




Low-resolution Face Recognition and Sports Training Action Analysis based on Wireless Sensors

Shiwu Chen^{1*} 

¹JiangXi University Of Engineering, Xinyu, JiangXi ,338000, China

Corresponding author: Shiwu Chen: chenshiwu197204@163.com

Abstract: In order to improve the accuracy of sports training action analysis, this paper combines wireless sensor technology and low-resolution recognition technology to construct a low-resolution face recognition and sports training action analysis system based on wireless sensors. The use of intelligent statistical features to construct a dictionary can reduce the dimension of the dictionary and improve the speed of feature detection. Moreover, intelligent feature recognition has the characteristics of spatial locality, direction selectivity and illumination invariance, which makes the constructed dictionary more robust to the influence of illumination and noise, thereby improving the recognition rate. The simulation study shows that the low-resolution face recognition and sports training action analysis system based on wireless sensors proposed in this paper has good practical results.

Keywords: Wireless sensor; low resolution; face recognition; physical training; action analysis.

DOI: <https://doi.org/10.14733/cadaps.2023.S12.152-171>

1 INTRODUCTION

For a human motion system with a relatively fixed training system and less degrees of freedom, the static analysis method is more effective. However, with the increase in the complexity of the training system, modern intelligent equipment tends to develop in the direction of multiple degrees of freedom, and the system movement is flexible and changeable. Generally, it is difficult to directly capture the relatively stable space training system vector loop of the human body system, and the drawbacks of the static analysis method become increasingly prominent. In view of the complex human training system, difficult analysis, and complex working conditions, some experts and scholars have proposed a meta-action theory research method, which studies the performance of the whole machine from the action layer, and gets rid of the limitations of the research method starting from the whole or individual limbs.

Meta-action is the smallest motion functional unit of human motion system, which generally includes several nodes, and the motion is output in the form of rotation or movement through the unit motion output element. The meta-action unit training system is relatively simple, and the size of the training system unit is moderate. Studying the performance of the whole machine from the meta-action layer greatly simplifies the workload of analyzing the complex human product training system. The motion analysis method based on the meta-action theory has obvious dynamic characteristics, which can effectively make up for the characteristics of the static analysis method. The dynamic analysis method based on meta-action has achieved good application results in the research of typical CNC machine tools, which provides theoretical and practical support and reference for the implementation of the method based on meta-action theory for complex human kinematics research. . From the perspective of method and concept, taking the movement process of meta-action as the object of attention, the complex movement is simplified and transparent, and the analysis method is more refined. On the other hand, the movement form of meta-action is simple and intuitive, and the technical difficulty in the process of actual movement monitoring and analysis is also less, and it is easier for people to accept and use.

The angle and shape of each joint of the human body control the macro posture of the human skeleton. Human joints are organizational structures connecting adjacent rigid bodies (limbs) of the human body, and each joint node constrains the relative motion of two adjacent rigid bodies (limbs) within its limited range [1]. Starting from the joint node, the relative motion between the adjacent rigid bodies connected by it is represented by some motion-related parameters and the limitation of the joint node on these parameters, and the human skeleton model can be embodied [2]. The real human joints are complex and delicate structures. In order to simplify the model, we can usually adopt an ideal joint model. The difficulty lies in the capture and recognition of low-resolution objects [3].

This paper combines wireless sensor technology and low-resolution recognition technology to build a low-resolution face recognition and sports training action analysis system based on wireless sensors to improve the effect of sports training action analysis.

2 RELATED WORK

The algorithm based on the principle of statistical analysis is Principal Component Analysis (PCA), which uses KL transform to reduce dimensionality and extract features for face image recognition [4]. The advantage of this algorithm is that it can effectively perform dimensionality reduction calculations for large data and multi-dimensional situations to obtain high resolution. The improvement on this algorithm is the SIFT algorithm, which can improve the robustness of the principal component analysis algorithm. , which has good classification performance for face recognition in complex environments [5]. The combined method of 2DPCA and MSD deletes and simplifies the method of pulse coupled neural network [6], all of which are improved to different degrees in different aspects. The shortcomings of these two methods are also obvious. When the input image is occluded, their recognition rate will decrease to varying degrees. After the research on the occlusion problem, researchers have proposed many improved algorithms, such as the algorithm of multi-classifier fusion based on sparse representation [7], and the algorithm based on scale-invariant characteristics [8]. However, although the occlusion problem can be effectively solved, there are still different degrees of features, so although face recognition has made great progress, the theoretical development is still far from practical application.

The principle of face detection algorithm based on template matching is to perform face detection through feature extraction of existing face models and feature matching of images that need face detection; with the improvement of computer hardware, the enhancement of computing power , Modern face models can already extract features from a large number of face models, but this method has very high requirements on computer hardware, and many machine learning algorithms

have been fully applied in it [9]. Literature [10] modeled facial organs, and then performed face image detection according to the template matching method. The limitation of this method is that the input image must be a frontal face image. Literature [11] conducts face detection through variability by studying the variability of face feature templates.

Since the image must require frontal input, researchers propose a fuzzy template matching algorithm [12], which can well solve the problem that the input image is not frontal input. In some complex environments, in addition to the influence of illumination and posture mentioned above, there are still many other interference factors in the face image, which requires researchers to improve the robustness of the existing face detection algorithm. A template matching algorithm that combines multiple features was born [13]. With the improvement of computer hardware performance and the application of 3D image acquisition equipment, the template matching method based on 3D face image features uses multi-level B-spline surface fitting and point cloud technology to extract median and horizontal contour lines, and obtain the overall information by mapping the control points of the 3D face to the depth image, and finally perform face detection through weighted summation [14].

In real life, the face images obtained by humans contain various complex noises [15]. Therefore, the above-mentioned methods based on prior knowledge and based on template matching have relatively low detection accuracy in natural scenes. In comparison, the face detection algorithm based on statistical learning selects a classification with a low false positive rate after comparing several classification rules. rules to detect faces. And how to obtain the classification rules with low false positive rate, this is done through the diversity selection of the sample database and the effective template matching of the trained classifier through the sample database. Face detection algorithm based on neural network [16], the input of this method is the feature vector of the face image, and the output is the category information of the face image. After comparing with the expected value, the positive weights and thresholds are retained. , although this method has high resolution, it cannot be generalized in practice due to the large amount of computation.

The face recognition system mainly includes two parts: face detection and face recognition. If under laboratory conditions, we can easily obtain face detection and face recognition in a single context. A high-resolution input image, thereby obtaining experimental results with high recognition rates [17]. However, in the actual natural environment, people often need to deal with the face image model in a complex environment. At this time, the incoming face image will be affected by various complex environments such as illumination changes, expression changes, posture changes and facial organ occlusion. These complex environments will seriously affect the signal-to-noise ratio of the input image. The recognition rate of face recognition will be greatly reduced. Even face recognition cannot be performed effectively [18]. Therefore, before we want to detect and recognize the target face image, we need to preprocess the image first: the purpose of the preprocessing of the face image is to reduce the interference signal of the image, reduce the influence of objective conditions, and improve the signal-to-noise of the image. ratio, so that the input image can meet the image conditions of face detection and face recognition later as much as possible [19].

3 LOW-RESOLUTION IMAGE FEATURE EXTRACTION ALGORITHM

This section summarizes the global and local feature extraction techniques commonly used in feature detection. It mainly introduces the definition, properties and application of each global or feature extraction algorithm in feature detection. At the same time, it is determined whether it can be applied in the touch screen feature detection algorithm according to the advantages and disadvantages of each feature and the characteristics of the low-resolution touch screen surface image.

- (1) Gray-level co-occurrence matrix

The gray-level co-occurrence matrix is the joint distribution of pixel gray-level pairs in the spatial domain under a specific step size and direction, which is a second-order statistic. The essence of the grayscale matrix is to start from the pixel with grayscale i at any position in space, and count the probability $p(i, j, d, \theta)$ of the pixel whose distance is $d = (dx^2 + dy^2)^{1/2}$ and the grayscale value is j at the same time. The probability $p(i, j, d, \theta)$ of occurrence of a grayscale pair (i, j) in a specific direction for an image $f(x, y)$ of size $N \times N$ can be expressed as the following formula.

$$p(i, j, d, \theta) = \{[(x, y), (x + dx, y + dy)] | f(x, y) = i, f(x + dx, y + dy) = j\} \quad (3.1)$$

Among them, x, y are the coordinates of the pixels in the image, $i, j = 0, 1, 2, \dots, L-1$ is the gray level of the image, and dx, dy are the offsets of the spatial position. d and θ are the generation step size and generation direction of the probability matrix P , respectively. Normally, θ takes $0^\circ, 45^\circ, 90^\circ, 135^\circ$ in four directions. The details are shown in the figure below.

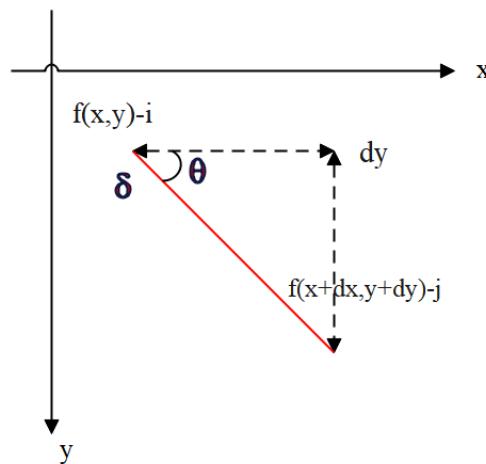


Figure 1: Schematic diagram of the gray-level co-occurrence matrix.

The GLCM obtained from the above formula (1) is a symmetric matrix. If the texture period of the image is large, that is, the texture is rough, the non-zero values of matrix $P(i, j, d, \theta)$ are concentrated near the main diagonal. If the texture period of the image is small, that is, the texture is finer, the non-zero values of matrix $P(i, j, d, \theta)$ are widely distributed. When the generation step is fixed, if the generation direction θ is the same as the direction of the texture in the image, the non-zero element values of the gray-level co-occurrence matrix $P(i, j, d, \theta)$ are mainly distributed near the main diagonal.

In applications such as texture recognition and classification, the gray-level co-occurrence matrix is not directly used as the image feature, but the secondary statistics extracted on the basis of the

gray-level co-occurrence matrix are used as the texture feature. Before extracting texture features, the gray-level co-occurrence matrix $P(i, j, d, \theta)$ needs to be normalized, as follows:

$$\hat{P}(i, j) = P(i, j) / R \quad (3.2)$$

Among them, R is the normalization constant. Six commonly used texture feature parameters are extracted from the normalized gray-level co-occurrence matrix $\hat{P}(i, j, d, \theta)$ (for the convenience of writing, $\hat{P}(i, j, d, \theta)$ is represented as a matrix about i, j), as follows:

1) Contrast

$$f_1 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [p(i, j, d, \theta)(i - j)^2] \quad (3.3)$$

If the gray value of the pixel pair i, j in the image has a large difference, the visual effect of the image is clearer and the contrast value f_1 is larger.

2) Correlation

$$f_2 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (ij\hat{P}(i, j) - \mu_1\mu_2) / \sigma_1^2\sigma_2^2 \quad (3.4)$$

Among them, there is $\mu_1 = \sum_{i=0}^{L-1} i \sum_{j=0}^{L-1} \hat{P}(i, j)$, $\mu_2 = \sum_{j=0}^{L-1} j \sum_{i=0}^{L-1} \hat{P}(i, j)$, $\sigma_1 = \frac{1}{L} \sum_{i=0}^{L-1} (i - \mu_1)^2 \sum_{j=0}^{L-1} \hat{P}(i, j)$ and

$$\sigma_2 = \frac{1}{L} \sum_{j=0}^{L-1} (j - \mu_2)^2 \sum_{i=0}^{L-1} \hat{P}(i, j)$$

. Correlation describes the similarity between row elements or column elements in the gray-level co-occurrence matrix. In texture images, the longer the gray value extends in a specific direction, the greater the correlation.

3) Entropy

$$f_3 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \hat{P}(i, j) \log \hat{P}(i, j) \quad (3.5)$$

Entropy is a measure of the amount of information an image has, and it can also reflect the complexity of the texture contained in the image. If the texture of the image is complex, the entropy value is large. Conversely, the more uniform the gray level in the image, the smaller the entropy value.

4) Homogeneity

$$f_4 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{\hat{P}(i, j)}{1 + (i - j)^2} \quad (3.6)$$

Homogeneity is also called inverse disparity, which is a measure of the local gray uniformity of an image. If the local grayscale of the image is uniform, the homogeneity f_4 takes a larger value.

5) Angular second moment

$$f_5 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \hat{P}^2(i, j) \quad (3.7)$$

The angular second moment reflects the thickness of the texture in the image. If the texture is rough, the value of f_5 is large, and if the texture is fine, the value of f_5 is small.

6) Variance

$$f_6 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - \mu)^2 \hat{P}(i, j) \quad (3.8)$$

Among them, μ is the mean value of the gray level co-occurrence matrix. When the value of f_6 is larger, the corresponding period of the texture in the image is larger.

In texture image analysis, Fourier transform realizes the conversion of image space domain and frequency domain, that is, the grayscale change and texture information in the spatial domain are converted into corresponding amplitude, phase angle and frequency information in the frequency domain. The power spectrum after Fourier transform not only reflects the change of energy in the image, but also reflects part of the texture features of the image. The one-dimensional discrete Fourier transform corresponding to the one-dimensional discrete function $f(x), x = 0, 1, 2, \dots, N$ is formula (3.9):

$$F(u) = \sum_{x=0}^{N-1} f(x) e^{-j2\pi ux/N}, (u = 0, 1, 2, \dots, N-1) \quad (3.9)$$

A two-dimensional Fourier transform is essentially a superposition of one-dimensional Fourier transforms on each row scan line and column scan line. That is, the discrete Fourier transform corresponding to an image $f(x, y)$ of size $M \times N$ is formula (3.10). Formula (3.10) can be rewritten as a complex number, that is, formula (3.11).

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(ux/M + vy/N)} \quad (3.10)$$

$$F(u, v) = R(u, v) + jI(u, v) \quad (3.11)$$

Among them, $R(u, v)$ is the real part and $I(u, v)$ is the imaginary part. Then, the amplitude, phase angle and power spectrum of the Fourier transform are defined as formulas (12) and (13), respectively:

$$M = \left(R^2(u, v) + I^2(u, v) \right)^{1/2}, P = \arctan(I(u, v) / R(u, v)) \quad (3.12)$$

$$|F(u, v)|^2 = F(u, v)F^*(u, v) = R^2(u, v) + I^2(u, v) \quad (3.13)$$

The SIFT operator is summarized into the following five steps:

The scale space $L(x, y, \sigma)$ of an image is obtained by convolving the Gaussian function $G(x, y, \sigma)$ with scale changes with the original image $I(x, y)$, that is

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (3.14)$$

Among them, "*" is the convolution symbol, and σ is the spatial scale factor. It describes the overall characteristics of the image. If σ is small, it describes the detailed features of the image. SIFT feature point detection is carried out on the difference of Do G scale space. The Do G scale space is obtained by subtracting different scale spaces, that is

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (3.15)$$

In order to make the SIFT feature rotation invariant, a certain strategy is used for all the key points extracted above (in the feature point area, the gradient histogram is counted, and the direction with the largest bin value of the histogram and more than 80% of the largest bin value is taken as the main direction.) to assign the main direction to it. In this way, each key point has three attributes: position, scale, and direction.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (3.16)$$

$$\theta(x, y) = \arctan((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (3.17)$$

Among them, $m(x, y)$ and $\theta(x, y)$ are the modulo value and direction of the position (x, y) in the image, as shown in Figure 2.

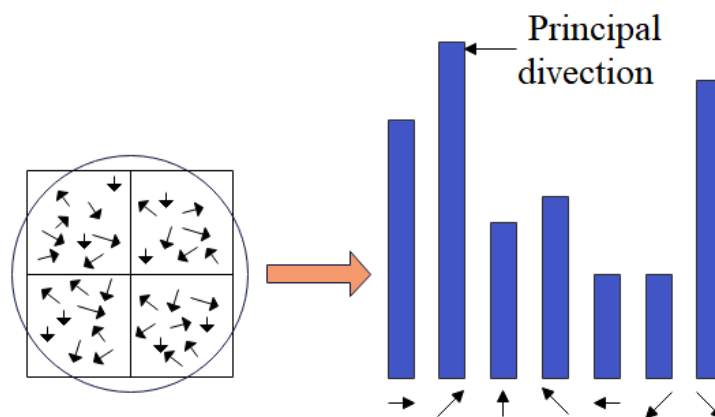


Figure 2: The main directions of key points are determined.

In general, it is assumed that there is an area of size 16x16, and this area is divided into 4x4 areas of 4x4 dimensions. A weighted gradient histogram is computed for each region, which is divided into 8 small patches. Then, the dimension of each feature point is 4x4x8=128 dimensions. The specific generation process is shown in Figure 3.

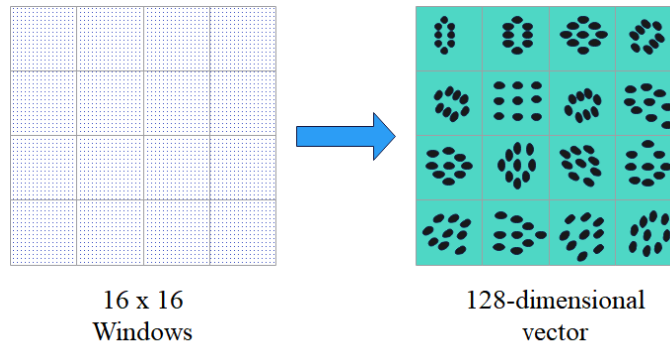


Figure 3: Description of the key points of the SIFT operator (among them, the red dots are the key points).

SIFT uses the DOG difference image, while SURF uses the Hessian matrix determinant to approximate the original image. In the description of SURF features, each pixel corresponds to a Hessian matrix. Since the SURF feature needs to construct a multi-scale space, the Hessian matrix corresponding to the pixel point $X(x, y)$ of the image on the σ scale is:

$$H(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{yx}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix} \quad (3.18)$$

Among them, $L_{xx}(X, \sigma), L_{xy}(X, \sigma), L_{yx}(X, \sigma), L_{yy}(X, \sigma)$ is the convolution of the second-order Gaussian partial differential equation and the original image at point (x, y) . In order to reduce the computation time and computational complexity, the SURF algorithm uses a block filter (or box filter) to approximate the second-order Gaussian partial differential. Therefore, the approximate Hessian matrix $\det(H_{approx})$ can be expressed as:

$$\det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy}) \quad (3.19)$$

Among them, the weight coefficient w usually takes the value of 0.9 ; D_{xx}, D_{xy}, D_{yy} is the value after the convolution operation between the box filter and the image.

Different from the SIFT algorithm, the SURF algorithm obtains the main direction of the feature point by counting the Harr wavelet features in the feature point field (for example, a circle with a radius of 6σ , and σ is the scale of the point). As shown in Figure 4, if we want to determine the main direction of the feature point where the center of the circle is located, we can select 60° as the rotation interval for rotation. At the same time, we calculate the sum of the horizontal and vertical Harr wavelet responses of all feature points in the rotation region corresponding to each interval. Finally, a vector with magnitude and direction can be obtained. The SURF algorithm selects the

sector where the maximum vector modulus value is located as the main direction of the feature point.

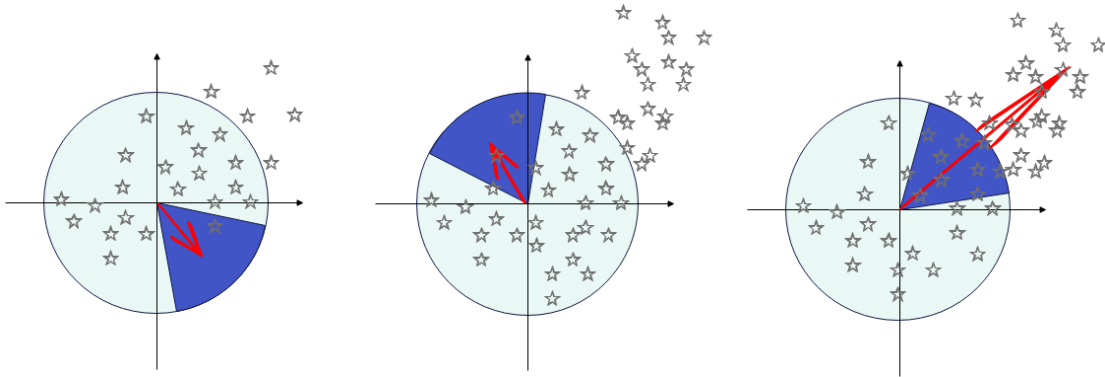


Figure 4: Figure 4 The main directions of SURF feature points are determined.

Different from the SIFT algorithm, the SURF algorithm takes a square box with a main direction of 60 around the feature point. The box has a main direction. Then, the square frame is divided into 4x4 sub-regions, and the sum of the Harr wavelet responses in the horizontal and vertical directions of the $5\sigma \times 5\sigma$ pixels in the sub-region is calculated to form a four-dimensional vector. The specific process is shown in Figure 5

$$v = \left(\sum dx, \sum dy, \sum |dx|, \sum |dy| \right) \quad (3.20)$$

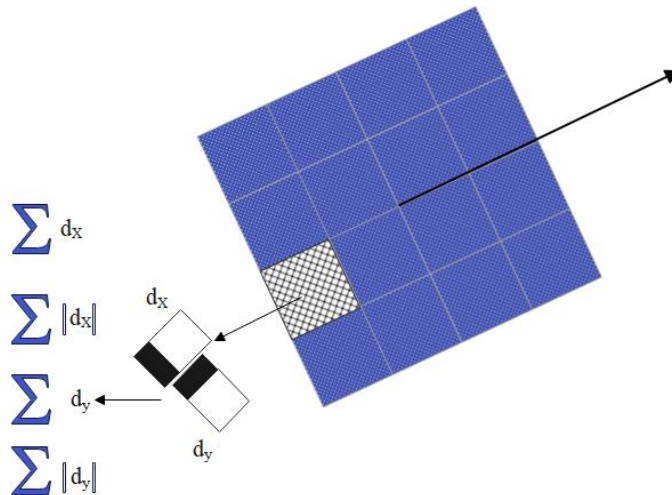


Figure 5: Generation of feature point descriptors.

Finally, the 4-dimensional vectors of the 4x4 sub-regions are sequentially connected, and the 64-dimensional feature vector of the feature point can be obtained.

A one-dimensional windowed short-time Fourier transform (STFT, or Gabor transform for short) is proposed, and the Gabor transform is defined as:

$$X(t, \omega) = \int_{-\infty}^{+\infty} x(s)g(s-t)e^{-j\omega s} ds \quad (3.21)$$

$$\psi_{t,\omega}(s) = g(s-t)e^{j\omega s} \quad (3.22)$$

Among them, $g(s)$ represents the window function, which can be Hanning, Hamming, Gaussian and other functions, and $x(s)$ is a given time domain signal. The essence of Gabor transform is the inner product of the time domain signal $x(s)$ and the basis function $\psi_{t,\omega}(s)$. A two-dimensional Gabor filter is defined as:

$$\psi(f_\mu, \theta_\nu, \gamma, \eta) = \frac{f_\mu^2}{\pi\gamma\eta} \exp\left[-\left(\frac{f_\mu^2}{\gamma^2}x_r^2 + \frac{f_\mu^2}{\eta^2}y_r^2\right)\right] \exp^{j2\pi f_\mu x_r} \quad (3.23)$$

$$x_r = x \cos \theta_\nu + y \sin \theta_\nu, \quad y_r = -x \sin \theta_\nu + y \cos \theta_\nu$$

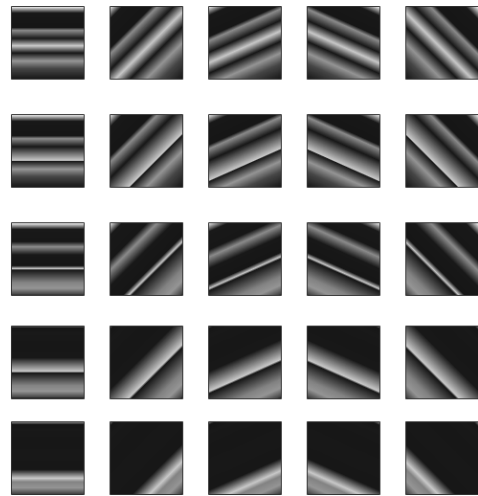
Among them, $f_\mu = f_{\max} / (\sqrt{2})^\mu$, $\theta_\nu = \frac{h}{8} \pi$, f_μ is the center frequency of the filter and θ_ν is the phase of the filter. μ and ν represent the scale and direction of the corresponding Gabor filter, respectively, γ and η are the wavelengths of the filter in the x (long axis) and y (short axis) directions, respectively, and x, y are the positions of the pixels in the spatial domain. Obviously, x, y can also be expressed as a function $G(f_\mu, \theta_\nu, \gamma, \eta)$ about the pixel position, that is, $G_{\mu,\nu}(x, y)$. The two-dimensional Gabor function is a complex function, that is

$$\psi_{\mu,\nu}(x, y) = \text{Re}(\psi_{\mu,\nu}(x, y)) + i \text{Im}(\psi_{\mu,\nu}(x, y)) \quad (3.24)$$

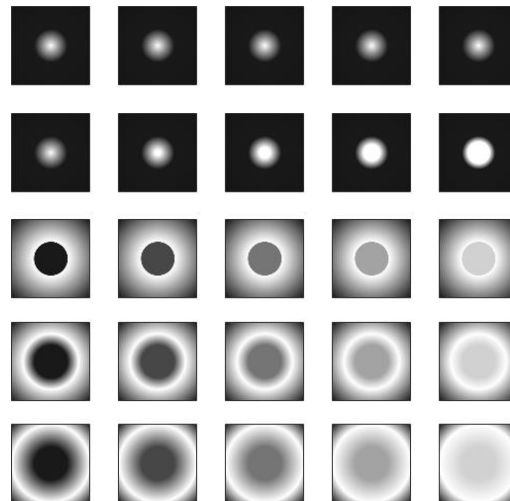
Among them, $\text{Re}(G_{\mu,\nu}(x, y))$ is the real part and $\text{Im}(G_{\mu,\nu}(x, y))$ is the imaginary part. In the application of Gabor filter, the most used is the amplitude response. Therefore, the above formula can be rewritten in the form of amplitude $M_{\mu,\nu}(x, y)$ and phase angle $\theta_{\mu,\nu}(x, y)$, as shown in formula (25). The real part response and amplitude response of Gabor filter are shown in Figure 6(a), 6(b), respectively.

$$\psi_{\mu,\nu}(x, y) = M_{\mu,\nu}(x, y)e^{i\theta_{\mu,\nu}(x, y)} \quad (3.25)$$

Gabor filter has the same receptive field as biological vision, and it has the advantages of spatial locality, illumination and rotation invariance. Therefore, Gabor filtering is widely used in the fields of computer vision and image processing. However, the selection of filter parameters has always been a difficult point in the application of Gabor filters.



(a) Real part response (25)



(b) Amplitude response (25)

Figure 6: Gabor filter response.

Among them, the main parameters of the Gabor filter are the highest frequency f_{\max} , the wavelength γ or η of the filter on the long or short axis, and the window size.

The size of the window in the Gabor filter is the Gabor filter template. Figure 7 shows the transformation of the real part response of the Gabor filter in the same direction and the same scale when the size of the filter template is $11 \times 11, 21 \times 21, 31 \times 31, 41 \times 41$ (other parameters are $f_{\max} = 0.25, \gamma = \sqrt{2}, \eta = \sqrt{2}$). It can be seen from the figure that the larger the template of the filter is, the more blurred the filtering effect of the image will be.

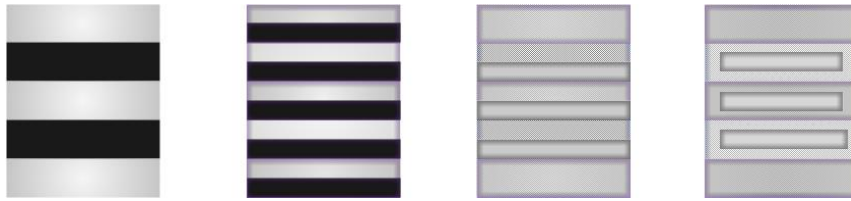


Figure 7: Effect of window size on real part response of 2D Gabor filter.

2) Influence of filter center frequency f_{\max} on texture wavelength

The center frequency f_{\max} is an important parameter of the filter, which determines the size of the filter wavelength. For example, in the case of the same scale and the same direction, the parameter is $\gamma = \sqrt{2}$ 、 $\eta = \sqrt{2}$, and the filter template size is 25×25 , as the center frequency f_{\max} (from 0.1 to 0.5, step size is 0.1) increases sequentially, the wavelength of the filter gradually decreases.

3) The effect of transverse wavelength γ and longitudinal wavelength η on the real part response of the two-dimensional Gabor filter. γ and η are the wavelengths of the filter on the x-axis and y-axis, respectively, η/γ is the spatial aspect ratio, when there is $\eta/\gamma = 1$.

b) The changes in the real part response of the Gabor filter is listed when there is $\gamma = 0.9\sqrt{2}, 1.9\sqrt{2}, 2.9\sqrt{2}, 3.9\sqrt{2}$ and $\eta = 0.9\sqrt{2}, 1.9\sqrt{2}, 2.9\sqrt{2}, 3.9\sqrt{2}$, respectively (the experiment is in the same scale and in the same direction, and there is $f_{\max} = 0.3$, the filter template is 25). In a), as the value of γ increases, the frequency range in the lateral direction of the coordinate axis becomes larger. In b), with the increase of η , the frequency range in the longitudinal direction of the coordinate axis becomes larger.

Any touch screen surface image $I(x, y)$ is given, and the sub-image block after overlapping and sub-blocking is $i_{\text{block}}(x, y)$. Its Gabor transform can be expressed as the convolution of $i_{\text{block}}(x, y)$ and the Gabor wavelet kernel function, as shown in formula (26). Among them, "*" is the convolution symbol. If the magnitude and phase angle of the output response $G_{\mu, \nu}(x, y)$ are assumed to be $M_{\mu, \nu}(x, y)$ and $\theta_{\mu, \nu}(x, y)$, respectively, $G_{\mu, \nu}(x, y)$ can be written as formula (27).

$$G_{\mu, \nu}(x, y) = i_{\text{block}}(x, y) * \psi_{\mu, \nu}(x, y) \quad (3.26)$$

$$G_{\mu, \nu}(x, y) = M_{\mu, \nu}(x, y) \cdot \exp(i\theta_{\mu, \nu}(x, y)) \quad (3.27)$$

According to the properties of Gabor filtering, different μ and ν can be selected to form different Gabor filter groups. This group of Gabor filters is used to expand the sub-image block, and its frequency domain feature information at different scales and different directions can be obtained.

This paper chooses 5 frequencies (that is $\mu = 0, 1, 2, 3, 4$) and 5 directions (that is $\nu = 0, 1, 2, 3, 4$),

so there are 25 Gabor filter functions in total. Because the amplitude information of the image reflects the energy spectrum of the image. Therefore, it is often used for texture recognition classification. Here, we use the statistical features of the magnitude matrix, such as the mean $m_{\mu,v}$, the standard deviation $\sigma_{\mu,v}$, the maximum value \max , and the energy value energy to represent the texture features of the image block. When there is only one statistical feature, the feature vectors of each sub-image block are arranged in the order of scale first and then direction. They are denoted by F_{mean} and F_{var} , as shown in formula (28). When there are two statistical features, the feature vector is the concatenation of the two feature vectors F_{mean} and F_{var} , which is denoted by $F_{\text{mean+var}}$, as shown in formula (29). In order to make the order of magnitude of the statistical features uniform, we normalize the feature vector, and the normalization formula is $F = (F - \text{mean}(F)) / \text{std}(F)$. F_{mean} , F_{var} and $F_{\text{mean+var}}$ are 25-, 25-, and 50-dimensional real vectors, respectively.

$$F_{\text{mean}} = (m_{0,0}, m_{0,1}, m_{0,2}, \dots, m_{4,4}), F_{\text{var}} = (\sigma_{0,0}, \sigma_{0,1}, \sigma_{0,2}, \dots, \sigma_{4,4}) \quad (3.28)$$

$$F_{\text{mean+var}} = (m_{0,0}, m_{0,1}, m_{0,2}, \dots, m_{4,3}, m_{4,4}, \sigma_{0,0}, \sigma_{0,1}, \dots, \sigma_{4,4}) \quad (3.29)$$

M normal images are selected for training to obtain a dictionary D, that is

$$D = [F_{1,1}, F_{1,2}, \dots, F_{i,j}, \dots, F_{M,1}, \dots, F_{M,N}], D \in R^{f \times N \times M} \quad (3.30)$$

Among them, N is the number of overlapping blocks of each image, and f is the dimension of the feature vector. For the convenience of expression, we uniformly express formula (30) as

$$D = [F_1, F_2, \dots, F_i, \dots, F_Q]. \text{ Among them, there is } Q = M \times N.$$

For the constructed dictionary D, it needs to be optimized to make it more compact and have better expressive ability. Here, we consider the requirement that every atom in D can be linearly represented by the other atoms as best as possible, that is:

$$\min \|z\|_0 \quad \text{s.t.} \quad D'z = F_i, \quad \text{and } z_i = 0, i = 1, 2, \dots, Q \quad (3.31)$$

Among them, D' is the matrix composed of other column atoms except F_i in D. In practical applications, in order to facilitate the use of the optimal method to solve the above formula, the l_1 norm is used to replace the l_0 norm. Therefore, the above formula can be written as:

$$\min \|z\|_1 \quad \text{s.t.} \quad D'z = F_i, \quad \text{and } z_i = 0, i = 1, 2, \dots, Q \quad (3.32)$$

For any touch screen surface image I(x, y) to be tested, after overlapping and segmenting, it is assumed to be divided into L blocks, and the Gabor feature of each block is extracted to form a feature vector matrix $F = [F_1, F_2, \dots, F_i, \dots, F_N], F_i \in R^f$. Among them, N is the number of

overlapping blocks for each image. According to the classification theory of sparse representation, each atom in the eigenvector matrix can be represented linearly under the optimized dictionary D_0 , and the sparse representation coefficient α_i can be obtained. Therefore, the sparse cost value S for this image is:

$$S = \sum_{i=1}^N \|\alpha_i\|_1 \quad (3.33)$$

Since dictionary D_0 is composed of Gabor features of normal images. Therefore, if I is featureless, it should be well sparsely reconstructed by dictionary D_0 , that is, the sparse cost value changes little. Therefore, in this paper, the mean value of the sparse cost value S of L normal images is used as the threshold to distinguish feature images from normal images, that is:

$$m_s = \frac{1}{L} \sum_L S, \text{ threshold} = m_s \quad (3.34)$$

The sparse representation coefficients of sub-blocks of normal images under dictionary D_0 are sparser, that is, the value of l_1 -norm is smaller. On the other hand, the value of the l_1 norm of the sparse representation coefficient of the feature image under the dictionary D_0 is large. Therefore, if the sparse cost value S of the image to be tested is greater than the threshold, the image to be tested is characteristic. The opposite is not true, that is:

$$\begin{cases} \text{Normal image} & S \leq \text{threshold} \\ \text{Defect image} & S > \text{threshold} \end{cases} \quad (3.35)$$

4 LOW-RESOLUTION FACE RECOGNITION AND SPORTS TRAINING ACTION ANALYSIS BASED ON WIRELESS SENSORS

The integrated learning structure architecture of the low-resolution face recognition system based on wireless sensors proposed in this paper is shown in Figure 8. The most important thing in the establishment of sports tracking system combined with wireless sensors is to establish the model and observation model of the moving target. The establishment of the motion model requires the construction of a camera model and a human body model, as shown in Figure 9. Figures 10 and 11 show the simulation examples of the low-resolution face recognition and sports training action analysis system based on wireless sensors in this paper (the pictures are from the network).

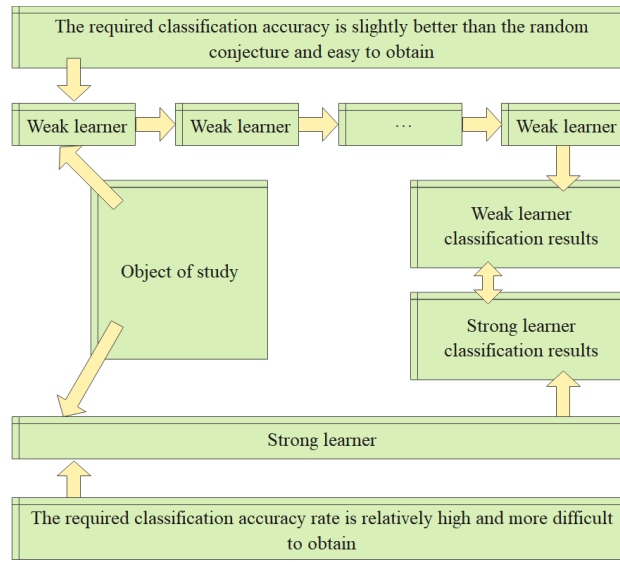


Figure 8: Integrated learning structure system.

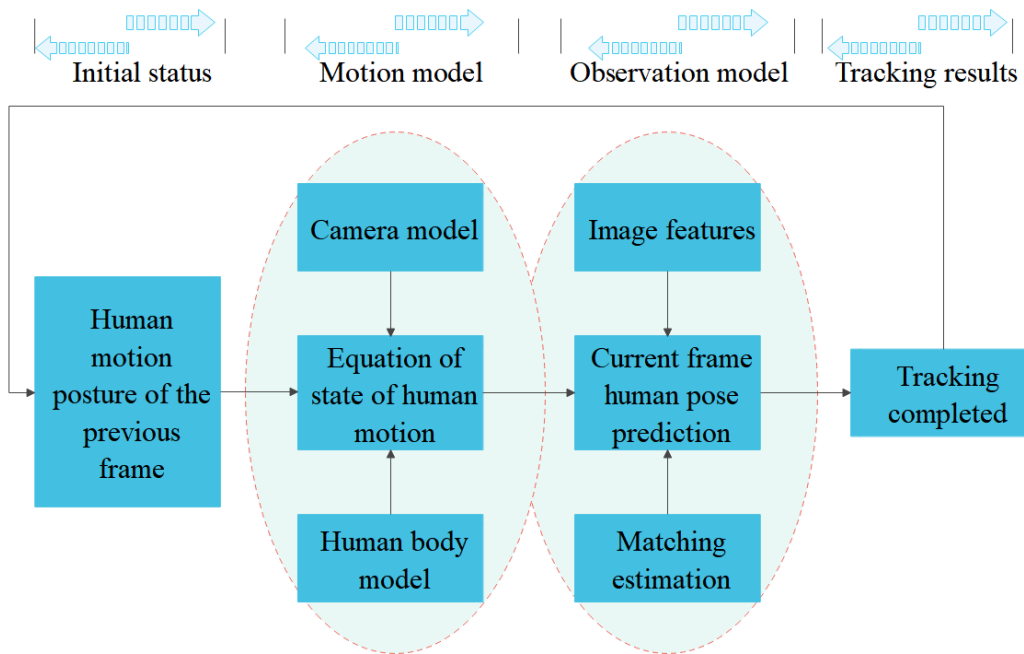


Figure 9: Principle block diagram of sports training action tracking.



(a) Original image



(b) Face recognition background removal



(c) Facial recognition digital processing

Figure 10: Example image of face recognition.



(a) Original image



(b) Background removal of sports training recognition



(c) Digitization of sports training recognition

Figure 11: Example image of sports training recognition.

On the basis of the above research, the low-resolution face recognition and sports training action analysis system based on wireless sensor proposed in this paper is simulated and verified, and the results shown in Table 1 and Table 2 are obtained.

<i>Number</i>	<i>Face recognition</i>	<i>Number</i>	<i>Face recognition</i>	<i>Number</i>	<i>Face recognition</i>
1	90.23	17	87.59	33	83.22
2	87.14	18	88.55	34	86.48
3	90.86	19	80.13	35	81.18
4	91.03	20	88.92	36	83.21
5	80.50	21	86.11	37	87.47
6	80.85	22	81.24	38	80.47
7	90.51	23	84.83	39	87.22
8	85.67	24	83.67	40	88.85
9	83.24	25	80.15	41	90.16
10	90.40	26	85.49	42	81.32
11	82.09	27	91.29	43	80.97
12	84.93	28	81.88	44	82.04
13	90.96	29	89.30	45	81.96
14	88.19	30	89.70	46	86.68

15	84.02	31	84.16	47	91.85
16	90.71	32	81.02	48	90.33

Table 1: Face recognition effect.

Number	Action analysis	Number	Action analysis	Number	Action analysis
1	81.57	17	78.10	33	82.03
2	81.01	18	76.10	34	81.51
3	86.76	19	81.92	35	83.49
4	84.62	20	83.73	36	80.32
5	86.97	21	85.66	37	86.83
6	80.97	22	87.32	38	87.62
7	85.71	23	77.10	39	83.44
8	80.07	24	87.82	40	76.43
9	84.65	25	77.07	41	78.44
10	83.91	26	85.98	42	81.65
11	86.62	27	87.18	43	86.08
12	84.10	28	85.48	44	84.96
13	85.56	29	87.43	45	83.50
14	82.68	30	79.94	46	79.09
15	83.31	31	83.50	47	78.12
16	79.98	32	81.79	48	87.33

Table 2: Analysis effect of sports training action.

It can be seen from the above research that the low-resolution face recognition and sports training action analysis system based on wireless sensors proposed in this paper has good practical results.

5 CONCLUSION

Sports training generally integrates a variety of functions, and often has a relatively complex sports rehabilitation system. When conducting training system and kinematic analysis, it would be too cumbersome to use the traditional method of studying from a single limb level, and the number of node units involved is huge and various, making it difficult to carry out in practical applications. Moreover, the traditional human motion analysis method relies on the entity node, starts from the static point of view, constructs the position vector loop according to the space training system of the human body system, and further uses the method of derivation and coordinate transformation to derive its motion parameter expression. This paper combines wireless sensor technology and low-resolution recognition technology to construct a low-resolution face recognition and sports training action analysis system based on wireless sensors. The simulation study shows that the low-resolution face recognition and sports training action analysis system based on wireless sensors proposed in this paper has good practical results.

Shiwu Chen, <https://orcid.org/0009-0005-5384-0163>

REFERENCES

- [1] Aso, K.; Hwang, D. H.; & Koike, H.: Portable 3D Human Pose Estimation for Human-Human Interaction using a Chest-Mounted Fisheye Camera, In Augmented Humans Conference 2021, 2021, 116-120. <https://doi.org/10.1145/3458709.3458986>
- [2] Bakshi, A.; Sheikh, D.; Ansari, Y.; Sharma, C.; Naik, H.: Pose Estimate Based Yoga Instructor, International Journal of Recent Advances in Multidisciplinary Topics, 2(2), 2021, 70-73.
- [3] Díaz, R. G.; Laamarti, F.; & El Saddik, A.: DTCoach: Your Digital Twin Coach on the Edge During COVID-19 and Beyond, IEEE Instrumentation & Measurement Magazine, 24(6), 2021, 22-28. <https://doi.org/10.1109/MIM.2021.9513635>
- [4] Ershadi-Nasab, S.; Noury, E.; Kasaei, S.; Sanaei, E.: Multiple human 3d pose estimation from multiview images, Multimedia Tools and Applications, 77(12), 2018, 15573-15601. <https://doi.org/10.1007/s11042-017-5133-8>
- [5] Gu, R.; Wang, G.; Jiang, Z.; Hwang, J. N.: Multi-person hierarchical 3d pose estimation in natural videos, IEEE Transactions on Circuits and Systems for Video Technology, 30(11), 2019, 4245-4257. <https://doi.org/10.1109/TCSVT.2019.2953678>
- [6] Hua, G.; Li, L.; Liu, S.: Multipath affinity stacked—hourglass networks for human pose estimation, Frontiers of Computer Science, 14(4), 2020, 1-12. <https://doi.org/10.1007/s11704-019-8266-2>
- [7] Li, M.; Zhou, Z.; Liu, X.: Multi-person pose estimation using bounding box constraint and LSTM, IEEE Transactions on Multimedia, 21(10), 2019, 2653-2663. <https://doi.org/10.1109/TMM.2019.2903455>
- [8] Liu, S.; Li, Y.; Hua, G.: Human pose estimation in video via structured space learning and halfway temporal evaluation, IEEE Transactions on Circuits and Systems for Video Technology, 29(7), 2018, 2029-2038. <https://doi.org/10.1109/TCSVT.2018.2858828>
- [9] Martínez-González, A.; Villamizar, M.; Canévet, O.; Odobez, J. M.: Efficient convolutional neural networks for depth-based multi-person pose estimation, IEEE Transactions on Circuits and Systems for Video Technology, 30(11), 2019, 4207-4221. <https://doi.org/10.1109/TCSVT.2019.2952779>
- [10] McNally, W.; Wong, A.; McPhee, J.: Action recognition using deep convolutional neural networks and compressed spatio-temporal pose encodings, Journal of Computational Vision and Imaging Systems, 4(1), 2018, 3-3.
- [11] Mehta, D.; Sridhar, S.; Sotnychenko, O.; Rhodin, H.; Shafiei, M.; Seidel, H. P.; Theobalt, C.: Vnect: Real-time 3d human pose estimation with a single rgb camera, ACM Transactions on Graphics (TOG), 2017, 36(4), 1-14. <https://doi.org/10.1145/3072959.3073596>
- [12] Nasr, M.; Ayman, H.; Ebrahim, N.; Osama, R.; Mosaad, N.; Mounir, A.: Realtime Multi-Person 2D Pose Estimation, International Journal of Advanced Networking and Applications, 2020, 11(6), 4501-4508. <https://doi.org/10.35444/IJANA.2020.11069>
- [13] Nie, X.; Feng, J.; Xing, J.; Xiao, S.; Yan, S.: Hierarchical contextual refinement networks for human pose estimation, IEEE Transactions on Image Processing, 28(2), 2018, 924-936. <https://doi.org/10.1109/TIP.2018.2872628>
- [14] Nie, Y.; Lee, J.; Yoon, S.; & Park, D. S.: A Multi-Stage Convolution Machine with Scaling and Dilution for Human Pose Estimation, KSII Transactions on Internet and Information Systems (TIIS), 13(6), 2019, 3182-3198. <https://doi.org/10.3837/tiis.2019.06.023>
- [15] Petrov, I.; Shakhuro, V.; Konushin, A.: Deep probabilistic human pose estimation, IET Computer Vision, 12(5), 2018, 578-585. <https://doi.org/10.1049/iet-cvi.2017.0382>
- [16] Szűcs, G.; Tamás, B.: Body part extraction and pose estimation method in rowing videos, Journal of computing and information technology, 26(1), 2018, 29-43. <https://doi.org/10.20532/cit.2018.1003802>

- [17] Thành, N. T.; Công, P. T.: An Evaluation of Pose Estimation in Video of Traditional Martial Arts Presentation, Journal of Research and Development on Information and Communication Technology, 2019(2),2019 114-126. <https://doi.org/10.32913/mic-ict-research.v2019.n2.864>
- [18] Xu, J.; Tasaka, K.; Yamaguchi, M.: Fast and Accurate Whole-Body Pose Estimation in the Wild and Its Applications, ITE Transactions on Media Technology and Applications, 9(1), 2021, 63-70. <https://doi.org/10.3169/mta.9.63>
- [19] Zarkeshev, A.; Csiszár, C.: Rescue Method Based on V2X Communication and Human Pose Estimation, Periodica Polytechnica Civil Engineering, 63(4),2019,1139-1146. <https://doi.org/10.3311/PPci.13861>