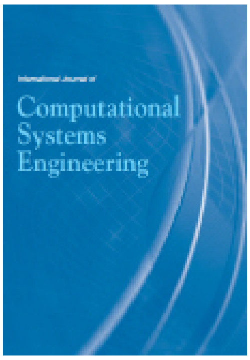


International Journal of Computational Systems Engineering

ISSN online: 2046-3405 - ISSN print: 2046-3391
https://www.inderscience.com/ijcsyse

Feature extraction of basketball player's foul action using machine vision

Xueliang Jia, Changguang Diao

DOI: 10.1504/IJCSYSE.2025.10071829

Article History:

Received:	26 December 2023
Last revised:	17 February 2024
Accepted:	04 March 2024
Published online:	18 June 2025

Feature extraction of basketball player's foul action using machine vision

Xueliang Jia* and Changguang Diao

Department of Physical Education,
Linyi Campus of Qingdao University of Technology,
Linyi 273400, Shandong, China
Email: ajiaxueliang@qut.edu.cn
Email: bdiaochangguang@qut.edu.cn
*Corresponding author

Abstract: The rising demand for basketball parallels the advancement of technology and the global economy, enhancing people's lifestyles. Modern basketball emphasises height, speed, accuracy, and strength, with increased physical contact. This intensifies gameplay, leading to inevitable mistakes. Foul regulations add unpredictability and challenge for players, coaches, and referees, influencing game outcomes. This study employs machine vision to analyse basketball foul movements. Experimental findings reveal patterns: the Chinese team committed 16 dribble defence fouls and 41 shooting defence fouls, while opponents committed 33 and 32, respectively. For ball defence, the Chinese team fouled five times in both preventing and defending, and four in pass prevention, compared to opponents' seven, 14, and five fouls, respectively. These results underscore the potential of machine vision in detailed detection of player fouls, offering insights for basketball development and gameplay refinement.

Keywords: machine vision; basketball players; foul action; feature extraction.

Reference to this paper should be made as follows: Jia, X. and Diao, C. (2025) 'Feature extraction of basketball player's foul action using machine vision', *Int. J. Computational Systems Engineering*, Vol. 9, No. 10, pp.20–30.

Biographical notes: Xueliang Jia is currently a physical education teacher at the Linyi Campus of Qingdao University of Technology. He has been working on the front line of physical education teaching for many years and has also led students to participate in various competitions as the head coach of the athletics team, achieving excellent results.

Changguang Diao is currently a full-time sports teacher at Linyi Campus of Qingdao University of Technology. He has led the team to train and participate in competitions as a football coach for many years, and has won the sixth place in the Shandong College Student Football League for many times.

1 Introduction

After years of development history of basketball, fierce physical confrontation has become a representative feature of competitive basketball. In the modern basketball game, foul is one of the most important factors that affect the winning or losing of the game and evaluate the player's ability (Gao and Hu, 2022). Fouls from different regions, courts, and times can fully reflect the team's basketball's concept of fouls, awareness of fouls, and ability to control fouls (Liu and Zheng, 2023). With the continuous improvement of the offensive level of competitive basketball, the defence has changed from passive to active, making the offensive and defensive conflicts in basketball more exciting and the game more difficult from beginning to end (Duan et al., 2022). Therefore, basketball players must perform reasonable or unreasonable technical actions in the game to achieve the purpose of controlling the ball in basketball games.

With the new changes to the offense rules, the secondary offense in the game of basketball is now more condensed, and the speed of offense and defence transitions has become faster (Wang and Chen, 2020). Basketball is a sport of intense physical confrontation. Since the physical confrontation between players becomes more intense in the game, fouls are naturally inevitable, and the tactical significance changes accordingly Song et al., 2020). Therefore, it's a challenge for coaches and players to maintain the intensity of the game and increase the motivation on offense and defence while staying as foul-free as possible. Mastering the characteristics and patterns of players' mistakes in the game makes an important contribution to the overall improvement of players' offensive and defensive performance (Yang et al., 2022). The positive effects of player mistakes during the game can support the team's tactics on both ends of the floor, thereby increasing the probability of winning to a certain extent.

In this paper, the characteristics of basketball players' foul actions based on machine vision are studied. The data showed that the Chinese team fouled 19 times for blocking, and the opponent fouled 19 times. The Chinese team fouled 64 times because of illegal hand, and the opponent fouled 82 times. The Chinese team fouled 8 times because of collision, and the opponent fouled five times. The Chinese team committed one technical foul and the opponent fouled three times. The Chinese team fouled two times for unsportsmanlike conduct and the opponent fouled one time. From the above data, it can be seen that after the research experiment of basketball players' foul action characteristics based on machine vision, it is of great significance to promote the development of current foul action monitoring.

2 Related works

This paper studied some techniques of foul action extraction, which can be fully applied to the research in this field. Hong et al. (2018) proposed a learnable temporal attention mechanism for automatic selection of important time points from action sequences. Seol et al. (2017) proposed a clinical question based on clinical semantic units and event causality patterns: an action relation extraction method. Hashimoto et al. (2017) introduced a human behaviour modelling method that designs human behaviour models based on stored data obtained during long-term monitoring of people. Kartmann et al. (2018) proposed a method for extracting physically plausible support relationships between objects based on visual information. Zhang and Zhang (2018) proposed a two-channel model to decouple spatial and temporal feature extraction. Maity et al. (2019) proposed an efficient method for human action recognition from sequences of contour images in videos. These methods provide some references for research, but due to the short time and small sample size of the relevant research, they have not been recognised by the public.

Based on machine vision, the following related materials have been reviewed to optimise the foul action extraction research. Dawood et al. (2017) aimed to develop an ensemble model based on image processing techniques and machine learning. Wang et al. (2017) developed a dynamic selection machine based on image processing technology. Tsai and Hsieh (2017) proposed a fast image alignment method using expectation maximisation technique. The research of Nouri-Ahmadabadi et al. (2017) developed an intelligent system based on machine vision and support vector machines. Zhang et al. (2020) proposed a hybrid convolutional neural network approach for fusion process monitoring. These methods provide sufficient literature basis for studying the feature extraction of basketball players' foul action by machine vision.

3 Overview of machine vision and motion feature extraction

In a fierce basketball game, various reasonable and unreasonable technical movements of the players appear alternately, and the referee must make an accurate and fair ruling quickly, which requires a high degree of concentration, but there are always distractions. Therefore, this paper makes the detection of basketball fouls more accurate by studying machine vision.

This article uses machine vision technology to study the feature extraction of players' foul actions in basketball games. It uses multi-sensor technology, the application of statistical learning methods, and the static and dynamic representation of target features. This method can more comprehensively and accurately capture the characteristics of basketball games. Complex movement patterns improve the accuracy of identifying athletes' foul behaviours. Feature extraction methods such as background subtraction and Gaussian mixture models are used to effectively overcome traditional limitations, cope with background noise and occlusion problems, and improve the robustness of the system. At the same time, through real-time processing and game data statistics, the machine vision system can obtain information in time during the game, provide referees with more timely auxiliary decisions, and improve the efficiency of the game process. By analysing the foul time period and area in detail, the system can more accurately locate and identify foul behaviour, providing an advanced and comprehensive solution for the accurate identification of fouls in basketball games.

Machine vision technology may bring several advantages in the fields of sports analysis, referee assistance and player evaluation. In sports analysis, using machine vision to accurately identify player fouls in basketball games can help sports analysts gain an in-depth understanding of the tactical and technical aspects of the game. Analysing the type, frequency and location of fouls can provide teams with key information for tactical improvement and opponent analysis. In terms of referee assistance, the machine vision system provides referees with real-time and objective data support, helping referees to more accurately determine foul behaviours during the game. This can improve the fairness of the game and the referee's decision-making level, and reduce disputes and mis-judgments. In terms of player evaluation, through detailed analysis of players' foul behaviours, coaches and team management can better evaluate players' performance and technical level. This has a guiding role in player training plans, tactical adjustments and player rotation plans.

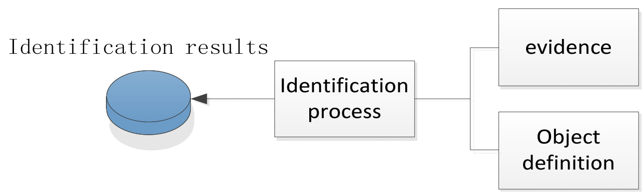
This research not only has practical applications in basketball, but also provides useful experience for technological innovation and data-driven decision-making in other fields. The use of machine vision technology has introduced advanced data analysis and processing methods into badminton, football and other sports fields, promoting the innovative application of technology in sports. The real-time data provided by machine vision systems provides an example for decision-making in safety monitoring and medical research. Through data analysis, machine vision systems provide objective basis for decision-making. This data-driven approach can promote more scientific management and decision-making practices in all walks of life, improving efficiency and accuracy.

3.1 Overview of machine vision

Computer vision, also known as machine vision, uses computers to simulate the visual function of the human eyes to extract information from images or image sequences for shape and motion recognition to recognise three-dimensional scenes and objects in the objective world. In a computer vision system, the input data is a grayscale matrix that represents a projection of a 3D scene. There can be multiple input fields, providing information from different directions, different viewpoints, and different points in time. The desired result is a symbolic description of the scene represented by the image. These are usually descriptions of object classes and relationships between objects, which may also include information such as spatial surface structure, shape, physical surface properties, colours, materials, textures, shadows, and light source locations (Miao et al., 2020).

With the development of computer technology, network technology and image processing technology, machine vision has become an integral part of modern industrial production, which includes expertise in pattern recognition, computer vision, digital image processing, machine learning and artificial intelligence. Machine vision systems can quickly capture large amounts of information and process them automatically with ease, facilitating integration with design and process control information, and enabling real-time identification and analysis. Now, machine vision technology has moved from laboratory to reality, which has been widely used in many fields, bringing huge economic and social benefits.

Figure 1 Object recognition process (see online version for colours)



It is not too early for machine vision to be proposed as an independent discipline, but it is by no means an independent research topic. Studying and mastering this subject require a comprehensive application of knowledge from other disciplines. In the field of machine vision, there are many interesting research directions that have long attracted many computer vision experts, who even dedicated their lives to some research tasks. Some of the important research areas are object tracking, object recognition, image processing, image segmentation, image classification, stereo vision, etc. The object recognition process is shown in Figure 1:

The problem of object recognition can be viewed as a process of debate. The image is the source of the evidence, and the conclusion is the determination of the existence of the object and its location. The identification method is the bridge from the evidence to the conclusion. This 'evidence' is often indeterminate because the characteristics of the object itself are not well-defined in the image, and errors may occur in extracting the image's features. Therefore, in practice, the reliability of each 'evidence' is determined by assigning a confidence level (Fernandez-Robles et al., 2022).

The research of machine vision mainly includes input devices, low-level vision, middle-level vision, high-level vision and system structure. Input devices include image processing hardware and digitising hardware. Low-level vision mainly uses various image processing techniques and algorithms to process the original input image, such as image enhancement, image filtering, edge detection, etc. The main task of intermediate vision is to recover depth, surface normality, contour, and other 2.5-dimensional scene information through stereo vision. The task of high-level vision is based on the original input image, the image baseline and the 2.5-dimensional map. It needs to complete the three-dimensional image reconstruction of the object in the central coordinate system and create the three-dimensional description of the object to identify the three-dimensional object and determine the position and orientation of the object. The system structure is not a concrete design example but a system model.

The image (original image) obtained by the imaging system cannot be directly used by the vision system because the vision system is interfered and restricted by various parties, and its image must be pre-processed by noise filtering or grayscale correction. The image acquisition process is shown in Figure 2:

Figure 2 Image acquisition process (see online version for colours)

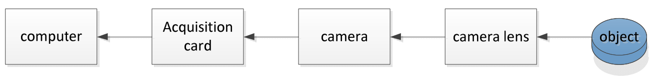


Image acquisition is to convert the visualisation and intrinsic properties of the measured object into data that can be processed by a computer, which directly affects the stability and reliability of the system. Optical systems, light sources, cameras, image processing devices (or image memory cards) are usually used to capture images of the object under investigation (Liu, 2022).

The foundation of machine learning is the study of statistical theory and data. Machine learning is the core technology of artificial intelligence. By simulating human learning behaviour, searching for rules from a large amount of data and judging unknown data according to the rules, this method makes computers more intelligent. Due to the shortcomings of artificial neural networks, it is easy to overfit with data learning, and the generalisation ability is poor. In traditional pattern recognition technology, it is usually only considered to improve the recognition accuracy by increasing the number of samples. This method only considers the fit of the classifier to the training samples, and in practice, the test samples cannot be accurately classified.

Support vector machines combine multiple binary classification problems, resulting in the final extension of multi-classification. Determining whether a sample is linearly separable is based on whether the sample can be divided by a linear function. The sample is said to be linearly separable if it can be split, and nonlinearly separable if it cannot. Then the general form of a linear discriminant function in D-dimensional space is as equation (1):

$$g(m) = \omega m + b \tag{1}$$

The decision function is as equation (2):

$$f(m) = \operatorname{sgn}(\omega \cdot m + b) \tag{2}$$

Normalising the discriminant function, the constraint is as equation (3):

$$n_i\left(\omega m_i + b\right) \geq 1, i = 1, 2, \ldots, y \tag{3}$$

The Lagrangian function is defined as equation (4):

$$L(\omega, b, a) = \frac{1}{2}\|\omega\| - \sum\nolimits_{i=1}^{y} \alpha_i \left[n_i(\omega \cdot m_i + b) - 1\right] \tag{4}$$

According to equation (4), the following constraint as equation (5) can be obtained:

$$\sum\nolimits_{i=1}^{y} \alpha_i n_i = 0 \alpha_i \geq 0 \tag{5}$$

If α^* is the optimal solution, then equation (6) and equation (7) can be obtained:

$$\omega^* = \sum\nolimits_{i=1}^{y} \alpha^* n_i m_i \tag{6}$$

$$b^* = n_i - \omega^* \cdot m_i \tag{7}$$

Among them, m is the test sample category determined by the optimal classification function.

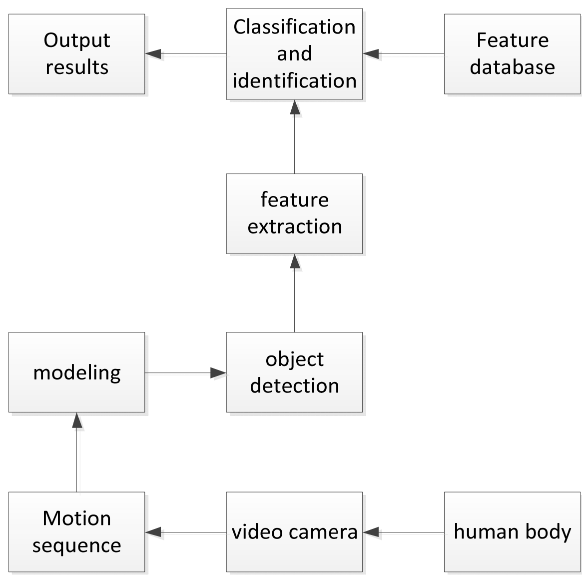
This article chooses the CNN model for feature extraction. When selecting the model, first of all, fully consider that foul behaviours in basketball games usually involve complex movements and positional relationships in space, and CNN has excellent spatial feature learning capabilities when processing image data. Second, local actions and key areas are very important for foul action feature extraction, and the local receptive field of CNN enables it to focus on the local structure of the image and accurately capture these details. Third, the complex actions and multi-level feature representation in basketball games require the model to have qualified shared weights and hierarchical structures. The shared weights and hierarchical structures in CNN enable it to effectively learn feature hierarchies in images. Foul behaviours in basketball games are mainly recognised through images. CNN performs well in processing image data. This adaptability makes CNN a suitable choice for processing basketball game images and extracting features from them.

During the training process, due to the variety of actions and scenes in basketball games, the use of data enhancement technology can help improve the generalisation ability of the model and increase the diversity of training data through random rotation, cropping, scaling, etc. Using pre-trained CNN models, such as those trained on large-scale image datasets, can accelerate model convergence and improve performance. The idea of transfer learning can be used for targeted training on basketball game images. The structure of the CNN may be adjusted according to experimental needs, such as adding or reducing convolutional layers and pooling layers, to adapt to the complexity of basketball game images. More importantly, hyperparameters such as learning rate and batch size can be adjusted to ensure that the model can effectively learn the characteristics of fouls in basketball games during the training process. In the experiment, the learning rate was set to 0.001, the batch size was 32 images, and the Adam optimiser was used for model optimisation.

3.2 *Overview of motion feature extraction*

Motion recognition is a broad field of research that encompasses multi-sensor technology, image processing, computer-aided design, pattern recognition, virtual reality, computer vision and graphics, visualisation techniques, and intelligent robotic systems. Methods for analysing and processing human motion perception usually include motion pattern recognition, motion pattern feature extraction and motion pattern recognition in complex backgrounds. As part of high-level processing, behaviour understanding and identity recognition are hot research areas that have received extensive attention in recent years, especially for human motion analysis and description, and thus for identity recognition. The proliferation of various motion detection algorithms has made human feature extraction and identification an important problem for today's systems (Liang et al., 2022).

Figure 3 Typical motion feature recognition system



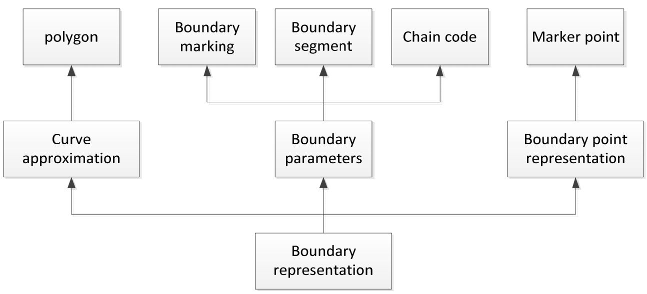
A standard motion feature recognition system consists of three main parts, image pre-processing and motion feature detection, motion feature extraction and classification. The three components include a camera, a host system, and a software package to process and identify people in moving video footage. A typical motion feature recognition system is shown in Figure 3:

First, a video surveillance system captures a sequence of video images of human movement. Second, image pre-processing and binarisation are performed on the detected video sequences of human moving objects. After these operations, the moving body image can be clearer and the background is single. Third, the typical features of human motion are extracted according to various rules and processed appropriately to ensure the consistency of these features with human motion patterns. Finally, the motion features to be identified are compared with feature templates in the feature database in order to manipulate and classify them (Liu, 2020).

Traditional feature recognition methods mainly use pattern matching, which is fast but not reliable, with a low recognition rate in complex environments. In recent years, feature recognition has mainly relied on statistical learning methods, that is, first building a set of rules based on knowledge, or using statistical learning methods to automatically learn corresponding rules and knowledge from patterns, and using the built or learned rules to classify patterns. Statistical learning methods overcome the shortcomings of traditional recognition methods and have good recognition performance in complex backgrounds and noisy situations, which is the future development direction of recognition technology.

The representation methods of target features are divided into static and dynamic features. The static representation of target features belongs to the perceptual layer, which mainly includes the shape, outer contour, colour information, and texture information of the target image. They can be divided into methods that represent features in terms of boundaries, regions, and changes (Yu et al., 2021).

Figure 4 Technology classification based on boundary representation

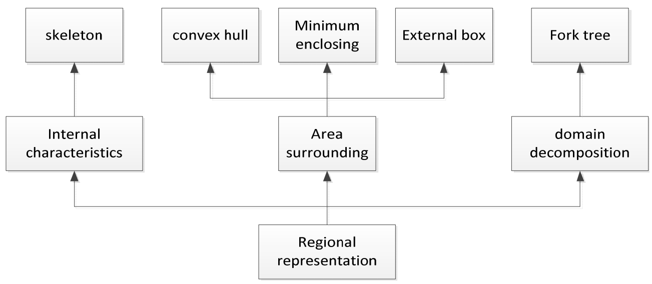


Boundary-based feature representations generally fall into three categories. The first category is to use boundary points to represent contours, usually using the marker point method. The second category is to use typical boundary parameters, such as chain codes, boundary segments, etc. The third category is to use curve approximation methods, of which the polygonal method is the most commonly used. The technology classification based on boundary representation is shown in Figure 4:

The use of barcode features for boundary representation is a method. The starting point is represented by coordinates, and the remaining points are described by calculated lines, which are often used to represent the boundaries of curves or regions. More accurate information is represented by using the surrounding eight adjacent dot links, which are used relatively frequently. Boundary features are also used to represent curves as approximations of polygons. For any closed curve, this representation can be considered to be valid if the number of points on the edge of the curve is equal or approximately equal to the number of edges of the polygon (Guo and Li, 2022).

Feature recognition typically uses differences between features of different objects and similarities between features within the same object. Region-based algorithms are the most common parallel methods for direct region detection. The idea of extracting structural features from moving objects is to first obtain meaningful sub-regions in the distribution structure of objects, where meaningful means that moving objects are divided into several sub-regions with topological or physical significance. Topological meaning refers to the spatial topological relationship between sub-regions, and the physical meaning refers to the independent motion characteristics of each sub-region. The technical classification based on region representation is shown in Figure 5:

Figure 5 Technology classification based on region representation



There are three region-based representations that use techniques such as region decomposition and region enclosing to approximate object features as well as interior features to represent objects. Spanning tree is used for region decomposition, outer box, minimum bounding rectangle, and convex graph are used for region enclosing. Skeleton features are used for internal features. Region decomposition is to decompose the target object into several simple units, where the enclosing region is represented by the geometric primitive filling, and the content feature is the collection of pixels within the region. The third method is to use internal region features, such as skeleton features, and it is also a common method to use regions to represent target features.

Feature extraction algorithms are unreliable in real-time due to background noise, shadows caused by lighting changes, errors caused by camera shake, and mutual and self-occlusion caused by moving objects. Frame difference method, optical flow method and background stretching method are the most commonly used feature extraction methods (Lei et al., 2022).

Image disparity is an algorithm that uses the absolute value of the difference in brightness between the previous image and the next image in a sequence of video images. It is great for highlighting changes between two images in an image sequence. Not only does it work with multiple moving targets, but it also works with moving cameras. The mathematical formula of the frame difference method is as equation (8):

$$G(m,n) = \left[f_{k+1}(m,n) - f_k(m,n) \right] \tag{8}$$

If the absolute value of the brightness difference obtained after discrimination is greater than the predetermined threshold, it is considered as the target area of human motion. Otherwise it is considered as the background area.

Moving objects on the retina of the human eye form a continuous sequence of images over time, giving the impression that the image 'flows' across the retina, hence the name 'optical flow'. At each time point, there are corresponding points in the image, which are connected by equation (9):

$$G(m,n,t) = G(m + \Delta m, n + \Delta n, t + \Delta t) \tag{9}$$

The right-hand side of equation (9) is expanded with Taylor formula, and the formula for constraining optical flow can be obtained as equation (10):

$$\frac{\partial G_t}{\partial m}\frac{dm}{dt} + \frac{\partial G_t}{\partial n}\frac{dn}{dt} + \frac{dG_t}{dt} = 0 \tag{10}$$

Equation (10) gives the spatio-temporal difference correspondence of the gradient of the moving target in time and space. The formula to minimise the function is defined as equation (11):

$$\varepsilon_{flow} = \sum\nolimits_m \left(u_x e_m + v_x e_n + e_t\right)^2 \tag{11}$$

Since the optical flow method is based on the Taylor formula, the optical flow method is very sensitive to light and noise, and it is computationally expensive to separate some objects (Wang et al., 2022).

For scenes where the camera position is fixed, the most commonly used object detection method is the background subtraction method. The basic principle of the background subtraction method is to construct a model about the background by analysing the video sequence and to compare the current frame image with the background model, detecting the moving target by comparing the results. The formula for background subtraction is as equation (12):

$$R_t(m,n) = \begin{cases} 0 \; \left| f_t(m,n) - b_t(m,n) \right| > T \\ 1 \; \left| f_t(m,n) - b_t(m,n) \right| \le T \end{cases} \tag{12}$$

The background difference method is a simple real-time algorithm that can capture a more complete image of a moving target than the image difference method. However, in practice, lighting and environmental conditions are constantly changing over time and are susceptible to interference between the current image and the background model. In order to overcome the interference caused by changes in the external environment, the background model needs to be continuously updated in practical applications.

The Gaussian mixture model is an extension of one of the Gaussian models and is suitable for unimodal scenes, that is, situations that background changes can be ignored. The Gaussian distribution function is expressed as equation (13):

$$P\left(M_t, \mu_t, \sum\nolimits_t \right) = \frac{1}{(2\pi)^{\frac{y}{2}}} e - \frac{1}{2}(m_t - \mu)^T (m_t - \mu) \tag{13}$$

In equation (13), Y represents the pixel colour dimension, and Y is 1 for grayscale images. For a single Gaussian model, the background model is usually updated using equation (14) and equation (15):

$$\mu(t+1) = (1-\alpha)\mu_t + \alpha M_t \tag{14}$$

$$\sum\nolimits_{t+1} = (1-\alpha)\sum\nolimits_t + \alpha\left(M_t - \sum\nolimits_t \right)\left(M_t - \sum\nolimits_t \right)^T \tag{15}$$

The single Gaussian model is suitable for situations where the background scene changes slowly and the scene is simple. If the background is complex or changes rapidly, the background pixel distribution shows a multimodal distribution, and the single Gaussian model cannot accurately describe the background. To effectively describe the background, a Gaussian mixture model can be built for these pixels. The mixed Gaussian model uses multiple weights to represent complex and changing scenes. Compared with the optical flow method, the mixed Gaussian model method has lower computational complexity, showing better results for outdoor target detection. The probability density function of the Gaussian mixture model is set as equation (16):

$$p(m) = \sum\nolimits_{i=1}^{X} \omega_i g_i(m) \tag{16}$$

In equation (16), m represents a point in the Y-dimensional space, and X represents the number of mixtures of the mixture Gaussian model. The larger the X value, the stronger the ability to deal with fluctuations. The larger the weight of each Gaussian function, the more dominant the corresponding Gaussian function is in the mixture model, and the sum of each weight is 1, as shown in equation (17):

$$\sum\nolimits_{i=1}^{X} \omega_i = 1 \tag{17}$$

The use of the mixture Gaussian model is divided into three steps, namely defining the model, updating the model and background judgment.

For this article, the CNN model can be generalised in large-scale programming to various game settings, focus angles and lighting conditions, but there are some possible limitations and areas that may need improvement. If a model is trained on a dataset of a specific basketball game, it may not be applicable to other basketball games or game settings of different sizes. Large scale programming may involve various types of games and venue settings, so the model needs to have a wider range of conditions. Focus angle and view training may have an impact on the performance of severe detection. If the model is at a specific angle, it may not effectively handle the situation at other angles. Considering that in large-scale programming, focus coordinates and settings may vary from site to site, the model needs to have better robustness. Changes in lighting conditions may have a negative impact on the performance of the model. If the model is trained under certain lighting conditions, it may have an impact on the brightness of other lighting conditions. For robustness, the model may need to be trained under various lighting conditions.

4 Extraction of foul features in basketball games

This experiment uses public images of professional basketball games in roboflow as a dataset. There are 2,917 images in total, with a size of 2.89G. They include eight types of fouls, including tactical fouls, physical fouls, walking with the ball, and collision fouls, ensuring the accuracy of the experiment diversity.

Experimental pre-processing includes data enhancement, noise reduction and background processing. The specific pre-processing steps used in this experiment are as follows:

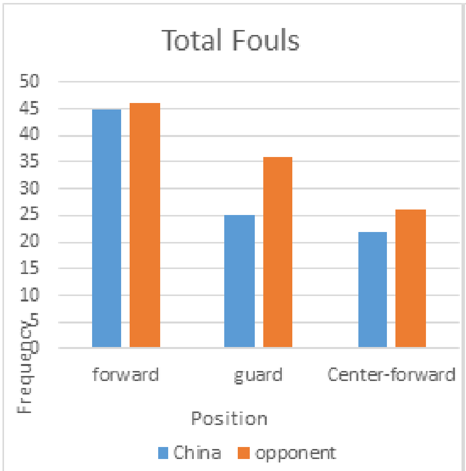
First, Gaussian filtering is used for noise reduction, which helps eliminate noise in the image and ensures that subsequent feature extraction steps are more accurate.

Second, threshold segmentation is used to separate the players and basketballs in the image from the background, which helps the model focus on the movements of the basketball players and reduce interference to the background. Finally, image enhancement processing is performed with random rotation and cropping.

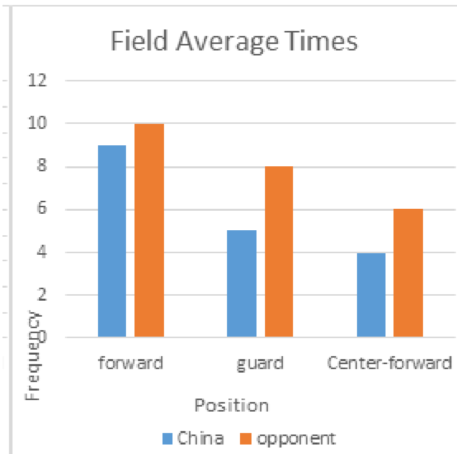
In feature engineering, this experiment focused on selecting some key attributes when extracting the features of foul behaviour, including movement trajectory, posture, ball position, etc. By tracking the trajectory of players, specific movement patterns can be captured, such as breakthroughs, man-to-man defence, etc., as these are important indicators of foul play. Observing an athlete's posture can provide information about their movements and intentions, as hand position and movement may suggest shooting or defensive actions. In addition, the position of the basketball can also be used as a reference for players' foul actions. Fouls usually involve illegal contact or interference with the ball, so the position and movement of the ball may be key features in identifying fouls.

In the feature extraction in this article, there may be some false positives and false negatives, which may be due to insufficient training data, insufficient feature extraction, or insufficient model complexity. If the selected feature extraction method fails to capture some key violation action features, it may lead to false negatives. In addition, if the complexity of the selected machine learning model is not sufficient to handle complex scenarios and actions, it may lead to false positives and false positives. Firstly, it can be ensured that the selected feature extraction method can cover various possible violations, including changes in different scenes and lighting conditions. Secondly, by increasing the diversity and scale of the dataset, it is possible to ensure that the training data covers various scenarios and actions, in order to reduce false negatives. The third option is to consider adjusting the threshold of the classifier to balance the trade-off between false positives and false negatives.

Figure 6 Statistics on the number of fouls by players in different positions in the Olympic men's basketball team (see online version for colours)



(a)



(b)

4.1 *Comparison of the number of fouls, time and areas in basketball games*

In this section, the video of the basketball match between China and the opponents in the Olympic Games was analysed. The video was strictly monitored using the feature extraction technology of machine vision. Competitive sports competitions were extremely intense. The form of the game was changing rapidly, and only with the help of video could the game situation be analysed more accurately. The statistics of the number of fouls in the Olympic men's basketball team are shown in Table 1:

Table 1 Statistics on the number of fouls in the Olympic men's basketball team

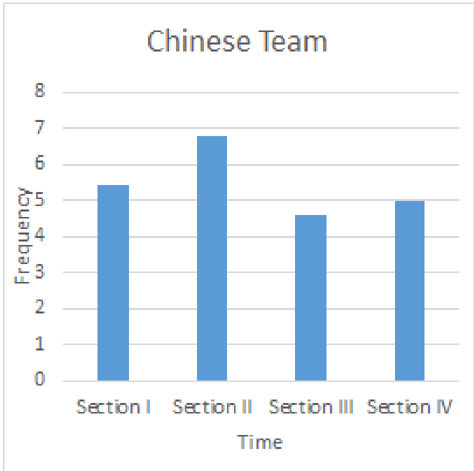
	China	*Opponent*	*China*	*Opponent*
Total number of fouls	18	22	90	106
Number of fouls of main force	9	12	45	56
Number of fouls of main force being fouled	15	13	75	61
Proportion of main force being fouled	71%	68%		
Number of fouls of substitutes	10	11	46	51
Proportion of fouls of substitutes	51%	48%		

It can be seen from Table 1 that the Chinese team committed 90 fouls in this Olympic Games, with an average of 18 per game. The fouls of main force totalled 45 times, averaging 9 times per game. The main force was fouled a total of 75 times, averaging 15 times per game. Substitutes fouled a total of 46 times, averaging 10 times. Opponents fouled a total of 106 times, averaging 22 per game. The main fouls of the opponents totalled 56 times, averaging 12 times per game. The main force of the opponents was fouled 61 times, averaging 13 times per game. Substitutes of the opponents fouled 51 times, averaging 11 per game. The statistics on the number of fouls by players in different positions in the Olympic men's basketball team are shown in Figure 6:

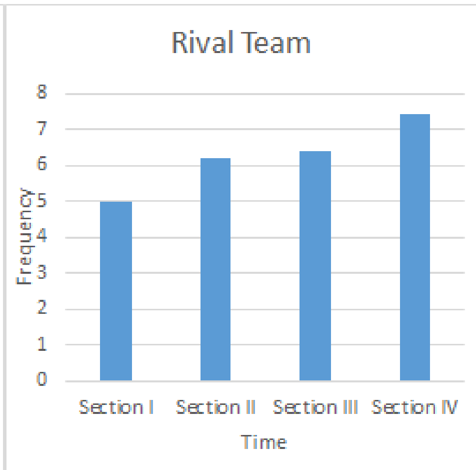
Figure 6(a) shows the total number of fouls by players at different positions, and Figure 6(b) shows the average number of fouls per game by players at different positions. As can be seen from Figure 6, the total number of fouls in the forward position of the Chinese team was 45 times, with an average of 9 times per game. The guard fouled 25 times, averaging five per game. The centre fouled 22 times, averaging four per game. In the opponent team, The forwards fouled 46 times, averaging 10 per game and the guard fouled 36 times, averaging 8 per game. The centre of the opponent fouled 26 times, averaging 6 per game. From the above data, it can be seen that the difference in the number of fouls between the forwards and centres on both sides is not large, and only the position of the guard has a significant difference. No matter they were in the team of China or the opponent, forward players fouled more than other position players. The statistics of the Olympic men's basketball foul time period are shown in Figure 7:

Figure 7(a) shows the foul time period of the Chinese team, and Figure 7(b) shows the foul time period of the opponent team. As can be seen from Figure 7, the Chinese team averaged 5.4 fouls per game in the first quarter and 6.8 per game in the second quarter, averaged 4.6 in the third quarter and five in the fourth. Opponents averaged five in the first quarter, 6.2 in the second, 6.4 in the third, and 7.4 in the fourth. It can be seen from the above data that the number of fouls by the Chinese team in the second quarter is significantly higher than that in other time periods, and the fourth quarter is the peak of the opponent's foul. The statistics of the foul area of the Olympic men's basketball team are shown in Figure 8:

Figure 7 Statistics of Olympic men's basketball foul time period (see online version for colours)



(a)



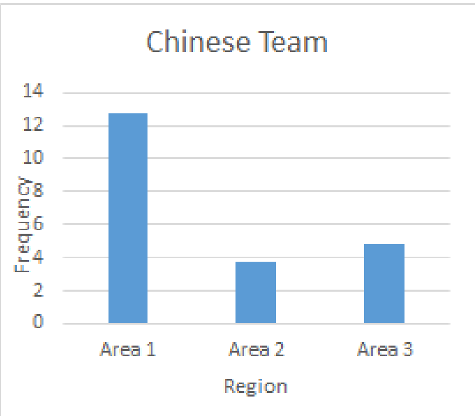
(b)

Figure 8(a) represents the foul area of the Chinese team, and Figure 8(b) represents the foul area of the opponent team. As can be seen from Figure 8, the Chinese team averaged 12.8 fouls in the first area, 3.8 fouls in the second area, and 4.8 fouls in the third area. The opponent team committed 9 fouls in the first and second districts, and 6 fouls in the third district. From the above data, the foul rate of the Chinese team in the first area was the highest among the three zones. In the number of fouls in the second area, the Chinese team was significantly less than the opponent team. In addition to the higher foul rate in the first area, the Chinese team had fewer fouls in the other two areas, and the opponent's fouls in the second and third areas were significantly higher than the Chinese team.

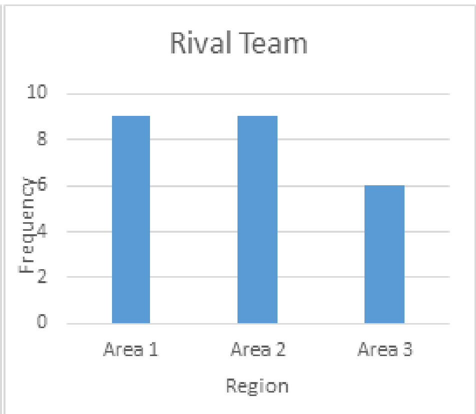
4.2 Comparison of foul types in basketball games

In this section, a more in-depth research and analysis of basketball game fouls based on the research in the previous section were conducted. As in the previous part, the video monitored by the feature extraction technology under machine vision was used to analyse the types of fouls in the game in detail. These analyses were a conduit for players to become aware of their mistakes and tactical problems. The statistics of the nature of fouls in the Olympic men's basketball team are shown in Table 2:

Figure 8 Statistics of Olympic men's basketball foul areas (see online version for colours)



(a)



(b)

As can be seen from Table 2, in the Chinese team, the fouls of blocking totalled 19 times, and the number of illegal use of hands was 64 times. There were 8 fouls of bumping into opponents and 1 foul of playing technique, with 2 unsportsmanlike fouls, in the team of China. The opponent team committed 19 blocking fouls, 82 fouls of illegal use of hands, 5 fouls of bumping into opponents, 3 fouls of playing technique, and 1 unsportsmanlike foul. It can be seen from the above data that the most fouls on both sides were illegal use of hands, with the Chinese team accounting for 71.8% and the opponent team accounting for 78.1%. The statistics of the types of defensive fouls of the Olympic men's basketball team are shown in Table 3:

Table 2 Statistics on the nature of fouls in the Olympic men's basketball team

	Total		*Proportion*	
	China	*Opponent*	*China*	*Opponent*
Foul of blocking	19	19	21.2%	18.1%
Illegal use of hands	64	82	71.8%	78.1%
Foul of bumping into opponents	8	5	8.9%	4.8%
Foul of playing technique	1	3	1%	2.9%
Unsportsmanlike foul	2	1	2.1%	1%

Table 3 Statistics of Olympic men's basketball defensive foul types

	Total		*Proportion*	
	China	*Opponent*	*China*	*Opponent*
Against dribbling foul	16	33	19.5%	34%
Against shooting foul	41	32	50.4%	33%
Against holding foul	5	7	5.9%	7.2%
Against no-ball foul	16	12	19.5%	12.3%
Against catching foul	5	14	5.9%	14.4%
Against passing foul	4	5	4.8%	5.1%

It can be seen from Table 3 that the Chinese team fouled 16 times against dribbling, 41 times against shooting, 5 times against holding the ball, 16 times against no-ball, 5 times against catching, and 4 times against passing. The opponent team fouled 33 times against dribbling, 32 times against shooting, 7 times against holding the ball, 12 times against no-ball, 14 times against catching the ball, and 5 times against passing the ball. It can be seen from the above data that the Chinese team and the opponent team committed more fouls at the point of defence shooting, with the Chinese team accounting for 50.4% and the opponent team accounting for 33%.

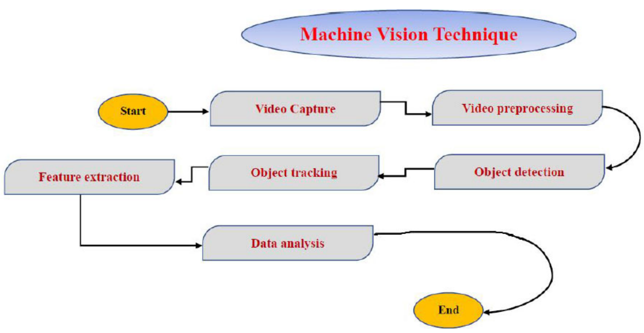
4.3 *Comparison of feature extraction performance of different models on players' foul actions*

In order to better evaluate the performance of models using machine vision technology in extracting features of athletes' foul actions, the CNN model used in this experiment is compared with the LSTM, random forest, decision tree, and support vector machine models, as shown in Figure 9. As can be seen from Figure 9, in this experiment, the accuracy, precision, and recall of the CNN model in extracting the foul action features of basketball players were 98.1%, 96.2%, and 98.5%, respectively, while the support vector machine model performed poorly in this experiment.

Figure 9 shows the comparison of F1 value, AUC, and response speed. As can be seen from Figure 9, the CNN model in this experiment performed the best, with the F1 value reaching 0.94 and the AUC reaching 0.95. Since the model is more complex, with a response speed as high as 1.07 s, but it takes 0.51 s shorter than the LSTM model.

This experiment fully considers the problem of real-time referee application assistance. The response time is good, and the problem of response delay and processing requirements is solved to a certain extent. This article significantly improves the inference speed and reduces latency of the model through in-depth optimisation, including reducing model size, reducing computational complexity, and using high-speed graphics processing units (GPUs).

Figure 9 Feature extraction performance of different models on athletes' foul actions and players' foul actions (see online version for colours)



In summary, the above experimental results indicate that in addition to identifying violations, a range of valuable insights and information can be explored through participants' movement data. First, by analysing players' performance, speed, heart rate and other data, the player's fatigue level can be assessed, which helps the coach make more informed substitution and rotation decisions during the game to maximise the retention of the team. Second, using sports data for biomechanical analysis can improve players' sports skills, discover potential sports injury risks, and formulate corresponding rehabilitation and training plans. Third, through analysis of players' movement trajectories and behaviours, players' strategy selection and execution effects during the game can be revealed, which helps guide the development of more effective game strategies and individual training plans. Fourth, based on sports data, a personalised physical fitness improvement plan can be developed for each team member, including speed, endurance, strength, etc.

5 Conclusions

With the improvement of basketball technique and tactics, the basketball game becomes more intense, resulting in more turnovers. Mistakes of players are also developing in a more subtle and professional direction, which is becoming an important factor in the game. Basketball games are inseparable from referees. The level of referees not only affects the game, but also affects the players' technical and tactical performance. In order to improve basketball, it is not only necessary to improve the technical level of players and coaches, but also to improve the application of referee rules, which is an important aspect to promote the development of basketball. Therefore, whether the referee understands the importance of the rules and correctly adjudicates fouls and violations to ensure fair play and smooth play is crucial for a normal game and the smooth running of both sides. Therefore, the research on machine vision and feature extraction technology in this paper can assist referees to monitor basketball players' foul actions, with important theoretical and practical significance.

Acknowledgements

Conceptualisation, X.J.; methodology, X.J. and C.D.; software, C.D.; validation, C.D.; formal analysis, X.J. and C.D.; investigation, X.J. and C.D.; resources, X.J.; data curation, X.J. and C.D.; writing – original draft preparation, X.J.; writing – review and editing, X.J. and C.D.; visualisation, S X.J. and C.D.; supervision, X.J.; project administration, X.J.; funding acquisition, X.J. All authors have read and agreed to the published version of the manuscript.

References

Dawood, T., Zhu, Z. and Zayed, T. (2017) 'Machine vision-based model for spalling detection and quantification in subway networks', *Automation in Construction*, September, Vol. 81, No. 9, pp.149–160.

Duan, Z.L., Yan, A.P. and Wang, T. (2022) 'Research and analysis on attack and defense levels of teams in NBA playoffs from 2020to 2021', *Journal of Anhui Sports Science*, Vol. 43, No. 4, pp.51–57.

Fernandez-Robles, L., Azzopardi, G. and Alegre, E. (2022) 'Machine-vision-based identification of broken inserts in edge profile milling heads', *Computer Engineering and Applications*, April, Vol. 44, No. 2, pp.276–283.

Gao, S.L. and Hu, L.H. (2022) 'Research on the Influence of the evolution of basketball rules on the development of basketball techniques and tactics', *Contemporary Sports Technology*, Vol. 12, No. 5, pp.191–194.

Guo, G.Z. and Li, J. (2022) 'Analysis of the application of artificial intelligence in contemporary computer information technology', *Electronic Test*, Vol. 36, No. 14, pp.132–134.

Hashimoto, K., Doki, K. and Doki, S. (2017) 'An extraction method of causality patterns between situation and human action with delay time for statistical modeling of human action', *IEEJ Transactions on Electronics Information and Systems*, Vol. 135, No. 7, pp.90–101.

Hong, Z., Xin, M. and Wang, S. (2018) 'End-to-end temporal attention extraction and human action recognition', *Machine Vision and Applications*, Vol. 29, No. 7, pp.1127–1142.

Kartmann, R., Paus, F. and Grotz, M. (2018) 'Extraction of physically plausible support relations to predict and validate manipulation action effects', *IEEE Robotics and Automation Letters*, Vol. 3, No. 4, pp.3991–3998.

Lei, G.Y., Liang, N. and Liu, C.M. (2022) 'Image texture feature analysis and extraction method', *Software Engineering*, Vol. 25, No. 7, pp.5–8.

Liang, X., Li, W.X. and Zhang, H.N. (2022) 'Review of research on human action recognition methods', *Application Research of Computers*, Vol. 39, No. 3, pp.651–660.

Liu, X.C. (2020) 'Motion video content classification algorithm based on big data analysis technology', *Modern Electronics Technique*, Vol. 43, No. 15, pp.70–73.

Liu, X.Y. (2022) 'Visual communication method of multi frame special effects image based on3D laser scanning', *Laser Journal*, Vol. 43, No. 5, pp.161–165.

Liu, Z.L. and Zheng, L. (2023) 'An analysis of referees' penalty calling features in men's 3x3 basketball games at the 14th National Game', *Bulletin of Sport Science and Technology*, Vol. 31, No. 1, pp.42–44.

Maity, S., Chakrabarti, A. and Bhattacharjee, D. (2019) 'Robust human action recognition using AREI features and trajectory analysis from silhouette image sequence', *IETE Journal of Research*, Vol. 65, No. 2, pp.236–249.

Miao, Y., Zhao, Z.S., Yang, Y.L., Xu, N., Yang, H.R. and Sun, Q. (2020) 'Survey of image captioning methods', *Computer Science*, Vol. 47, No. 12, pp.149–160.

Nouri-Ahmadabadi, H., Omid, M. and Mohtasebi, S.S. (2017) 'Design, development and evaluation of an online grading system for peeled pistachios equipped with machine vision technology and support vector machine', *Information Processing in Agriculture*, Vol. 4, No. 4, pp.333–341.

Seol, J.W., Yi, W. and Choi, J. (2017) 'Causality patterns and machine learning for the extraction of problem-action relations in discharge summaries', *International Journal of Medical Informatics*, February, Vol. 98, No. 2, pp.1–12.

Song, L.L., Li, Q. and Xu, H.L. (2020) 'Research on the influence of the changes of the new rules of basketball on the application of basketball technology', *Journal of Guangzhou Sport University*, Vol. 40, No. 4, pp.81–84.

Tsai, D.M. and Hsieh, Y.C. (2017) 'Machine vision-based positioning and inspection using expectation–maximization technique', *IEEE Transactions on Instrumentation and Measurement*, Vol. 66, No. 11, pp.2858–2868.

Wang, J.P., Zhu, R., Zhang, G., He, X.B. and Cai, R.P. (2022) 'Image flow measurement based on the combination of frame difference and fast and dense optical flow', *Advanced Engineering Sciences*, Vol. 54, No. 4, pp.195–207.

Wang, L.L. and Chen, W.F. (2020) 'Study on the requires of offensive integrity for the improvement of basketball game rules', *Physical Education Review*, Vol. 39, No. 4, pp.99–100.

Wang, Q., Chen, B. and Zhu, D. (2017) 'Machine vision-based selection machine of corn seed used for directional seeding', *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 48, No. 2, pp.27–37.

Yang, H., Hou, J.L., Lu, Z. and Xiang, L.X. (2022) 'Classification of grades of contemporary world men's basketball team and characteristics of its winning indexes', *Sichuan Sports Science*, Vol. 41, No. 5, pp.73–78.

Yu, W.Q., Yu, J., Bai, M.Y., Xiao, C.B. (2021) 'Video object detection using fusion of SSD and spatiotemporal features', *Journal of Image and Graphics*, Vol. 26, No. 3, pp.542–555.

Zhang, K. and Zhang, L. (2018) 'Extracting hierarchical spatial and temporal features for human action recognition', *Multimedia Tools and Applications*, Vol. 77, No. 13, pp.16053–16068.

Zhang, Y., Hong, G.S. and Ye, D. (2020) 'Powder-bed fusion process monitoring by machine vision with hybrid convolutional neural networks', *IEEE Transactions on Industrial Informatics*, Vol. 16, No. 9, pp.5769–5779.