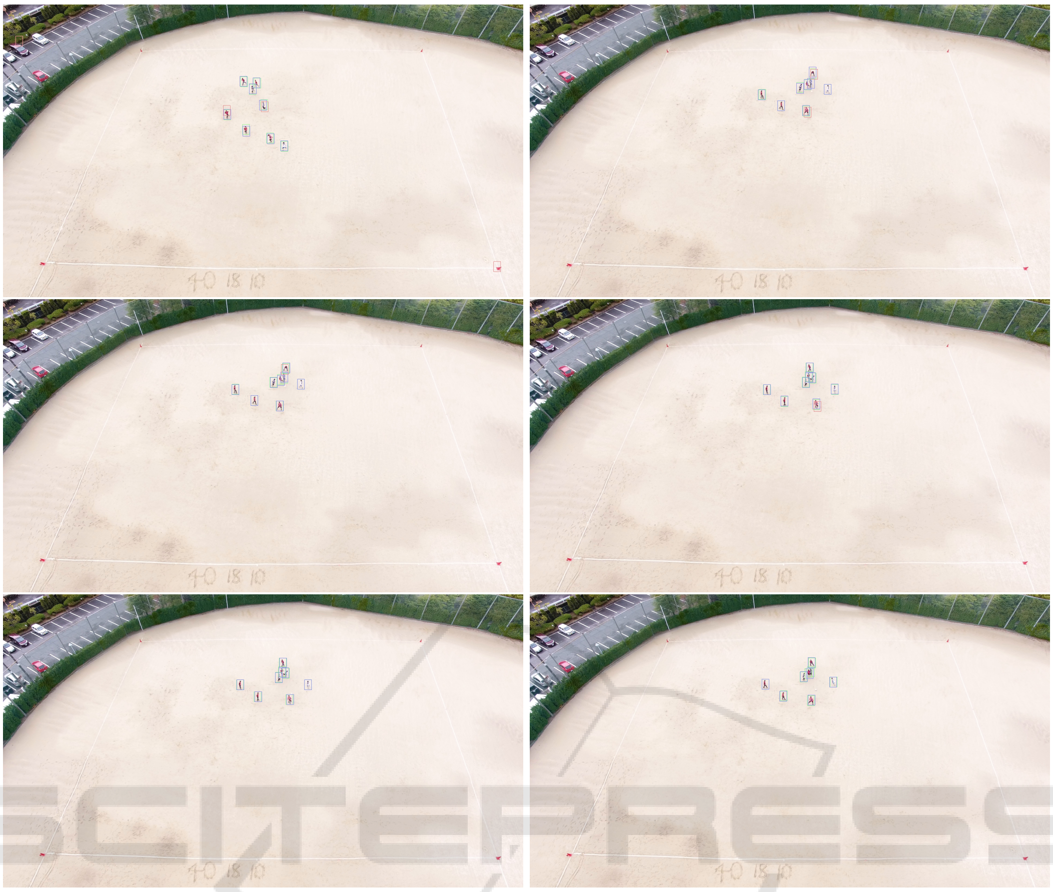

Figure 14: Detection examples for several numbers of weak classifiers.

was included in the training process to construct a strong classifier.

At the 10^{-1} false positive per image (FPPI) that is often used to compare the detection performance, 125, 150, 175, and 200 weak classifiers showed similar performance. This result means that the number of weak classifiers can be reduced by up to 37.5% without any degradation in detection performance.

5.2 Detection Examples

Fig. 14 shows the detection examples when the number of classifiers used in a strong classifier are changed. The number of weak classifiers were 10, 30, 50, 100, 150, and 200 for the top left, top center, top right, bottom left, bottom center, and bottom right images, respectively.

Surprisingly, in the top left regions of the image, false positives rarely occurred even though only color features were adopted to construct a strong classifier for detection. In these examples, a false positive occurred only in the case of 10 weak classifiers.

6 CONCLUSION

This paper presented a novel dataset composed of actual images captured by a camera mounted on a drone. The dataset includes 5000 images of a mini game of soccer involving eight people. Locations of humans in these images were manually marked using a novel software tool named QtKukeiKakuKun that was developed using python and Qt for this research.

Experimental results using the actual dataset showed that excellent detection accuracy could be achieved by a detector composed of informed-filters using only color features. The detection accuracy is maintained even when the number of weak classifiers are reduced from 200 to 125. This indicates that the processing speed of detection by a classifier composed of informed-filters can be improved without any degradation of detection accuracy.

ACKNOWLEDGMENT

The research results have been partly achieved by "Research and development of Innovative Network Technologies to Create the Future", the Commissioned Research of National Institute of Information and Communications Technology (NICT), JAPAN.

REFERENCES

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 580–587.

Hara, S., Yomo, H., Miyamoto, R., Kawamoto, Y., Okuhata, H., Kawabata, T., and Nakamura, H. (2017). Challenges in Real-Time Vital Signs Monitoring for Persons during Exercises. *International Journal of Wireless Information Networks*, 24:91–108.

Miyamoto, R., Nakamura, Y., Ishida, H., Nakamura, T., and Oki, T. (2019). Comparison of object detection schemes using datasets of sports scenes.

Miyamoto, R. and Oki, T. (2016). Soccer Player Detection with Only Color Features Selected Using Informed Haar-like Features. In *Advanced Concepts for Intelligent Vision Systems*, volume 10016 of *Lecture Notes in Computer Science*, pages 238–249.

Miyamoto, R., Yokokawa, H., Oki, T., Yomo, H., and Hara, S. (2017). Human detection in top-view images using only color features. *The Journal of the Institute of Image Electronics Engineers of Japan(in Japanese)*, 46(4):559–567.

Oki, T. and Miyamoto, R. (2017). Efficient GPU implementation of informed-filters for fast computation. In *Image and Video Technology*, pages 302–313.

Oki, T., Miyamoto, R., Yomo, H., and Hara, S. (2019). Detection accuracy of soccer players in aerial images captured from several viewpoints. *MDPI J. Funct. Morphol. Kinesiol.*, 4(1).

Redmon, J. and Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. pages 6517–6525.

VS-PETS (2001). Proc. IEEE international workshop on visual surveillance and performance evaluation of tracking and surveillance. http://ftp.pets.rdg.ac.uk/.

Zhang, C. and Viola, P. (2007). Multiple-instance pruning for learning efficient cascade detectors. In *Proc. Advances in Neural Information Processing Systems*.

Zoph, B., Vasudevan, V., Shlens, J., and Le, Q. (2018). Learning transferable architectures for scalable image recognition. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 8697–8710.