

Fundamental Research on Athlete Positions Estimation in Indoor Sports at Various View

Iori Iwata

*Graduate School of Sustainable of Systems Science, Komatsu University, Komatsu, Ishikawa 923-8511, Japan
Email: 24311001@komatsu-u.ac.jp*

Yoshihiro Ueda

*Faculty of Production Systems Engineering and Sciences, Komatsu University, Komatsu, Ishikawa 923-8511, Japan
Email: yoshihiro.ueda@komatsu-u.ac.jp*

Kazuma Sakamoto

*Faculty of Production Systems Engineering and Sciences, Komatsu University, Komatsu, Ishikawa 923-8511, Japan
Email: kazuma.sakamoto@komatsu-u.ac.jp*

Riku Kaiba

*Graduate School of Sustainable of Systems Science, Komatsu University, Komatsu, Ishikawa 923-8511, Japan
Email: 24211007@komatsu-u.ac.jp*

Abstract

In recent years, data collection for tactical analysis in sports has become increasingly prevalent. In sports such as volleyball, basketball, and soccer, where player positioning is closely linked to scoring opportunities, research has been conducted to visualize player positions using various technological approaches. Notably, numerous research has focused on enhancing tactical analysis by estimating player positions through image recognition methodologies. These approaches typically rely on images captured by one or more cameras. From these images, specific reference points on the court are identified and transformed into a bird's-eye view using image transformation algorithms such as projective transformation to visualize player positioning. This process requires the selection of four reference points on the court, preferably encompassing the entire playing area. However, capturing these four reference points from ideal viewing angles is often infeasible in many venues. Additionally, live game images frequently feature zoomed-in views of players and shifts in camera angles as the ball is tracked, creating challenges for consistent analysis. These limitations restrict tactical analysis to videos specifically recorded for such purposes, excluding archival video not originally intended for research and thus limiting data diversity. To address these challenges, this research aims to develop a system that broadens the applicability of tactical analysis by utilizing video captured in stadiums where a part of views is feasible, as well as historical video data. Focusing on volleyball, the proposed approach automatically identifies reference points based on the coordinates of the net and court lines to estimate player positions. This system seeks to enable robust and efficient analysis across diverse video sources, enhancing the scope and utility of tactical insights.

Keywords: volleyball, player position, bird's-eye view, projective transformation

1. Introduction

In recent years, Japan has introduced the Third Basic Plan for Sports [1], which is to be implemented over a five-year period from FY2022 to FY2026. The plan places an emphasis on leveraging advanced technologies and digital transformation (DX) [2] to effect revolutionary change in the domain of sports. In this context, considerable effort has been made in the field of sports to acquire player position data for the purpose of supporting tactical analysis [3], [4], [5]. This practice is particularly prevalent in sports such as soccer, volleyball, and basketball, where player positioning is closely tied to tactical strategies and scoring opportunities. In these sports, a considerable number of teams employ dedicated analysts, alongside coaches and managers, to track player positions at every moment and conduct in-depth tactical analysis.

In volleyball, an indoor sport, data analytics software such as DataVolley [6] is widely used to assist in tactical analysis. This software, which is primarily utilized by professional teams, requires input such as player positions, play details, and back numbers. By accumulating and analyzing this data, the software provides actionable data that contributes to tactical decision-making. However, it is necessary to manually input the data, and as a result, the adoption has been delayed in amateur teams. Recently, advancements in technology have enabled the estimation of player positions using camera images or other sensors, allowing for easier visualization of player movements. Typically, this involves creating a bird's-eye view and plotting player positions after estimating them from camera footage. Projective transformation, a common method used for this purpose, requires at least four reference points. Existing research [1], [2] use the four corners of the court as reference points. Nevertheless, this methodology restricts the scope of investigation to a

predefined angle of view, necessitating adjustments whenever the angle undergoes a change. Furthermore, conventional techniques often rely on video footage encompassing the entire court, which limits their scope of application.

Therefore, this research aims to create a player position estimation system that can handle various angles of view.

2. Methodology

(Fig. 1) illustrates the research process, which is represented by a flow diagram. The input is an image of a volleyball player, and the output is a bird's-eye view. The input images are processed using YOLOv9[7] (hereinafter referred to as YOLO), a type of object recognition algorithm, to recognize players. The lines and the net in the image are visually inspected, and their coordinates are manually obtained. The coordinates are obtained by visually locating the lines and nets in the image and tracing them with the mouse. Next, the number of line intersections is calculated, and reference points for projective transformation are determined based on this number. The established reference points are employed to transform the coordinates through a projective transformation, thereby generating a bird's-eye view. In this study, the YOLO algorithm is employed to identify individuals from match videos and ascertain their locations within the images. The individuals are represented by bounding boxes, as illustrated in (Fig. 2), and the upper left and lower right coordinates of the square are extracted. In this case, the upper left coordinate is (0,0), and thus, the coordinate value calculated from Eq. (1) is utilized as the individual's position within the image.



Fig.2 Person detection with YOLOv9

2.1. Method of reference points selection for various view

First, the coordinates of the four reference points for the projective transformation are obtained. In this research, a linear equation of the volleyball court line is obtained. In this case, depending on the angle of view, only a portion of the line may be used, but this is not a problem because the line is extended for use. In such cases, the intersection can be found by extending the lines. After obtaining multiple court lines on the image, the intersections are calculated. After acquiring several court lines in the image, the intersection points are calculated. In general, there can be anywhere from 0 to 6 intersections within an image.

If there are six intersections, as illustrated in (Fig. 3), the entire court is reflected and the coordinates of the four corners of the court can be obtained, so they are used as they are. If there are five or fewer intersections, the straight-line equation of the acquired court lines is extended within the image area as shown in the dotted line in (Fig. 4), and further intersections are searched for. If a new intersection is detected, as indicated by the blue dot in (Fig. 4), the point in question should be selected.

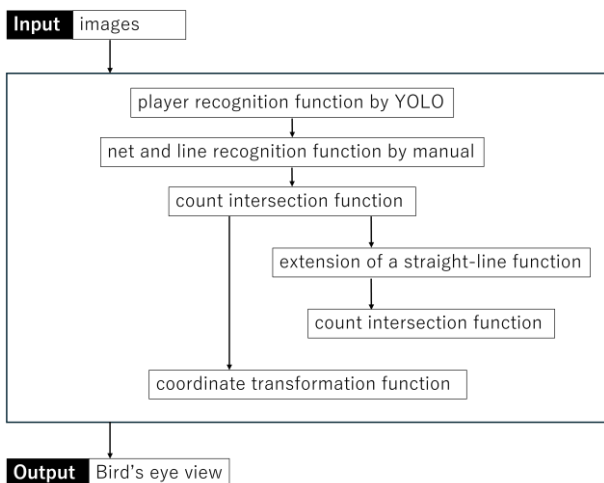


Fig. 1 Flow diagram

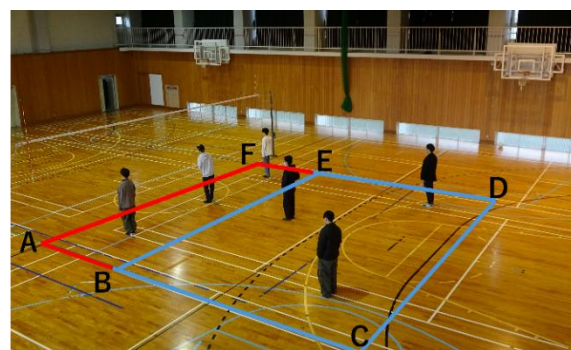


Fig.3 Taken image with virtual court line

$$\begin{cases} x = (x_{min} + x_{max})/2 \\ y = y_{max} \end{cases} \quad (1)$$

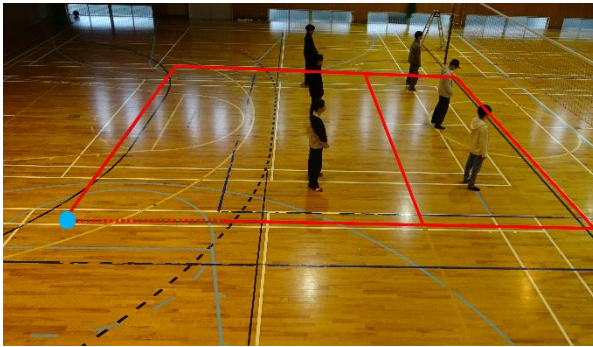


Fig.4 Creating intersection by extension

After this operation is completed, the number of intersections is checked. When the number of intersections is less than four, as illustrate in (Fig. 5), the players are hardly seen in the image and it is difficult to use the image for tactical analysis, so “Please change the angle of view” is represented to encourage the viewer to change the angle of view.

Next, the reference point is established based on the number of intersections identified. In this investigation, the angular perspectives captured within the blue frame in (Fig. 3), do not encompass the net and, thus, were excluded from the tactical analysis. As a result, these angles were excluded from research. When there are four or five intersections, the selected reference points are among A, B, E, and F in (Fig. 3). The selection process is carried out accordingly. The point closest to the straight line of the net and to the left of the net is A, the other is F, the point closest to A that is not F is B, and the point closest to F that is not A is E. When there are six points of intersection, A, B, E, and F are defined in the same way, and the point on the straight-line AB is C and the other is D.



Fig.5 Example with 4 or less intersections

2.2. Creating a bird's eye view

Using the coordinates obtained in section 2.1 as a reference point, the projective transformation is used to transform the image into the bird's-eye view shown in (Fig. 6) to estimate the position of the players. The projective transformation is performed with reference to [8]. The number of intersections has three patterns from Section 2.1: four, five, and six. For four or five of them,

convert A, B, E, F in (Fig. 3) to A', B', E', F', in (Fig. 6) for six, convert A, C, D, F to A', C', D', F' when there are six.

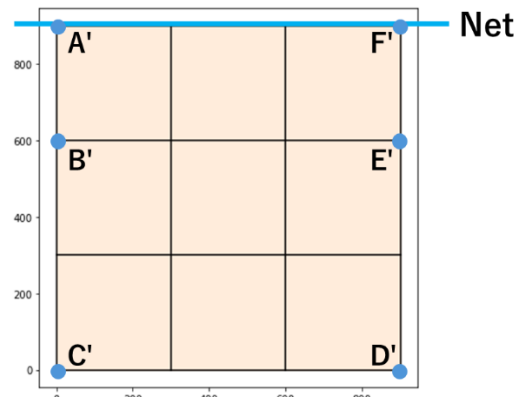


Fig.6 bird's-eye-view

3. Experimental Details

This experiment was conducted in the gymnasium at the Awazu Campus of Komatsu University. As shown in (Fig. 7), multiple cameras were installed to capture images from various angles. Examples of images taken from each angle are presented in (Fig. 8). In this experiment, to obtain the correct position of the players, although videos are typically used to estimate player positions, this experiment utilized still images to evaluate accuracy. The ground truth position of each player on the court was measured using a tape measure and defined as the correct position. The positional arrangement of the players during this experiment is illustrated in (Fig. 9). The error between the player positions estimated by the system and the ground truth positions was calculated and used for evaluation.

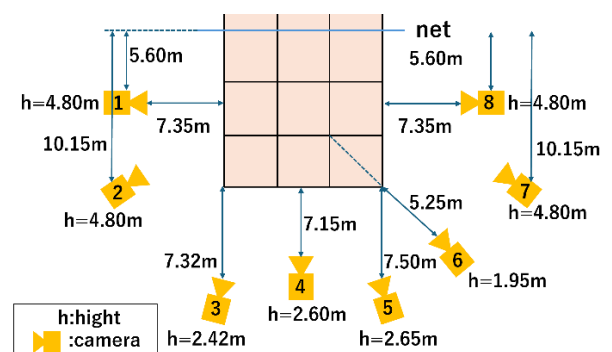


Fig.7 Shooting method

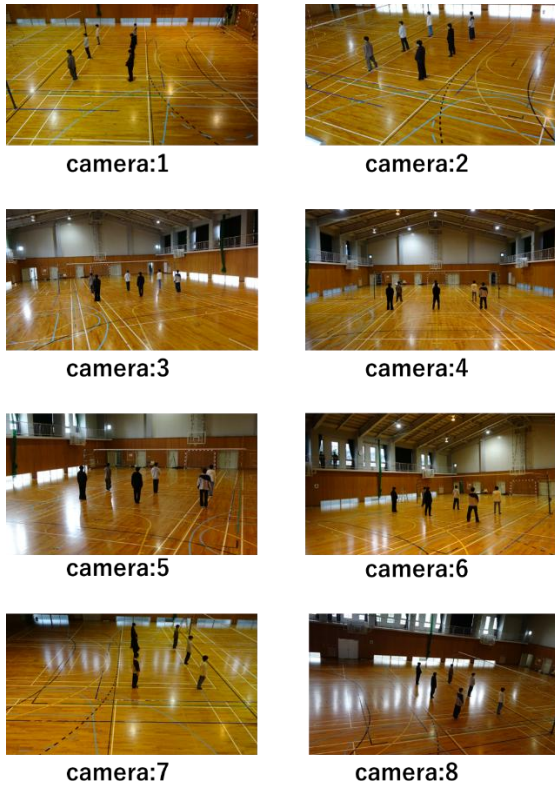


Fig. 8 Images taken from each camera angle of view

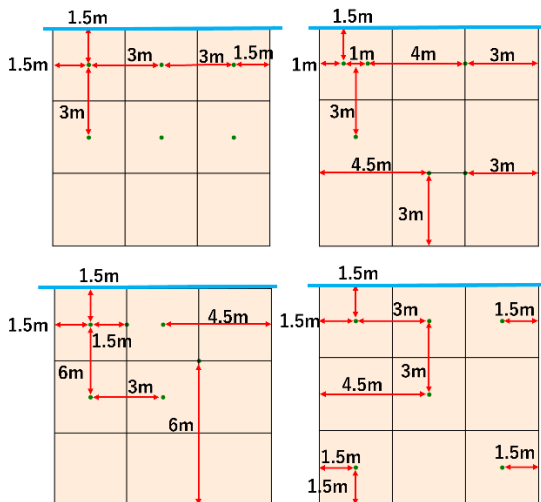


Fig.9 All positions pattern

4. Results and Discussion

The results of this research are summarized in (Table 1). (Table 1) presents a subset of data, displaying the coordinates and ground truth positions of 351 successful plots. (Fig. 10) shows an example of the output bird's-eye view. The red dots (Fig. 10) are the player positions estimated by the system, and the green dots are the correct positions. The line connecting (0,900) and (900,900) is the net position. The x-axis direction in this figure is the net horizontal direction, and the y-axis is the net vertical direction. These plots were derived from a total of 64 images captured by cameras 1 through 8. It is important

to note that plots located outside the court were excluded from this experiment. This exclusion was intended to avoid plotting coaches, managers, spectators, or other individuals not directly involved in the game.

Table 1. Error with the correct position

	x1	y1	x2	y2	dx	dy	distance
1	359	209	349	190	0.13	0.34	0.36
2	677	219	676	191	0.02	0.50	0.50
:	:	:	:	:	:	:	:
351	361	383	349	354	0.22	0.52	0.56
avg.					0.21	0.21	0.30

unit: x1, y1, x2, y2(pix) dx, dy(meter)

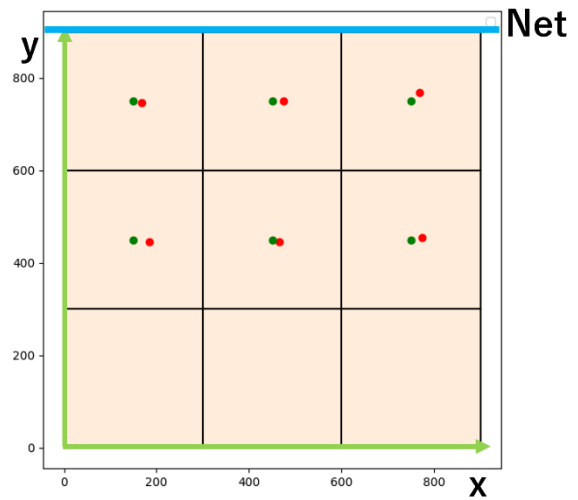


Fig. 10 An example of a bird's-eye view

As (Table 1) shows, the average error of the 351 players in this experiment was 0.21 meters in the x-coordinate, that is, in the direction of transition to the net, and 0.21 meters in the y-coordinate, that is, in the vertical direction of the net. The distance of the error is 0.30 meters. Since the shoulder width of a typical male is between 0.4 meters and 0.45 meters, we consider this accuracy to be sufficient for tactical analysis. As shown in (Fig. 11), the bounding box can accurately surround the frame of the person in the YOLO detection. If a person is surrounded by a frame that is too large, the error is the result of the deviation from the original frame.

(Table 2) summarizes the accuracy from each camera angle. The analysis revealed that the footage captured by cameras 3, 4, and 5, which are positioned near the rear of the court, exhibited larger average errors in the y-direction compared to the x-direction displacement across the entire court. In the image, the x-axis corresponds to the depth in the real world. Since this study uses monocular player position estimation, the error in the y-direction is larger than in the x-direction. Camera 4, which captures the court directly from behind, was found to have the highest error, likely due to the inherent difficulty in accurately estimating player positions along the y axis from this angle. Conversely, accuracy improved as the angle of view approached a top-

down perspective, as seen with cameras 1 and 2. However, camera 8, despite being positioned at the same height as cameras 1 and 2, did not achieve similar levels of accuracy. (Fig. 12) shows an image captured by Camera 8, where YOLO algorithm was employed to identify the players. Additionally, (Fig. 13) represents the bird's-eye view at the same time. In this experiment, slight discrepancies in the camera's field of view were observed. For instance, while Camera 1 captured the feet of occluded players, allowing for high-accuracy player positioning, Camera 8 did not capture the feet of the player in the center of the left column, as shown in (Fig. 12). Instead, the bottom edge of the bounding box was displayed around the player's waist. If the bounding box had enclosed the feet, it would have been possible to estimate the player's position more accurately. This discrepancy is thought to have contributed to the error in Camera 8's field of view.

This reduction in YOLO's detection accuracy caused by player occlusion likely contributed to the overall decrease in experimental accuracy. However, when examining images from Camera 4's field of view at the same moment, the two players occluded in Camera 8's field of view were successfully enclosed by bounding boxes. These findings suggest that leveraging multiple fields of view could enable the creation of a system more robust to occlusions.

Table 2. Error of each camera angle of view (meters)

camera's No.	dx	dy	distance
1	0.19	0.10	0.21
2	0.11	0.20	0.23
3	0.08	0.21	0.22
4	0.08	0.40	0.41
5	0.22	0.31	0.38
6	0.14	0.22	0.26
7	0.25	0.08	0.26
8	0.35	0.16	0.38



Fig. 11 Example of people detection

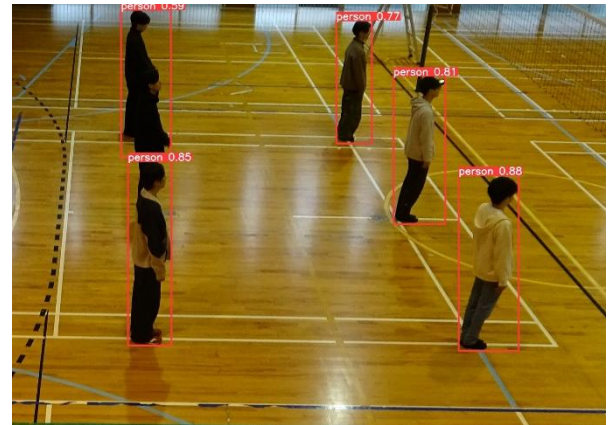


Fig 12 Player Occlusion

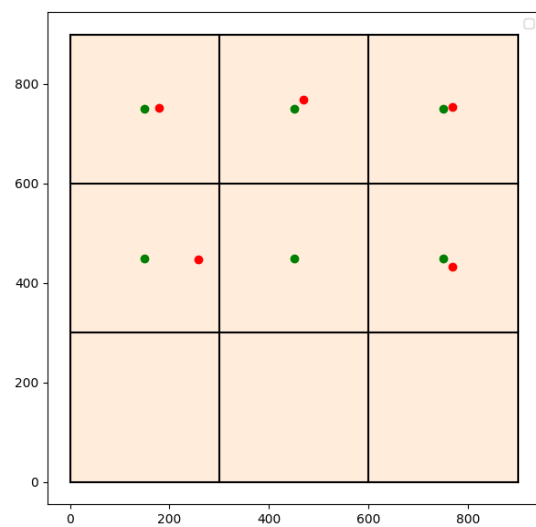


Fig. 13 Estimation of player position at occlusion

Next, we compared the results based on the number of intersections detected in the images. (Table 3) shows the average error when estimating the results of performing a projective transformation using A, B, E, and F as the reference points, and when performing a projective transformation using A, C, D, and F as the reference points. The results indicate that differences in error were observed along the y-axis. Since this comparison does not involve variations in detection accuracy by YOLO, the observed error is considered to arise during the projective transformation process. Generally, when using projective transformation to map shapes, accuracy tends to improve as the area of the quadrilateral formed by the reference points increases. Furthermore, points located inside the quadrilateral defined by the reference points are expected to have higher transformation accuracy compared to those located outside the quadrilateral. These two factors are considered to have contributed to the observed differences in accuracy.

Table 3. Difference in reference point and error

Reference point for the perspective transformation	dx	dy	distance
A, B, E, F	0.21	0.24	0.32
A, C, D, F	0.21	0.15	0.26

5. Conclusion

In this research, we aimed to broaden the scope of tactical analysis by estimating player positions from various viewing angles, rather than relying solely on a fixed angle of view. As detailed in Section 4, the overall error in this experiment was 0.30 meters. Although there is a difference of about 0.2 meters in error depending on the angle of view, it is possible to create a bird's-eye view even if the image was taken at an arbitrary angle of view. This result demonstrates that the proposed method can generate bird's-eye views from diverse angles, thereby contributing to tactical analysis. However, certain viewing angles may result in player occlusion, which can impede accurate player detection by YOLO. Furthermore, this experiment assumed that players remained on the ground; when players are airborne, such as during jumps, the estimation error is expected to increase. To address these limitations, future work will focus on developing a system that integrates multiple viewpoints to enhance robustness against occlusion and enable high-precision position estimation for airborne players. Such advancements would significantly improve the system's applicability for tactical analysis across a broader range of scenarios. Then, the detection of lines and nets is performed by an object recognition algorithm, aiming for full automation.

References

1. Japan Sports Agency, The Third Sport Basic Plan, https://www.mext.go.jp/sports/b_menu/sports/mcatetop01/list/1372413_00001.htm
2. Ministry of Internal Affairs and Communications, 2021 White Paper on Information and Communication Definition of Digital Transformation, <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r03/html/nd112210.html>
3. C.-C. Hsu, H.-T. Chen, C.-L. Chou, and S.-Y. Lee, 2D Histogram-based player localization in broadcast volleyball videos, *Multimed. Syst.*, vol. 22, no. 3, Jun (2016), pp. 325–341.
4. G. Tsiotas, DEEP LEARNING MODELS FOR THE AUTOMATIC ANALYSIS OF VOLLEYBALL PLAYS,
5. J. Theiner, W. Gritz, E. Muller-Budack, R. Rein, D. Memmert, and R. Ewerth, Extraction of Positional Player Data from Broadcast Soccer Videos, in 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA: IEEE, Jan (2022), pp. 1463–1473.
6. DataProject, DataVolley,
7. C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao, YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information, arXiv, 2024.
8. Y. Luo et al., A Review of Homography Estimation: Advances and Challenges, *Electronics*, vol. 12, no. 24, Jan (2023).

Authors Introduction

Mr. Iori Iwata



informatics.

He received the M.S. degree with the Graduate school of Sustainable Systems Science from Komatsu University, Japan, in 2024, where he is currently pursuing the Ph.D. degree at the Graduate school of Sustainable Systems Science. His research interests include image recognition, tactical analysis, and sports

Prof. Yoshihiro Ueda



productivity improvement, labor saving, automation using data science, and human centric system development.

He received the Ph.D. degree in Mathematical Information Science from Kanazawa University, Japan, in 2001. Since 2021, he has been a professor at the Faculty of Production Systems Engineering and Sciences, Komatsu University, Japan. His research interests include

ASST. Kazuma Sakamoto



informatics.

He received the B.S., M.S., and Ph.D. degrees in Informatics from Kansai University, Japan, in 2015, 2018, and 2021, respectively. Since 2021, he has been an assistant professor at the Faculty of Production Systems Engineering and Sciences, Komatsu University, Japan. His research interests include web mining, natural language processing, image processing, and sports

Mr. Riku Kaiba



computer vision and sports informatics.