Human Posture Recognition in Intelligent Healthcare

Xinyue Yang\textsuperscript{a}, Xiaoyang Ren\textsuperscript{b}, Meng Chen\textsuperscript{c}, Luqi Wang\textsuperscript{d}, Yuhao Ding\textsuperscript{e}

Zhongnan University of Economics and Law

\textsuperscript{a}1987543919@qq.com, 15717172075.86, \textsuperscript{b}1164578066@qq.com, 18953573439.86, \textsuperscript{c}1184945444@qq.com, 15921619931.86, \textsuperscript{d}641517890@qq.com, 13545747916.86, \textsuperscript{e}1023776337@qq.com, 18507155147.86.

\textbf{ABSTRACT.} With the rapid growth of medical technology, intelligent healthcare has gradually integrated into our lives, and its core technology is the human body motion state detection. According to the basic tools being used, we divide this technology into two categories: vision-based detection and sensor-based detection, and take the fall monitoring as an example for elaboration. Finally, different methods are compared and analyzed, the difficulties that need to be solved in this field are summarized, the future trends are included too.

1. INTRODUCTION

Intelligent healthcare refers to the realization of interaction between patients and medical institutions to gradually achieve informationization by creating a medical information platform in health record field by means of the Internet of Things. In the near future, the healthcare industry will incorporate more technologies for the realization of real intelligence and informatization. But one of the most important technologies for the implementation is motion detection technology. Nowadays, the motion detection technology applied in the field of intelligent healthcare at home and abroad roughly falls into two types, one is vision-based detection, and the other is sensor-based detection\textsuperscript{[1]}. The former is judged by performing playback and video location technologies on the saved video frames. By installing multiple cameras in the environment in which the detected object is located, computer vision technology is used to combine the motion process with changes in human body morphology, and the motion state of the detected object is further determined by analyzing changes in the posture of the human body. In a dynamic environment, the 3D simulated pictures of the detected object’s motion are available for real-time display, and the judgment results are more accurate. Sensor-based motion detection is determined by a combination of analysis and integration of human posture data measured by the sensor. It is mainly through the collection of node fusion data for measurement in each major joint parts of human body that the human posture information is analyzed, followed with the judgment of the motion state. Such detection is low in cost and convenient to carry, with which the real-time monitoring of the motion state is realizable.

In this paper, the vision and sensor-based motion detection methods are sorted and summarized respectively, and takes the fall detection as an example. The first chapter introduces the research background and the research status at home and abroad. The second chapter and the third chapter introduce the vision-based motion state detection and sensor-based motion state detection according to the different detection methods; The fourth chapter analyzes and compares two detection methods; The fifth chapter looks forward to the future research trends in this field based on the research status of fall detection.
2. VISION-BASED MOTION DETECTION

The vision-based motion state detection mainly refers to the image acquired by the camera, uses some means to detect and track the motion of the human body, extracts parameters that can represent the motion information of the human body, analyzes the motion state of the human body, and finally achieves an understanding of human behavior. The key steps can be divided into three parts: target detection, feature extraction and fall detection (action classification).

Fig.1. Basic process based on visual inspection

2.1 Target Detection

Motion target detection aims to find the kinetic region in video and provide a reference region for the subsequent action analysis and other tasks. The most traditional target detecting algorithms include inter-frame difference method, background subtraction method and optical flow method.

The basic principle of inter-frame difference method[2] is to obtain the difference image by subtracting the images of two successive frames for two-dimensionalization through threshold processing, and finally to acquire the location of the target contour through morphological processing. Some scholars come up with the three-frame differentiate method based on the traditional two-frame differentiate method, which performs the differentiate operation on two consecutive frames in the neighboring three frames, and finally operates the differentiate result to obtain the foreground object.

The basic principle of background subtraction method[3] is to make differentiate operation between the preselected background and the current frame. If it is greater than the threshold, it is considered as the target area. If it is smaller than the threshold, it is considered as the background area. On account of the background subtraction, some further propose algorithms such as linear Kalman filtering[4], statistical average method[5], Gaussian Mixture Model (GMM) [6], and Vibe[7], improve the detection efficiency and render more accurate detection results.

The optical flow method[8] is to represent the motion state of object in the image by calculating the instantaneous velocity of a motion pixel in the video sequence. The basic calculation is as follows: suppose that the pixel \((x, y, t)\) of a certain frame in the video moves to \((x + \Delta x, y + \Delta y, t + \Delta t)\) after the elapse of time \(\Delta t\), since the motion of pixel does not affect the grey level of the pixel, hence \(I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)\) and the optical flow of the pixel is obtained by the derivation of time:

\[
\frac{\partial t}{\partial x} \Delta x + \frac{\partial t}{\partial y} \Delta y + \frac{\partial t}{\partial t} = 0
\]

(1)

\[
\frac{\partial t}{\partial x} V_x + \frac{\partial t}{\partial y} V_y + \frac{\partial t}{\partial t} = 0
\]

(2)
$V_x$ and $V_y$ represent the instantaneous velocity of the pixel along the X and Y directions respectively. Since most of the existing optical flow methods are complex in computation, which are greatly affected by noise, light source, shielding and other factors, ordinary hardware is unavailable and real-time is hard to satisfy. Therefore, generally optical flow method is used in conjunction with other algorithms.

2.2 Feature Extraction

The feature expression of image is a process to map the pixel of original image to the distinguishable dimensionality spatial data, which is a step of great concern to break the relationship between the underlying pixel and the high-level semantics. Judging from whether it can be obtained through self-learning, it can be divided into artificial design-based features and learning-based features.

2.2.1 Feature expression based on artificial design

Feature expression based on artificial design refers to the feature extracted by manual design. Due to the differences between visual features and visual computing, they are roughly divided into four types: gradient feature, pattern feature, shape feature and color feature. The first type describes a moving target by acquiring gradient information close to a specific key point with excellent characteristics such as rotation, scale invariability and typical algorithms of SIFT[9], HOG[10], etc. Pattern feature is a description obtained by analyzing the relative disparity between the local area of an image, usually used to represent the image texture information. Typical algorithms include Gabor[11], LBP[12] and Haar-like[13], etc. Shape feature is a model-based target detection, generally describing the target contour with classical algorithms such as DPM[14], Shape context[15], KASs. The color feature is obtained by calculating the probability distribution of local image attributes (such as grayscale and color). The classical algorithms include Color names[16] and entropy-based salient feature[17].

2.2.2 Feature expression based on machine learning

The application of machine learning to vision-based target detection and tracking mainly includes regression and detection.

The regression result is the exact coordinate of human joint. Alexander Toshev et al. [18] first applied deep learning to the estimation of human posture as a regression problem, and used DNN[19] to locate the entire human joints in a large space. On the basis of human motion image input, an initial coordinate of the joint is obtained using the convolutional neural network transformed by function, and the posture is estimated more accurately by means of cascade. Only local views are used to estimate joints in the estimation of posture, while problems of no great application benefit in practical use do play a certain promotion role.

Using thermograph analysis, different human parts should be given different response values in detection in the corresponding thermograph. The higher the value is, the higher the probability that pixel is the target joint becomes. This method has been dominant in the present human posture recognition researches. The main research direction lies in how to obtain more accurate thermographs that are close to the real marking situation, especially in some undesirable conditions or in the case of distinguishing multiple postures. The main tool used of the detection methods is CNN[20], which is a multi-layer neural network adept in processing images, especially the machine learning problems related to large images. By virtue of a series of methods, convolutional network successfully reduces the dimensionality of image recognition with a large data size, and finally guarantees its training availability. The single-layer convolutional neural network contains two processes, convolution and subsampling, whose implementation procedure is shown in Figure 2:
In the convolution process, different features of the signal are extracted by introducing different convolution kernels to observe the specific patterns of input signal. Dimensionality reduction of the feature graph is mainly performed in subsampling in the form of average or maximum pooling generally. The single-layer convolutional neural network is stacked layer by layer to form CNN where the output characteristics of underlaying layer serve as the input signals of the upper layer.

2.3 Fall Detection
As the feature expression is extracted, the target or background appearance model is established on the basis of feature expression, further classified to obtain the different motion states. There are generally two classification methods, one is by threshold classification and the other is by classifier.

2.3.1 Thresholding
The key of thresholding[21] is to judge a fall by setting one or multiple threshold values. If the received signal exceeds or drops below the preset value, it is judged as a fall with the main expression as below:

\[ f(x, y) = \begin{cases} 
1 & \text{sign} > T \\
0 & \text{sign} \leq T 
\end{cases} \]

(3)

Where, \( f(x, y) \) is the judging result, \( f(x, y) = 1 \) means that the received signal surpasses the preset value, and its state is judged to be a fall. \( f(x, y) = 0 \) means that the received signal does not exceed the preset threshold, judging that its state is not a fall. The selection of threshold value \( T \) is diverse, which is set as required. Medrano et al. [22] set thresholds based on feature mean and variance; Magnenat-Thalmann et al. [23] obtained data thresholds based on human height and then derived velocity thresholds; Planinc et al. [24] based on joint points and ground direction and spine. The judging process of thresholding is simple and fast, but greatly affected by the selection of threshold value. Inappropriate selection of threshold will result in a pre-judgment. Thresholding is usually used jointly with other methods.

2.3.2 Classifier
Classification refers to the construction of a classification function or model based on existing data. This function or model maps the data records in database to one of the given categories, and is applicable to data forecasting. The current classifiers in use include Support Vector Machine (SVM), decision-making tree, logistic regression, Naive Bayes and neural network, etc.

The main idea of support vector machine is to randomly generate a hyperplane in the feature space, translate it in a certain direction until the hyperplane distributes pixels in different categories on both sides. Capezuti et al.[25] used hierarchical SVM to differentiate the fall action and other actions in line with the height variation of the intermediate node for consecutive frames in the detected video. The way of falling is further subdivided: falling while standing and falling off a chair. Bian et al.[26] used SVM to determine the falling directions of head, body, hip and trunk on the basis of RDT.

The main idea of Naive Bayes (NB) is to solve the probability of each category emerging under the condition of a given pending sorting item and the category with the largest probability is taken as the category of sorting item. Athitos et al.[27] used Bayes to judge the height, speed and frame duration number of the head, and a probability calculation is conducted on the plus or minus characteristic to judge the fall.
K-nearest Neighbor (KNN) is to classify the category of one sample into that of the majority of K-Nearest samples in the feature space. Kepski et al.[28] first used moving average method to extract the length-width ratio of human bounding box, further presenting a method to be isolated from human body with a lower calculated amount by using KNN classifier.

3. SENSOR-BASED MOTION MONITORING

Sensor-based motion detection largely relies on the inertial sensor. Node sampling plus fusion data measurement is performed by means of band binding and other methods in the main joints of human body (such as the joints of head, neck, waist and limbs) to detect the motion state. The basic process is shown in Figure 3 below:

![Basic Flow of Motion Detection](image_url)

3.1 Analysis of Human Posture

In the analyzing process of human posture, some minor mechanical vibrations usually trigger high-frequency noise. To suppress these high-frequency noises, it is necessary to filter process the collected original data. By means of recursive average filtering, Kalman filtering and digital filtering, errors and faults in real-time processing and computer operation may be reduced.

Body posture analysis is vital in the inertial sensor that studies the kinematics parameters of different human actions by establishing the cartesian coordinate system OXYZ. With the center of mass of human body as the origin, the front is defined as the positive direction of X-axis, the left-hand side as the positive direction of Y-axis, and the vertical direction as the positive direction of Z-axis. At the same time, the deflection angle around X-axis is defined as roll, the deflection angle around Y-axis as pitch, and that around Z-axis as yaw. The changes in body posture resulted from different human actions are regarded as the rotation and translation of human body in the OXYZ coordinate system[29]. The coordinate system of body posture is as shown in Figure 4:
Fig. 4. Coordinate System of Human Posture

Since different kinematics parameters emerge in different parts of human body in motion, the corresponding accelerated velocity, angular velocity and magnetic deviation are accessible by using the inertial sensor to get ready for posture analysis.

3.1.1 Analysis of static posture

The analysis of static posture is based on the fact that the gravity acceleration of human body is 1g in static state or low-velocity movement, and the acceleration in other directions is almost 0 when the roll angle $\theta$ and pitch angle $\gamma$ are obtained by calculating the acceleration component of gravity in different axial directions. The main formula are as follows:

$$
\theta = \left(-\frac{a_x}{\sqrt{a_y^2 + a_z^2}}\right)
$$

$$
\gamma = \arctan\left(\frac{a_y}{a_z}\right)
$$

The roll angle $\theta$ and pitch angle $\gamma$ are obtained by the accelerated velocity prior to the transformation of coordinates when a posture matrix between the two angles is obtained and expressed as:

$${C_n^b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \times \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} =
$$

$$
\begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ \sin\theta & \cos\theta & \sin\cos\theta \\ -\sin\theta & -\sin\theta & \cos\cos\theta \end{bmatrix} \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix}
$$

(6)

Where, $b$ is the carrier coordinate system, $n$ is the geographic coordinate system, $H_x$, $H_y$ and $H_z$ are the magnetic deviators output by the inertial sensor with the following relationship:

$$
[H_x \ H_y \ H_z]^T = {C_n^b}[H_x \ H_y \ H_z]^T
$$

(7)

Since the basic rotation matrix is a unit orthogonal matrix, here the following formula exist:

$$
\begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} = \left({C_n^b}\right)^{-1} \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} = \left({C_n^b}\right)^T \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} =
$$
Combining the above formula, the angle of rotation $\Phi$ in static state is obtained [30].

$$\Phi = \frac{H_x}{H_y}$$

### 3.1.2 Analysis of quaternion posture

Static state only accounts for a small part of human motion, in which the majority of human body is in disorder, thus the analysis of dynamic posture is the most frequently used. Quaternion method [31] is the most common in dynamic posture analysis. By gathering the original data of gyroscope, it calculates with quaternion. Based on the collected accelerometer and magnetometer signals, by virtue of the direction between geographic coordinate system and carrier coordinate system of the gravity field and geomagnetic field, cosine is transformed to calculate the corresponding posture angle.

Quaternion $q$, $q = [q_0, q_1, q_2, q_3]^T$, $|q|^2 = q_0^2 + q_1^2 + q_2^2 + q_3^2 + \lambda^2 = 1$ are defined and the transformational relation between quaternion and posture matrix is:

$$\begin{bmatrix} \lambda \\ q_0 \\ q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} \cos \frac{\gamma}{2} \cos \frac{\theta}{2} \cos \frac{\phi}{2} + \sin \frac{\gamma}{2} \sin \frac{\theta}{2} \sin \frac{\phi}{2} \\ \sin \frac{\gamma}{2} \cos \frac{\theta}{2} \sin \frac{\phi}{2} - \cos \frac{\gamma}{2} \sin \frac{\theta}{2} \sin \frac{\phi}{2} \\ \cos \frac{\gamma}{2} \sin \frac{\theta}{2} \cos \frac{\phi}{2} + \sin \frac{\gamma}{2} \cos \frac{\theta}{2} \sin \frac{\phi}{2} \\ \sin \frac{\gamma}{2} \sin \frac{\theta}{2} \sin \frac{\phi}{2} - \cos \frac{\gamma}{2} \cos \frac{\theta}{2} \cos \frac{\phi}{2} \end{bmatrix}$$

By means of data fusion of accelerometer, gyroscope and magnetometer, a more accurate posture angle is solved. Every time a posture algorithm is performed, quaternion is updated correspondingly depending on the fused sensor data. Posture angles $\gamma$, $\theta$ and $\Phi$ are calculated with the updated quaternions.

$$\gamma = -\arctan\left(\frac{2(q_1q_2 - q_0q_3)}{q_0^2 + q_2^2 - q_1^2 - q_3^2}\right)$$

$$\theta = \arcsin\left(2(q_3q_2 + q_0q_1)\right)$$

$$\Phi = -\arctan\left(\frac{2(q_1q_3 - q_0q_2)}{q_0^2 + q_3^2 - q_1^2 - q_2^2}\right)$$

In comparison to the static posture analysis, the calculated amount of quaternion method is small with no singularity, thus it is widely used in real-time posture calculation of motion control system. There are a variety of applications where Quaternion is widely used as an analytical method. Pierononi et al. [32] designed a system to calculate posture and height of center of gravity using MEAS sensors such as MPU6050, HMC5883L, MS5611-01BA for data processing with Kalman filtering. The human body acceleration data are feature extracted by SVM so that quaternion is obtained and converted into euler angle, thereby detection of human direction posture is possible. Rico-Azagra et al. [33] used an explicit complementary filter (ECF) expressed with quaternions to fuse accelerometers or gyroscopes and achieve an accurate attitude estimation of the unmanned ground vehicle employing low-cost inertial measurement units. Brigante et al. [34] put forward a wearable modular system that captures real-time human motion and estimates attitude and gyroscopic bias by using a quaternion-based extended Kalman filter.

### 3.2 Fall Detection

The data of human posture analysis are used for motion detection while a prominent usage of motion detection is to judge the falling state, determining whether the body falls by calculation of features. The
mainstream falling detecting algorithms fall into three kinds: multi-level thresholding, intelligent classification (mainly based on support vector machine) and machine learning-based detection while the third algorithm incorporates two branches: KNN algorithm and BP neural network. The first algorithm needs to store all the training datasets before classification, which makes the operation processing extremely difficult and prone to wrong judgment. The latter is unable to observe the learning process, rendering unexplainable output results and affecting the reliability of results, and it may even fail to achieve the learning purpose. Therefore, this section mainly studies the previous two methods, namely multi-level thresholding and support vector machine.

3.2.1 Multi-level thresholding detection

In the multi-level thresholding, the general human action is first divided into two kinds of violent (jogging, falling, jumping) and ordinary action based on the minimum value of resultant acceleration. The minimum value of the resultant acceleration threshold \( \text{acc}_t \) is further determined. In the case of \( \text{acc} > \text{acc}_t \), it belongs to ordinary action, otherwise it is violent. The resultant angular velocity and resultant attitude angle basically classify the violent action. Next, the resultant angular velocity and resultant attitude angle thresholds, namely the maximum \( \text{gyr}_t \) and \( \text{angle}_t \) are determined. In case of \( \text{gyr} > \text{gyr}_t \) or \( \text{angle} > \text{angle}_t \), it is a falling action, and otherwise it is jumping or jogging. To improve the distinction of action, the intersection of judgment of three thresholds is taken. The resultant acceleration, resultant angular velocity and resultant attitude angle at time \( t \) are recorded as \( \text{acc}(t) \), \( \text{gyr}(t) \) and \( \text{angle}(t) \) respectively, and the basis of judging the upcoming falling are: \( \text{acc}(t) < \text{acc}_t \) and \( \text{gyr}(t) > \text{gyr}_t \) plus \( \text{angle}(t) > \text{angle}_t \)[35][36]. As shown in Figure 5:

![Fig.5. Multi-level Thresholding Detection](image)

Thresholding is widely used. Lindeman[37] proposed three thresholding detecting methods: the sum of acceleration vectors in XY-plane is greater than 2g. The sum of velocity vectors of each spatial component prior to collision is greater than 0.7m/s, and the sum of acceleration vectors of all spatial components is greater than 6g, which adds a basic threshold to the sensor-based fall detecting algorithm and lays the theoretical foundation for the multi-threshold detecting method. Bourke[38] designed a triaxial accelerometer combined with biaxial gyroscope, and set its threshold value to judge whether it is in the falling state. Li Na et al. [39] used Kalman filter to identify the steady and unsteady state, and adopted self-adapting thresholding to identify the running and walking action in the steady state.

3.2.2 Support vector machine detection

SVM makes full use of external factors for statistical analysis and supplements adjustment factors with some weight, especially reducing the complexity of algorithm by linear growth.

The classification interface in Figure 6 correctly distinguishes two kinds of interfaces, and the distance from the optimal page to the two nearest points is the largest, as shown by \( H_1 \) and \( H_2 \) in the Figure. By casting the sample data into classification interval, distinguishing samples and establishing gaps, the entire empirical value reaches the minimum, and more nonlinear data are input when the elongation factors are amplified[40].
Using support vector machine is also very common, and many professionals improved and optimized it on the basis of ordinary vector machine detection. Liu Yong et al. [41] proposed a fall detection technology based on acceleration and dip angle changes, performing SVM data sample training to detect human posture by making use of feature vector. Yang et al. [42] used one class support vector machine (OCSVM) of semi-supervised learning algorithm to detect the unfulfilled falling events. Zhang Jinqiao et al. [43] used AXDL345 sensor and MSP430F149 single-chip as the device data acquisition module and data processing module, proposing a multi-level fall detecting algorithm fusing acceleration and attitude angle. This study supplemented the theoretical research made by Liu Yong et al. [44] that is not complete with ADLs data analysis.

4. COMPARISON AND ANALYSIS

In general, the main detection process of vision and sensor-based motion state detection system are roughly the same. Both are collecting the action features of human body first followed with analytical screening of the collected results and information plus the extraction of corresponding feature values. A series of algorithmic processing distinguishes human motion behaviors. In the information extraction stage, the former mainly uses one or more cameras for video image collection with the conventional camera or a Kinect somatosensory camera [45]; the latter mainly relies on inertial sensors, using a simplex three-axial acceleration sensory sensor or nine-axial sensor to detect accelerated and angular velocities, and the corresponding detecting algorithms are used to process data. In the action classification stage, the main methods for both vision and sensor-based are similar, which generally fall into thresholding method and support vector machine. In view that the detection effect of support vector machine method is superior to the multilevel thresholding method, but the response time lags behind. As the processor performance is further optimized and lifted, the operation speed gets faster and the efficiency is higher. Support vector machine also proves to be an excellent fall detecting algorithm whose main development tends to reduce the sensitivity to missing data and find a general solution to nonlinear problems.

Of various human detection systems, the vision-based detection system has unique superiorities, which is able to detect falling with the help of the increasingly popular home intelligence monitoring device [46]. Relative to the human motion information acquired by sensor, its accuracy is higher with flexible design and simply obtained information; since the obtained image contains plentiful information, this user-friendly system is convenient to use. Main problems facing this method including [47]:

1. The design based on 3D human model is complex and computationally intensive, while other falling detection methods are easily interfered by body sheltering, illumination variation and shadow, although with a relatively small calculation amount.
2. The detecting effect for actions similarly to falling and in parallelism to cameras are to be improved, thus it is necessary to take a certain measure to reduce the
false drop rate. (3) Since it is difficult to obtain the real human falling data directly, all methods are performed under the simulated experimental conditions.

The motion detecting method based on wearable sensors realizes falling detection relying on human motion information through the motion sensor device with a relatively low cost to use. Its wearing convenience enables the user to monitor and measure the human motion information in real time, posing no restrictions on the environment without revealing user’s privacy, thus it is worth being promoted. The single three-axial acceleration sensor is simple and easy to use, only requiring a simpler method to collect fewer human posture parameters. By optimizing and contrasting the algorithm, the detecting accuracy can be improved. The combined nine-axial sensor is able to capture more human parameters to render more accurate and faster analysis. Its main disadvantages are as follows: (1) The relatively low cost of sensor restricts the manufacturing process of various sensors, affecting the accuracy of the data collected. Therefore, to guarantee the data accuracy, in addition to the complementary advantages by using each sensor in the later stage, it is necessary to compensate and calibrate the data at the initial stage of data acquisition[48]. (2) During the measurement process, the influence of mechanical vibration and installation environment will inevitably cause noises to the sensors [49].

5. CONCLUSION AND PROSPECT

Motion state detection technology serves as a crucial component of the smart home medical monitoring field. The two major implementation technologies today incorporate the vision-based and sensor-based motion state detection, both of which are developing to a mature and dominant position. The basic comparison is shown in Table 1.

|                           | Based on the visual | Based on the sensor |
|---------------------------|--------------------|--------------------|
| **Hardware**              | Camera             | Inertial sensor    |
| **Information acquisition speed** | Fast              | Fast              |
| **The amount of information** | Sufficient       | Insufficient      |
| **Cost**                  | High               | Low               |
| **Fault tolerance**       | High               | Low               |
| **Amount of calculation** | Large              | Petty             |
| **The user experience**   | Relatively poor (privacy exposed) | Relatively good (easy to wear) |
| **Acquisition precision** | High               | Low               |

The sensor-based detection method mostly adopts the thresholding method, thus being convenient to design and use, low in cost, easy and flexible to install. It is also suitable to carry with a wide application range, unlikely to expose the privacy of users. However, the sensor needs to be carried for long, posing a certain impact on human activities. The vision-based detection method does not require long-time wearing and acquires abundant human motion information. With sophisticated image processing technology, a variety of falling detection methods with high accuracy are designed, which can detect falling with the help of the commonly used intelligent monitoring equipment. However, it is computationally intensive, the acquisition of relatively accurate results requires a large amount of calculation or training, and there may be cases when the judgment is inaccurate in an environment with sheltering. On the whole, each of the two methods is a double-edged sword, which should be used in conjunction with specific scenarios.
The future development of the sensor-based detection technology mainly lies in three aspects[50]. The first one improves the manufacturing process of the sensor and improve the measurement accuracy and stability of the sensor. The second fuses multi-sensor data. Targeting at the problem that single sensor data cannot obtain acceptable recognition accuracy, various sensors including acceleration sensor, magnetic sensor, gravitational acceleration sensor, gyro sensor, and air pressure sensor are integrated to collect the motion state data of people indoors and go through data processing. The third aspect combines the sensor with the multi-layer neural network for multiple learning on the multi-sensor data, adjust network parameters, and form the indoor motion state recognition model with a strong recognition capability.

Besides, the future development direction of computer vision-based detection technology mainly lies in three aspects. The first fuses the multi-scale information in feature extraction, deep mining the hidden connections within data and improving the accuracy of data judgment. The second aspect adds the spatial-temporal information of feature occurrence while using the classification algorithm, considering the fusion of multi-source complementary information of different levels such as features and decisions from different dimensionalities such as time and reasoning to improve the detecting accuracy. The third aspect links the technology to the environmental layout to solve the state detection problem when the object is sheltered.

Moreover, with the popularization of smartphone and the growth of elderly users, it has become a possible future trend to apply the smartphone-based falling detection[51]. The method of collecting motion state data through smartphone is very effective and close to life in the experiment and real environment. Although the sensor accuracy in smartphone is low, it is convenient to gather a large amount of motion state information, and to identify the high-accuracy motion state by extracting features in different motion states. As artificial intelligence advances, smart mobile phone will be making greater progress and the calculated quantity of data will be growing. Consequently, the on-line identification of motion state will eventually be realized. With smartphone, integration of processes such as data acquisition, processing, model training and classification recognition is available, and a complete motion state identification system is formed. What’s more, machine learning is also a tendency for future development in the field of falling detection, and the computing resources required by its specific application pose a severe challenge for the processor and memory of mobile phones. How to implement an efficient algorithm with low resource demands on smartphone is also a task to be probed into.

REFERENCES

[1] Gravina, R., Alinia, P., Ghasemzadeh, H., & Fortino, G. (2017). Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. Information Fusion, 35, 68-80.

[2] Zhang, C., Tian, Y., & Capezuti, E. (2012, July). Privacy preserving automatic fall detection for elderly using RGBD cameras. In International Conference on Computers for Handicapped Persons (pp. 625-633). Springer, Berlin, Heidelberg.

[3] Seki, M., Fujiwara, H., & Sumi, K. (2000). A robust background subtraction method for changing background. In Proceedings Fifth IEEE Workshop on Applications of Computer Vision (pp. 207-213). IEEE.

[4] Brown, R. G., & Hwang, P. Y. (1992). Introduction to random signals and applied Kalman filtering (Vol. 3). New York: Wiley.

[5] Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123(3), 585-595.

[6] Ijima, E. P., & Chalavadi, K. M. (2016). Human action recognition using genetic algorithms and convolutional neural networks. Pattern recognition, 59, 199-212.

[7] Barnich, O., & Van Droogenbroeck, M. (2010). ViBe: A universal background subtraction algorithm for video sequences. IEEE Transactions on Image processing, 20(6), 1709-1724.

[8] Bao, W., Li, H., Li, N., & Jiang, W. (2009, April). A liveness detection method for face recognition based on optical flow field. In 2009 International Conference on Image Analysis and Signal Processing (pp. 233-236). IEEE.
[9] Ke, Y., & Sukthankar, R. (2004). PCA-SIFT: A more distinctive representation for local image descriptors. CVPR (2), 4, 506-513.
[10] Llorca, D. F., Arroyo, R., & Sotelo, M. A. (2013, October). Vehicle logo recognition in traffic images using HOG features and SVM. In 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013) (pp. 2229-2234). IEEE.
[11] Jain, A. K., & Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. Pattern recognition, 24(12), 1167-1186
[12] Wang, X., Han, T. X., & Yan, S. (2009, September). An HOG-LBP human detector with partial occlusion handling. In 2009 IEEE 12th international conference on computer vision (pp. 32-39). IEEE.
[13] Mita, T., Kaneko, T., & Hori, O. (2005, October). Joint haar-like features for face detection. In Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1 (Vol. 2, pp. 1619-1626). IEEE.
[14] Yan, J., Lei, Z., Wen, L., & Li, S. Z. (2014). The fastest deformable part model for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2497-2504).
[15] Belongie, S., Malik, J., & Puzicha, J. (2001). Shape context: A new descriptor for shape matching and object recognition. In Advances in neural information processing systems (pp. 831-837).
[16] Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D., & Li, S. Z. (2014, September). Salient color names for person re-identification. In European conference on computer vision (pp. 536-551). Springer, Cham.
[17] Klein, D. A., & Frintrop, S. (2011, November). Center-surround divergence of feature statistics for salient object detection. In 2011 International Conference on Computer Vision (pp. 2214-2219). IEEE.
[18] Toshev, A., & Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1653-1660).
[19] Qian, Y., Fan, Y., Hu, W., & Soong, F. K. (2014, May). On the training aspects of deep neural network (DNN) for parametric TTS synthesis. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 3829-3833). IEEE.
[20] Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 806-813).
[21] Sahoo, P. K., Soltani, S. A. K. C., & Wong, A. K. (1988). A survey of thresholding techniques. Computer vision, graphics, and image processing, 41(2), 233-260.
[22] Igual, R., Medrano, C., & Plaza, I. (2013). Challenges, issues and trends in fall detection systems. Biomedical engineering online, 12(1), 66.
[23] Bian, Z. P., Chau, L. P., & Magnenat-Thalmann, N. (2012). A depth video approach for fall detection based on human joints height and falling velocity. International Conference on Computer Animation and Social Agents.
[24] Planinc, R., & Kampel, M. (2013). Introducing the use of depth data for fall detection. Personal and ubiquitous computing, 17(6), 1063-1072.
[25] Zhang, C., Tian, Y., & Capezuti, E. (2012, July). Privacy preserving automatic fall detection for elderly using RGBD cameras. In International Conference on Computers for Handicapped Persons (pp. 625-633). Springer, Berlin, Heidelberg.
[26] Bian, Z. P., Hou, J., Chau, L. P., & Magnenat-Thalmann, N. (2014). Fall detection based on body part tracking using a depth camera. IEEE journal of biomedical and health informatics, 19(2), 430-439.
[27] Zhang, Z., Liu, W., Metsis, V., & Athitsos, V. (2012, November). A viewpoint-independent statistical method for fall detection. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012) (pp. 3626-3630). IEEE.
[28] Kepski, M., & Kwolek, B. (2012, September). Human fall detection by mean shift combined with depth connected components. In International Conference on Computer Vision and Graphics (pp. 457-464). Springer, Berlin, Heidelberg.

[29] Kelly Jr, P. B., Schoendorfer, D. W., & Simmons, J. L. (2003). U.S. Patent No. 6,611,783. Washington, DC: U.S. Patent and Trademark Office.

[30] Shi, G., Zou, Y., Jin, Y., Cui, X., & Li, W. J. (2009, February). Towards HMM based human motion recognition using MEMS inertial sensors. In 2008 IEEE International Conference on Robotics and Biomimetics (pp. 1762-1766). IEEE.

[31] Carminati, M., Ferrari, G., Grassetti, R., & Sampietro, M. (2012). Real-time data fusion and MEMS sensors fault detection in an aircraft emergency attitude unit based on Kalman filtering. IEEE Sensors Journal, 12(10), 2984-2992.

[32] Pierleoni, P., Belli, A., Maurizi, L., Palma, L., Pernini, L., Paniccia, M., & Valenti, S. (2016). A wearable fall detector for elderly people based on ahrs and barometric sensor. IEEE sensors journal, 16(17), 6733-6744.

[33] Rico-Azagra, J., Gil-Martínez, M., Rico-Azagra, R., & Maisterra, P. (2016). Low-cost attitude estimation for a ground vehicle. In Robot 2015: Second Iberian Robotics Conference (pp. 121-132). Springer, Cham.

[34] Brigante, C. M., Abbate, N., Basile, A., Faulisi, A. C., & Sessa, S. (2011). Towards miniaturization of a MEMS-based wearable motion capture system. IEEE Transactions on industrial electronics, 58(8), 3234-3241.

[35] Chen, J., Kwong, K., Chang, D., Luk, J., & Bajcsy, R. (2006, January). Wearable sensors for reliable fall detection. In 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference (pp. 3551-3554). IEEE.

[36] Mubashir, M., Shao, L., & Seed, L. (2013). A survey on fall detection: Principles and approaches. Neurocomputing, 100, 144-152.

[37] Lindeman U. Evaluation of a Fall Detector Based on Accelerometers: a Pilot Study[J]. Med Biol Eng Comput, 2005.

[38] Bourke, A. K., & Lyons, G. M. (2008). A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. Medical engineering & physics, 30(1), 84-90.

[39] Li, N. (2013). Research on wearable health monitoring system based on human activity recognition. Beijing University of Technology.

[40] Zhang, T., Wang, J., Liu, P., & Hou, J. (2006). Fall detection by embedding an accelerometer in cellphone and using kfd algorithm. In IJCSNS International Journal of Computer Science and Network Security.

[41] Bourke, A. K., & Lyons, G. M. (2008). A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. Medical engineering & physics, 30(1), 84-90.

[42] Hong, Y. J., Kim, I. J., Ahn, S. C., & Kim, H. G. (2010). Mobile health monitoring system based on activity recognition using accelerometer. Simulation Modelling Practice and Theory, 18(4), 446-455.

[43] Zhang, T., Wang, J., Xu, L., & Liu, P. (2006). Fall detection by wearable sensor and one-class SVM algorithm. In Intelligent computing in signal processing and pattern recognition (pp. 858-863). Springer, Berlin, Heidelberg.

[44] Yang, K., Ahn, C. R., Vuran, M. C., & Aria, S. S. (2016). Semi-supervised near-miss fall detection for ironworkers with a wearable inertial measurement unit. Automation in Construction, 68, 194-202.

[45] Stone, E. E., & Skubic, M. (2014). Fall detection in homes of older adults using the Microsoft Kinect. IEEE journal of biomedical and health informatics, 19(1), 290-301.

[46] Baek, W. S., Kim, D. M., Bashir, F., & Pyun, J. Y. (2013, January). Real life applicable fall detection system based on wireless body area network. In 2013 IEEE 10th Consumer Communications and Networking Conference (CCNC) (pp. 62-67). IEEE.
[47] Suh, C., & Ko, Y. B. (2008). Design and implementation of intelligent home control systems based on active sensor networks. IEEE Transactions on Consumer Electronics, 54(3), 1177-1184.
[48] Yang, L., Ren, Y., & Zhang, W. (2016). 3D depth image analysis for indoor fall detection of elderly people. Digital Communications and Networks, 2(1), 24-34.
[49] Zhang, X., Li, Y., Mumford, P., & Rizos, C. (2008, November). Allan variance analysis on error characters of MEMS inertial sensors for an FPGA-based GPS/INS system. In Proceedings of the International Symposium on GPS/GNSS (pp. 127-133).
[50] López-Nava, I. H., & Muñoz-Meléndez, A. (2016). Wearable inertial sensors for human motion analysis: A review. IEEE Sensors Journal, 16(22), 7821-7834.
[51] Zhang, C., Tian, Y., & Capezuti, E. (2012, July). Privacy preserving automatic fall detection for elderly using RGBD cameras. In International Conference on Computers for Handicapped Persons (pp. 625-633). Springer, Berlin, Heidelberg.