CrossMark

Fitness training driven by image target detection technology

Lei Wang, Tuojian Li[*], Jinhai Sun and Xianliang Zhang

Abstract

The fitness training system needs to capture the training staff dynamics in real time, but it is difficult to capture the training staff dynamics during the actual training process. Based on this, this study uses the physical characteristics of fitness trainers as indicators for image target detection. According to the human body will dissipate more heat during the fitness process, this study uses infrared capture as the basis of image capture detection technology, uses FCM clustering algorithm as the fuzzy image background segmentation algorithm, and uses k-means clustering analysis to study the gray histogram and propose a composite classification feature tracking method for trainer image tracking. Combined with the experimental research, the research shows that the research method utilizes the advantages of the composite classification feature to improve the detection rate of the human target. Therefore, it is a real-time and very effective infrared image human detection algorithm.

Keywords: Target detection, Fitness, Training

1 Introduction

Detecting and tracking moving targets in fitness videos can provide a lot of help for professionals. For example, the fitness instructor can use the relevant data extracted from the fitness video to conduct the fitness system research, and the fitness personnel's trajectory is mapped to the detection model through the detection, tracking, and classification of the fitness personnel, thereby analyzing the fitness strategy. The athlete can improve the training level through three-dimensional reconstruction of these data and realistic design, simulation, and analysis of technical actions. Simultaneously, the athlete can analyze the relevant video of the fitness video captured by the camera to extract the target of the region of interest, thereby making a more accurate analysis and maximizing the training effect. At present, image detection technology has been applied to many aspects, but research on fitness training is still rare. Therefore, it needs to be extended on the basis of relevant research.

The Fitness target detection work for Fitness videos mainly includes site detection, player detection and ball detection. The main extraction methods are roughly divided into three categories: optical flow method, frame

difference method, and background difference method. Kong proposed a method for site color adaptive detection of the course area, which looks for the main area in the histogram and then estimates the mean and variance of the area. At the time of detection, the two-color spaces of RGB+HIS are used to complement each other, the former as the control space and the latter as the basic space [1]. Wang uses the Gaussian mixture model to automatically obtain the color of the field in fitness video using the EM algorithm [2]. In the player recognition, Xiao et al. initially tested the automatic modeling and detection of the color of the two players' jerseys [3]. Using the site segmentation image as a mask, Purnomo et al. used the Ada Boost method to detect players, and its performance was greatly improved compared to the previous method. It also collects samples through automatic player detection, uses unsupervised clustering to learn jersey colors, and then classifies players, which achieves an automatic high-performance classification of both players and referees. Early ball testing basically used color template matching technology [4]. Luo et al. proposed a new idea. Firstly, the size of the sphere was inferred according to the size of the player's area, then the aspherical area was filtered out, and finally the Kalman filter was used to track the area containing the ball [5]. Suwa et al. used two stages to detect and identify the

* Correspondence: tuojianli@yeah.net
Shandong University, Jinan 250061, Shandong, China

Wang *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:102

Page 2 of 9

ball. First, the Hough detection algorithm was used to detect the area that may contain the ball, and then the neural network classifier was used to find the area containing the ball. In order to further increase the speed, this method introduces background subtraction method and ball tracking technology [6].

In fitness video, another interesting task is the tracking of the target, mainly tracking the players and the ball. According to the matching principle, the existing tracking algorithms are divided into four categories: model-based tracking, contour-based tracking, region-based tracking, and feature-based tracking. In order to improve the search speed of the target, mathematical tools are used such as Kalman filter, Condensation algorithm, particle filter algorithm, Mean Shift algorithm, dynamic Bayesian network, and so on. Thayaparan uses a trajectory-based method to track football, uses the Viterbi algorithm to detect and track the ball, and uses the least squares method to obtain the difference function of the ball's motion trajectory. According to the interpolation function, some errors in the process of detecting and tracking the ball by the Viterbi algorithm are eliminated, and the ball position information at the missed detection is supplemented. This method does not require an accurate template of the ball and has good robustness and can also achieve good results when the ball is occluded [7]. Willett et al. use 3D modeling to track and use the generalized frustum, elliptical column, and ball 3D model to describe the structural details of the human body, which can accurately restore the target's trajectory and shape, and also restore 3D information under target occlusion. However, the establishment of a three-dimensional target model requires a large number of model parameters, and the model matching process is more complicated. Therefore, this method is only suitable for a small number of specific types of target tracking [8]. Li uses Kalman filters to track multiple targets. It can track new targets and track original targets and corrects the trajectories by least squares [9]. In order to solve the tracking problem of the occlusion player, Rozantsev et al. used the region tracking combined with the color histogram and the particle filtering algorithm to achieve the player's tracking, overcome the incomplete occlusion of the target, and improve the accuracy of the target tracking. Since particle filters have excellent effects on occlusion tracking, in recent years, particle filter tracking algorithms have also achieved a lot of results. Although the particle filter has good tracking effect and robustness, it is not studied in this paper because of the large amount of computation, poor real-time performance, and high requirements for machine configuration [10].

It can be seen that detecting and tracking moving targets in fitness videos is of great significance, which is conducive to the automatic analysis of Fitness videos, thus providing advanced tools and means for fitness training. In fitness videos, athletes' movements are irregular, and they change in posture during exercise. The particularity of fitness video poses many challenges for moving target detection and tracking. Therefore, the main research purpose of this paper is to analyze and research the current motion detection and tracking methods and propose an effective algorithm to realize the detection and tracking algorithm of moving targets in volleyball video and use the tracking target position information to generate motion. Therefore, the main research purpose of this paper is to propose an effective algorithm to realize the detection and tracking algorithm of moving targets in volleyball video by analyzing the current motion detection and tracking methods and use the tracked moving target position information to generate the moving trajectory of moving targets, which is convenient for higher-level video data analysis and behavior decision making.

2 Research methods

This study mainly analyzes the image detection of the fitness system. However, during the fitness process, the human body will dissipate more heat, which is different from the general state. Therefore, this study uses infrared sensing for image acquisition.

2.1 Fuzzy clustering segmentation

In grayscale images, the distribution of background and objects tends to overlap with each other, which makes the spatial information of pixels and their adjacent domains not fully utilized. Based on this, we can use image fuzzy clustering segmentation method based on two-dimensional histogram. In this method, the gray information of the infrared image and the spatial information between the neighborhoods are combined to construct a two-dimensional histogram, which makes the object and the background easier to distinguish.

The gray level of the original image $I(x, y)$ is set to L and the size is set to $N \times N$. The number of gray levels and the size of $h(x, y)$ obtained by $I(x, y)$ using 3×3 or 5×5 dot matrix are the same as $I(x, y)$. At this point, the two images can form a two-tuple consisting of the gray level of the pixel and the average gray level of the neighborhood of the pixel. Each binary group belongs to a point on a two-dimensional plane, and all points are $L \times L$. The frequency at which the two groups (s, t) appear is $f_{\_}(s,t)$, which represents the number of two-dimensional points when the gray level of $I(x, y)$ is s and the gray level in $h(x, y)$ is t.

Wang et al. EURASIP Journal on Image and Video Processing (2018) 2018:102

Page 3 of 9

$$\sum_{s=0}^{L-1}\sum_{t=1}^{L-1}f_{s,t} = N \times N \qquad (1)$$

The FCM clustering algorithm is a gray-based iterative optimization process. It continuously iterates through the clustering center and membership function to find the cluster center value and membership value which can minimize the objective function and use this as the threshold image for optimal segmentation. Assume that the training sample set can be expressed as Eq. (2). c is the predetermined number of categories, and $v_i(i = 1, 2, \Lambda c)$ is set. $u_{ik}\{i = 1, 2, \Lambda c_k = i = 1, 2, \Lambda n\}$ is the membership function of the k_{th} sample to the i_{th} class, and there is a relationship as shown in the Eq. (3). The objective function of FCM can be expressed as Eq. (4).

$$X = \{x_k, k = 1, 2\Lambda, n\} \qquad (2)$$

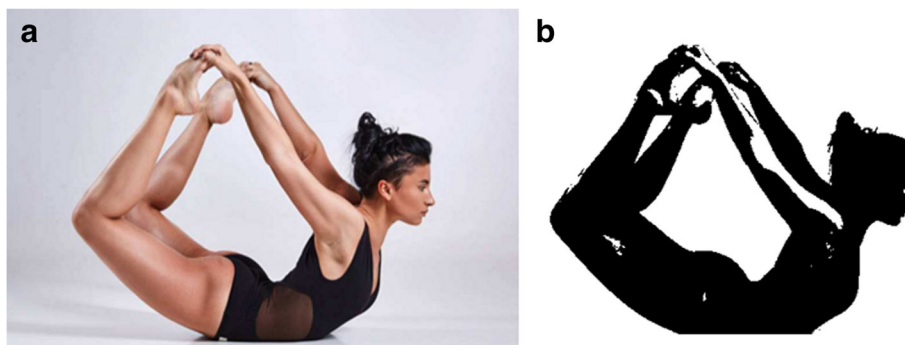$$0 \le u_{ik} \le 1, 0 \le \sum_{K=1}^{N}u_{ik} \le n \qquad (3)$$

$$J_m(U, v) = \sum_{i=1}^{c}\sum_{k}^{n}u_{ik}^{m}||x_k - v_i||^2 \qquad (4)$$

Among them, $v = (v_1, v_2, \Lambda, v_c)$ and $U = \{u_{ik}\}$ belong to the fuzzy weighted index, and the segmentation diagram shown in Fig. 1 can be obtained by the objective function.

In order to effectively segment the target area of the human body, by analyzing the gray scale distribution in multiple infrared images, it can be seen that the infrared image is insensitive to light compared to visible light images and is only sensitive to temperature. Therefore, its grayscale distribution is relatively obvious. For some scenes, the gray distribution on the infrared image has a single-peak distribution, while the gray scale distribution of the image under visible light conditions is messy and has multiple peaks. The complex background image shown in Fig. 2 is selected, and the results obtained are shown in Fig. 2.

By comparison, the target temperature of the human body in the infrared image tends to be stable, but due to the different wearing of the human body, the distribution is not a fixed value, which is distributed over a small temperature range. On the image, it appears that the target gray value is distributed within a small segment of a certain gray value. Considering that the unimodal distribution background model cannot simultaneously deal with noise, illumination changes, and other factors in the image, we can describe the pixel feature values in the image. The eigenvalue is the brightness of the pixel. At the same time, we model the brightness of the pixels in the same position in the continuous image sequence in the time domain as a series of time series propagation in the image. In the modeling process, by using multiple weighted processes to simulate the background of complex changes to eliminate the sudden changes of light in the actual scene, as well as the impact of roadside tree swing, the robust tracking of moving targets in complex background is well realized. It has many advantages, such as low computational complexity, weakened background interference sensitivity, and adaptability. Although the gray scale distribution of the infrared image is a single-peak distribution, the segmentation of the foreground object is not well accomplished. Related research shows that although the adaptive hybrid model segmentation algorithm is a fast segmentation algorithm, it can also segment the foreground target, but the effect is not very good.

Therefore, in this study, the histogram is first divided into multiple regions for multi-cluster analysis, and then the distribution of clustering centers of different categories after clustering is analyzed. Finally, the maximum point is found from it and the target is well separated from the image as the image segmentation threshold. The K-means clustering algorithm is an iterative optimization process that assigns each sample to the class that is closest to its nearest neighbor class. The algorithm has the advantages of relatively simple and fast

Fig. 1 Segmentation result graph. **a** Fitness scene **b** Segmentation graph

Fig. 2 Image segmentation in a complex background. **a** Fitness picture **b** Unimodal distribution
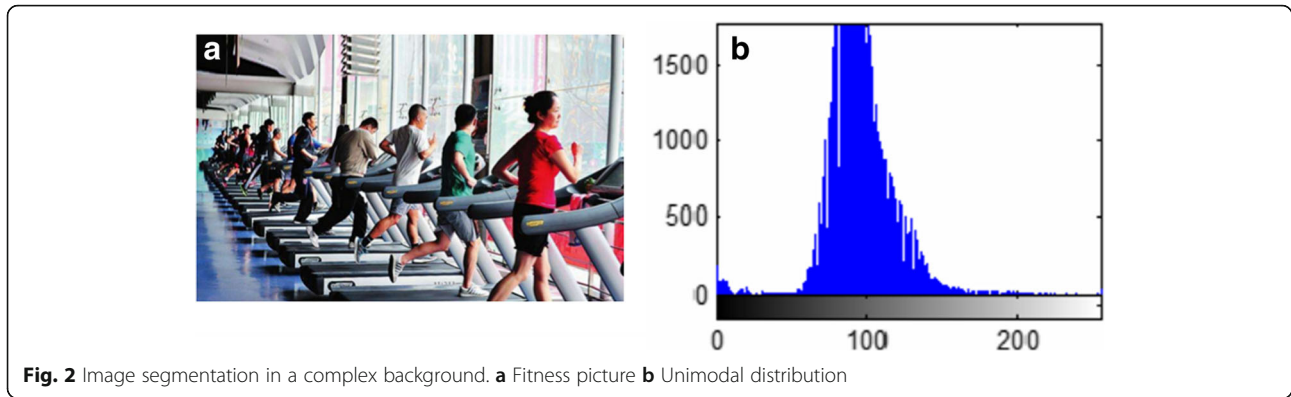
clustering speed. The K-means clustering center analysis algorithm execution process is: First, it is divided into multiple initial cluster centers by the standard deviation of image gray distribution in the effective area of the histogram. Secondly, k-means clustering is performed on different gray values. Finally, the relationship between the cluster space K_i and its two adjacent cluster spaces K_{i-1} and K_{i+1} is considered. The principle of the K-means clustering center analysis method is as follows: the cluster center values before and after clustering are u_i^0, u_{i-1}^0, u_{i+1}^0, and u_i, u_{i-1}, u_{i+1}, respectively. u_i^0, u_{i-1}^0, u_{i+1}^0 have a linear relationship before clustering, and the relationship satisfies:

$$l^0 = \frac{u_{i+1}^0 - u_i^0}{u_i^0 - u_{i-1}^0} \tag{5}$$

After clustering, the relative relationship of u_i, u_{i-1}, and u_{i+1} can be expressed as:

$$l = \frac{u_{i+1} - u_i}{u_i - u_{i-1}} = \frac{\left(u_{i+1}^0 - u_i^0\right)}{\left(u_i^0 - u_{i-1}^0\right)} \frac{(\Delta u_{i+1} - \Delta u_i)}{(\Delta u_i - \Delta u_{i-1})} \tag{6}$$

If the pixels in the cluster space K_{i-1}, K_i, K_{i+1} belong to the same type of target, then the value of Δu_{i11}, Δu_i, Δu_{i+1} will be very small after clustering, so that $l \approx 1$ can be obtained from the analogy of Eqs. (5) and (6). The meaning is that the linear relationship between the three types of cluster centers belonging to the same type of target before and after clustering should not be destroyed, and the relative change is small. If the central value of one of the cluster spaces has a large variation range before and after clustering, the linear relationship will be destroyed at this time, and the trend of the cluster center value will definitely show a significant turning. The target grayscale can be segmented by the actual grayscale threshold of the category pixel set at the turning point. In order to reduce the computation time, we usually use the method of averaging the center values of

two adjacent spatial clusters at the turning point to select the threshold.

The absolute value of the relative incremental difference of the cluster center values is used to find the center of the histogram cluster turning point. The relationship is expressed by Eq. (7):

$$CR = \left| \frac{u_{i+1} - u_i}{u_{i+1}} - \frac{u_i - u_{i-1}}{u_i} \right| = \frac{u_i^2 - u_{i+1} u_{i-1}}{u_{i+1} u_i} \; i$$
$$= 2, \Lambda \, N{-}1 \tag{7}$$

A category corresponding to the maximum point of CR near the saturation direction is used as a turning point. After the turning point is selected, the two adjacent cluster center values are averaged at the point, and the average value is used as the threshold value to perform threshold binarization of the infrared image to obtain the foreground target. Since the brightness of the human target is generally high in the infrared image, we can reduce the amount of data that can be processed by analyzing only the pixels above the average brightness of the image. This method can not only speed up the process of the algorithm but also make the transition point of the cluster center more prominent.

2.2 Improved k-means clustering analysis
The core part of the algorithm is the selection of the turning point of the cluster center trend after k-means clustering analysis of the gray histogram. An important factor affecting the outcome is the choice of the measure function. We assume that when the clustering N cluster centers no longer change and are in a straight line, that is:

$$\Delta = u_2 - u_1 = \Lambda = u_i - u_{i-1} = u_{i+1} - u_i$$
$$= u_N - u_{N-1} \; i = 2, \Lambda \, N{-}1 \tag{8}$$

When Δ is a constant, the theoretical linear relationship is not destroyed, and there is no turning point.

However, due to $u_1 < \Lambda < u_i < u_i < u_N$, Eq. (7) can be converted to $CR(i) = |\frac{\Delta^2}{u_{i+1}u_i}|$, and then $CR(2) > CR(3) > \Lambda > CR(N-1)$ can be known, that is to say, the turning point is always the second category and the theory is different. Therefore, this paper improves the measure function of determining the turning point of the histogram cluster center. We assume that the K_i center to be tested is the turning point of the histogram cluster center trend; then, there is

$$CR = \left| \frac{u_{i+1}-u_i}{u_{i+1}} - \frac{u_i-u_{i-1}}{u_i} \right| = \frac{u_{i+1} + u_{i-1}}{u_i} - 2 \, i$$
$$= 2, \Lambda \, N-1 \tag{9}$$

If $CR = 0$ at this time when the linear relationship is not destroyed, the offset between the cluster center value that the measure function is transformed into K_i and the theoretical center value can be used as a measure function:

$$d = |u_i - u_{it}| = \left| u_i - \frac{u_{i+1} + u_{i-1}}{2} \right| \quad i = 2, K, N-1 \tag{10}$$

At this time, d is $N-2$ column vector $d(a)$ corresponds to $u(2)$. The process of determining the turning point is transformed into a process of finding the maximum point of the criterion d. Once the turning point is determined, in order to reduce the error and reduce the amount of calculation, we usually take the average value of the two adjacent cluster center values at the turning point as the threshold value and binarize the whole image. Since the brightness of the human target in the infrared pedestrian image is generally higher than the background brightness, we can also analyze only the pixels above the average brightness of the image, thereby reducing the amount of data that can be processed and improving the efficiency of the algorithm. In addition, the trend of the resulting cluster center is more obvious.

2.3 Composite classification feature tracking

For the probability density function $f(x)$, the $f(x)$ kernel function estimation expression for a set of sample points A in the known d-dimensional space is:

$$\hat{f}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x_i-x}{h}\right) w(x_i)}{h^d \sum_{i=1}^{n} w(x_i)} \tag{11}$$

Among them, $w(x\_i)$ represents a weight assigned to the sample point x_i and $K(x)$ is a kernel function and satisfies $\int k(x)dx = 1$. The profile function $k(x)$ of the kernel function $K(x)$ is defined such that $K(x) = \cdot k(x)$, i.e., $g(x) = \cdot k(x)$, whose corresponding density kernel function can be expressed as $G(x) = g(||x||^2)$. The

estimate of the gradient $\Delta f(x)$ of the probability density function $f(x)$ is:

$$\hat{\Delta} f(x) = \Delta \hat{f}(x)$$
$$= \frac{2 \sum_{i=1}^{n} (x-x_i) k' \left(\left\| \frac{x_i-x}{h} \right\|^2 \right) w(x_i)}{h^{d+2} \sum_{i=1}^{n} w(x_i)} \tag{12}$$

From the above definition, $g(x) = k(x)$, $G(x) = g(|(|x|)|^2)$, the above formula can be converted to obtain the formula (13):

$$\hat{\Delta} f(x) = \frac{2}{h^2} \left[\frac{\sum_{i=1}^{n} G\left(\frac{x_i-x}{h}\right) w(x_i)}{h^d \sum_{i=1}^{n} w(x_i)} \right]$$
$$\times \left[\frac{\sum_{i=1}^{n} (x_i-x) G\left(\left\| \frac{x_i-x}{h} \right\|^2 \right) w(x_i)}{\sum_{i=1}^{n} G\left(\frac{x_i-x}{h}\right) w(x_i)} \right] \tag{13}$$

From the above, we can draw:

$$M_h(x) = \frac{1}{2} h^2 \frac{\nabla \hat{f}_k(x)}{\hat{f}_G(x)} \tag{14}$$

It can be seen from the Eq. (14) that the offset mean vector M_h (x) calculated by the kernel function G at the x point is proportional to the gradient of the probability density function $(f\_k)\hat{} (x)$ estimated using the kernel function K after normalization. Among them, the normalization factor is the probability density estimate of the kernel function G at point x. Therefore, the mean shift vector $M_h(x)$ always points to the direction in which the probability density increases the most.

By placing the x of Eq. (13) outside the summation number, the following equation can be obtained and then the first term on the right side of the above formula is denoted as $m\_h(x)$.

$$M_h(x) = \frac{\sum_{i=1}^{n} G\left(\frac{x_i-x}{h}\right) w(x_i) x_i}{\sum_{i=1}^{n} G\left(\frac{x_i-x}{h}\right) w(x_i)} \tag{15}$$

$$m_h(x) = \frac{\sum_{i=1}^{n} G\left(\frac{x_i-x}{h}\right) w(x_i) x_i}{\sum_{i=1}^{n} G\left(\frac{x_i-x}{h}\right) w(x_i)} \tag{16}$$

Given an initial point x, the kernel function G(X), and the tolerance error ε, the mean shift algorithm is then looped through the following three steps until the end condition is met. (1) $m_h(x)$ is calculated; (2). $m_h(x)$ is assigned to x; (3) if $||m_h(x) - x|| < \varepsilon$, the loop ends, and if not, continue with (1). On this basis, the infrared human tracking algorithm flow can be expressed as follows: (1) The target template is selected while the particle set is initialized. (2) The

Wang *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:102

Page 6 of 9

current frame is acquired. (3)Through the state transition model, the particle state is predicted to obtain the particle set $\{(s\_k^{\wedge}(i), 1/N)\}\_(i = 1)^{\wedge} N$. (4) According to the infrared target brightness characteristics and motion information, each particle is drifted by the mean shift algorithm to obtain a new particle set $\{(s\_k^{\wedge'}(i), 1/N)\}\_(i = 1)^{\wedge} N$. (5) Through the target brightness feature and the motion information feature, the observation model is established, and the particle weight $\{w\_k^{\wedge}((i))\}\_(i = 1)^{\wedge}N$ is calculated. (6) The particle weights are normalized and resampled. (7) The current frame target state is estimated; (8) The next frame is collected.

3 Results

For the research method and the measurement function of this paper, the unfiltered image and the 3×3 median filtered image were respectively segmented. The corresponding improvement function is obtained from the reference, and combined with the algorithm of the present study, the results are shown in Table 1.

Through the detection of the fitness process of the elderly, the results are shown in Fig. 3. Among them, (a)–(e) represent the original infrared image, the pre-improved measure function, the median filter + the improved pre-measurement function, the improved post-measurement function, and the median filter + the improved measure function, respectively.

In the classification process, the detected human candidate targets are normalized to 2050 pixels. These candidate targets fall into two categories: the human target is marked as 1, and the non-human target is marked as 0. In the experiment, the selection of support vector parameters C, γ and histogram series parameters has a great influence on infrared human body detection. In order to better complete infrared human detection, we first need to set numerical values for them. In this paper, we set the search space of C, γ as $C : 2^{-5} - 2^{11}$, $C : 2^{-5} - 2^{1}$. Then, we obtained 240 human samples and 240 non-human samples as training samples, 150 human samples and 150 non-human samples as test samples by manual sorting in the

image of two test sets. At the same time, we randomly divide the training samples into ten groups for cross-contrast test and obtain the detection rate of the test samples under different histogram series, as shown in Table 2, and draw the obtained results into the statistical graph of Fig. 4.

Figure 5 shows the effect of infrared human detection results that partially used composite classification features for two test sets. From the detection results, the composite classification feature proposed in this study can better eliminate the complex background and halo interference in the infrared human detection algorithm, correctly detect the human target, and have a certain effect on the adhesion between human bodies.

4 Analysis and discussion

From Fig. 3, we can see that the infrared maps (a), (d) have relatively more noise and the difference in threshold is also larger. However, the change range of threshold variation of image (b), (c), (e) before and after filtering is not large, indicating that it is mainly affected by the measure function. From the effect of the figure, the improved measurement function segmentation results are relatively complete and have good noise immunity.

In the case of linear inseparability, we map the sample to a high-dimensional feature space and use the original space function to implement the inner product operation in the high-dimensional feature space. In this way, we transform the nonlinear problem into a linear problem of another space to obtain the attribution of a sample. The theory of functionals shows that as long as a kernel function satisfies the Mercer condition, it corresponds to the inner product of a certain space. Therefore, as long as we can find the appropriate inner product function on the optimal classification plane, we can solve the linear indivisible classification problem. The key technology of the support vector machine is the selection of the kernel function. Low-dimensional space vector sets are generally difficult to divide. In this case, we need to map it to high-dimensional space. However, although this

Table 1 Comparison of segmentation threshold before and after algorithm improvement

Image serial number	Pre-improved measure function	Median filtering + pre-improved measure function	Improved measure function	Median filtering + improved measure function
1	152.120	118.410	118.410	118.410
2	139.220	139.000	122.150	122.160
3	144.140	143.490	125.460	125.530
4	156.190	129.060	129.060	129.110
5	110.380	110.200	110.000	110.100

Fig. 3 a–e Fitness tracking image
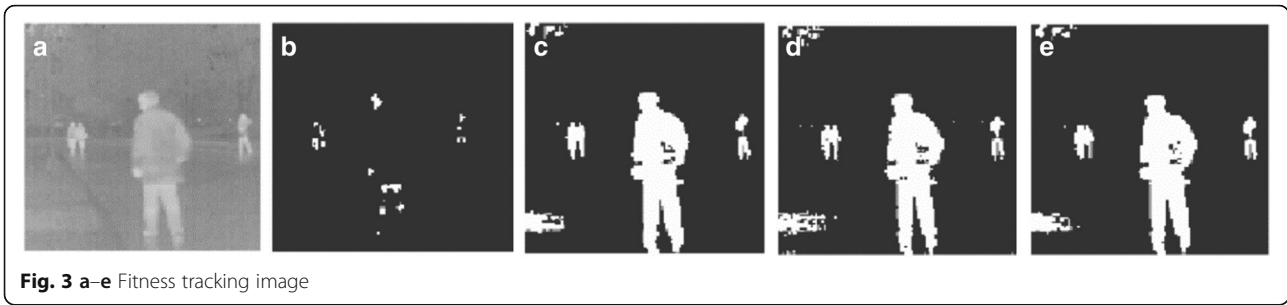
can solve the linear inseparable classification problem, it also brings about an increase in the computational complexity. For this, it needs to be solved by a kernel function. We can say that as long as we choose the appropriate kernel function, we can get the classification function of high-dimensional space. Grayscale information is a commonly used feature in infrared images that is often used to detect candidate target regions and template human targets from test sets and training sets. The template human target is selected by constructing an SVM model for the training set for three different kernel functions. These selected template human targets must be normalized prior to entering the classifier for human and non-human targets.

It can be seen from Table 2 that when the histogram is 25, the detection rate of positive and negative samples reaches the highest point. At this time, the obtained SVM classifier has a penalty factor $C = 2048$ and a kernel function width $\gamma = 2048$.
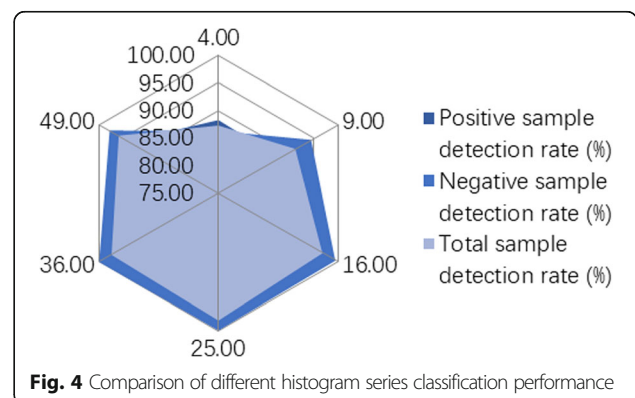
As far as the basic idea is concerned, mathematical morphology processing is to measure and extract the objects in the image with certain connected structural elements, which not only simplifies the image data but also maintains the basic physical features in the image, thus achieving the purpose of image analysis and recognition. This method generally processes the binarized image. Basic mathematical morphology operations include corrosion, expansion, and open and closed operations. In this study, the histogram is first divided into multiple regions for multi-cluster

analysis, and then the distribution of clustering centers of different categories after clustering is analyzed. Finally, the maximum point is found and used as the image segmentation threshold to separate the target from the image. The K-means clustering algorithm is an iterative optimization process that assigns each sample to the class that is closest to its nearest neighbor class. The algorithm has the advantages of relatively simple and fast clustering speed. The K-means clustering center analysis algorithm execution process is as follows: First, the tentative linear division into a plurality of initial cluster centers in the effective region of the histogram is performed by using the standard deviation of the gray scale of the image as the brightness. Secondly, k-means clustering is performed on different gray values. Finally, the relationship between the cluster space K_i and its two adjacent cluster spaces K_{i-1} and K_{i+1} is considered.

It can be seen from Fig. 5 that the classification feature algorithm proposed in this study can better eliminate the complex background and the halo interference, correctly detect the human target, and have a certain effect on the adhesion between human bodies. The application of composite classification features in infrared human target detection can not only improve the correct detection rate, but also reduce the missed detection rate and false detection rate to varying

Table 2 Detection rate of test samples under different histogram series

Histogram series	4.00	9.00	16.00	25.00	36.00	49.00
Positive sample detection rate (%)	88.30	87.80	94.30	96.50	94.80	93.90
Negative sample detection rate (%)	86.30	94.40	99.50	99.90	100.00	97.80
Total sample detection rate (%)	87.30	91.10	96.80	98.10	97.40	95.90

Fig. 4 Comparison of different histogram series classification performance

Wang *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:102

Page 8 of 9

Fig. 5 **a**–**c** Composite classification feature detection results

degrees. In the experiment, the combination of the direction gradient histogram feature and the other two features produces different results. For example, when we keep the characteristics of the direction gradient histogram unchanged and increase the physical features, the number of false detections and missed inspections is reduced, but the number of false positives is significantly reduced. However, when the inertia feature is increased, the number of missed inspections is significantly reduced, but the number of false positives is increased. When the three features are used simultaneously, the number of missed inspections and the number of false positives are significantly reduced. Therefore, the use of composite classification features can complement each other and comprehensively describe the characteristics of the human body in the infrared image, thereby improving the detection rate of the human body.

Through the above analysis, it can be seen that the method utilizes the advantages of the composite classification feature and improves the target detection rate of the human body, which is a real-time and effective detection infrared image human detection algorithm.

5 Conclusions

This study mainly analyzes the image detection of fitness systems. However, during the fitness process, the human body will dissipate more heat, which is different from the general state. Therefore, this study uses infrared sensing to collect images. In this study, an image fuzzy clustering segmentation method based on two-dimensional histogram is used. In this method, the gray information of the infrared image and the spatial information between the neighborhoods are combined to construct a two-dimensional histogram, which makes the object and the background easier to distinguish. In this study, the histogram is first divided into multiple regions for multi-cluster analysis, and then the distribution of clustering centers of different categories after clustering is analyzed. Finally, the maxima points are found and are used as the image segmentation threshold to separate the target from the image. In the case of linear inseparability, we map the sample to a high-dimensional feature space and use the original space function to implement the inner product operation in the high-dimensional feature space. In this way, we transform the nonlinear problem into a linear problem of another space to obtain the attribution of a sample. Combined with the experimental research, it can be seen that the method utilizes the advantages of the composite classification feature and improves the target detection rate of the human body, which is a real-time and effective detection infrared image human detection algorithm.

About the Authors
Lei Wang, born in Jinan, Shandong Province, is a Doctoral student in Shandong University. Her main research directions are youth fitness and neural network.
Tuojian, Li, born in Shandong Province, is an Assistant Researcher, and his main research directions are Physical activity and Youth fitness.
Jinhai Sun, born in Shandong Province, is a Professor in Shandong University. His main research directions are sports management and system engineering.
Xianliang Zhang, born in Shandong Province, is an assistant researcher. His research interests include sports human science.

Authors' contributions
LW, TL, and JS designed the research, LW performed the research, and XZ wrote the paper. JS and XZ contribute same to this article and could be

consider as common corresponding author. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. W. Kong, J. Yu, Y. Cheng, et al., Automatic detection technology of sonar image target based on the three-dimensional imaging[J]. J. Sens. **2017**(2), 1–8 (2017)
2. Y. Wang, Y. Li, Representation and organization for spatial data in LBS[J]. J. Earth Sci **25**(3), 544–549 (2014)
3. D. Xiao, J. Fu, X. Deng, et al., Design and test of remote monitoring equipment for bactrocera dorsalis trapping based on internet of things[J]. Trans. Chin. Soc. Agric. Eng. **31**(7), 166–172 (2015)
4. F.A. Purnomo, P.I. Santosa, R. Hartanto, et al., Implementation of Augmented Reality Technology in Sangiran Museum with Vuforia[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing **333**(1), 012103 (2018)
5. G. Luo, A small target detection technology for forest-fire prevention based on infrared monitoring system[J]. Boletin Tecnico **55**(9), 140–147 (2017)
6. K. Suwa, K. Yamamoto, M. Tsuchida, et al., Image-based target detection and radial velocity estimation methods for multichannel SAR-GMTI[J]. IEEE Trans. Geosci. Remote Sens. **55**(99), 1–14 (2017)
7. T. Thayaparan, L. Stanković, I. Djurović, Micro-Doppler-based target detection and feature extraction in indoor and outdoor environments[J]. Frequenz **345**(3–4), 700–722 (2015)
8. R.M. Willett, M.F. Duarte, M. Davenport, et al., Sparsity and structure in hyperspectral imaging: sensing, reconstruction, and target detection[J]. IEEE Signal Process. Mag **31**(1), 116–126 (2014)
9. H. Li, Limited magnitude calculation method and optics detection performance in a photoelectric tracking system[J]. Appl. Opt. **54**(7), 1612–1617 (2015)
10. A. Rozantsev, V. Lepetit, P. Fua, On rendering synthetic images for training an object detector ☆[J]. Comput. Vis. Image Underst. **137**, 24–37 (2015)