Fall detection via human posture representation and support vector machine

Kaibo Fan, Ping Wang, Yan Hu and Bingjie Dou

Abstract

Accidental falls of elderly people are a major cause of fatal injuries, especially for those living alone. We present a novel vision–based fall detection approach that analyzes an extracted human body using described human postures. First, a human body extracted by a background subtraction technique is located by a minimum area-enclosing ellipse. Then, a normalized directional histogram is developed around the center of the ellipse to represent a human posture by multi-directional statistical analysis. After that, 12 static and 8 dynamic features are derived from the normalized directional histogram. These features are fed into a directed acyclic graph support vector machine to distinguish four closely related human postures (standing, crouching, lying, and sitting). A fall-like accident is detected by counting the occurrences of lying postures in a short temporal window. After conducting majority voting, a fall event is determined by immobility verification. From the experimental results, an overall accuracy of 97.1 is obtained for recognition of the four postures, and only 1.0% of postures are misclassified as lying postures. Our fall detection system achieves up to 95.2% fall detection accuracy on a public fall dataset.

Introduction

Falls and fall-related injuries are a major public health problem for aging populations all over the world. According to the statistics of the World Health Organization,[1] approximately 28%–35% of people aged 65 and over fall every year, and this rate increases to 32% or even 42% for those over 70 years old. This makes falls one of the five most common causes of death among elderly people.[2–4] The medical expenses caused by the immense number of falls occurring every year have become a heavy burden for the population not only of China but also of other countries with the aging problem. It has already been proven that the medical consequences of a fall incident are highly dependent on the response and rescue time. However, commercial fall detection systems are mostly based on wearable sensors, which elderly people may forget to wear. It is not very convenient for them to carry these devices at all times. The main drawback of such devices is obvious. The demand for fall detection surveillance systems, which can automatically monitor and analyze abnormal behaviors of elderly people, has increased within the healthcare industry with the rapid growth of the population of the elderly. Thus, a highly accurate automatic fall detection system may become a

School of Electrical and Information Engineering, Tianjin University, Tianjin, P.R. China

Corresponding author:
Kaibo Fan, School of Electrical and Information Engineering, Tianjin University, No. 92 Weijin Road, Tianjin 300072, P.R. China.
Email: fankaibo@tju.edu.cn

significant part of the smart living environment for elderly people living alone.

More specifically, the monocular vision–based approach plays an irreplaceable role in fall detection. Monocular vision–based systems are very cheap and easy to set up. Moreover, many other activities besides fall incidents can be detected simultaneously with less intrusion. Each event lasts for a short period of time and contains different types of postures. During this short period, human postures change considerably with high velocity, especially in a fall incident. Due to this observation, a fall event can be detected by distinguishing these postures. Other actions, such as crouching, lying, and sitting, may have some postures similar to a fall incident, but they have entirely different semantic contents. Knowing only postures is not enough to distinguish these similar motions accurately; therefore, in our approach, we also consider the temporal information between postures.

This article proposes a novel posture–based framework for fall detection inspired by the observation that different motions have different characteristics reflected by postures. For distinguishing a fall incident from other daily activities, the majority voting strategy is utilized to combine several classified postures in a short temporal window. Our main contributions are summarized as follows:

(1) The foreground human body extraction is reported in section "Human body extraction." The human silhouette is refined and covered by a minimum area-enclosing ellipse, which is more compact than the traditional rectangular box for describing human posture.
(2) To describe the postures of a fall incident, a normalized directional histogram (NDH) is developed from the covered silhouette by multi-directional statistical analysis. In light of the NDH, 12 static and 8 dynamic features are also derived for posture recognition.
(3) To detect a fall incident, fall-related postures (standing, crouching, lying, and sitting) are classified by a directed acyclic graph support vector machine (DAGSVM). We accumulate the lying postures in a short temporal window to filter fall-like accidents. Once a fall accident is verified by immobility detection, the final decision is made.

The remainder of this article is organized as follows. Section "Related work" provides an overview of related works. In section "Human body extraction," we briefly describe the extraction of the human body implemented with the help of a background model. Section "Human posture description" gives an elaborate explanation of the posture description, which includes the

human body location, the NDH representation of postures, and the statistical characteristics of the NDH. Section "Fall event detection" explains the classifier of the fall detection, including the postures related to fall incidents, the directed acyclic graphic strategy of multiclass support vector machine (SVM) used for the four postures classification, and the fall event validation. In section "Experimental results," several experiments are conducted to show that our algorithm works not only in the case of fall incident detection but also in detecting other complicated human postures. Some guidelines on fall detection are presented in section "Discussion." Finally, section "Conclusion" contains the conclusion.

Related work

Current solutions to the fall detection problem can be roughly divided into two categories: non-computer vision–based methods and computer vision–based methods:[5,6]

(1) *Non-computer vision–based methods.* There are many non-computer vision–based methods of fall detection, such as sensitive floor tiles,[7] simple sensors,[8] and wearable sensors.[9–11] As falls cannot be detected at locations not equipped with specialized tiles, these tiles should be installed everywhere in the living room. These simple sensors merely provide some raw data and do not give sufficient information of a fallen person. The error rate of the method is also very high. The problem of wearable detectors is that the elderly may easily forget to wear them. Moreover, such a device is useless once the battery power is used up. Recently, to alleviate this problem, some researchers have proposed the use of acoustic and Doppler radar for fall detection.[12,13] Moreover, there is a new trend toward the use of mobile fall detection systems.[14,15] These systems also make it possible to prevent and anticipate many other health hazard situations apart from fall detection and notification.
(2) *Computer vision–based methods.* Great progress has been made in computer vision and image processing techniques in recent decades, which has opened up new opportunities to improve fall detection systems. Image processing plays an indispensable role in fall detection systems because of its evident advantages. According to the principles of fall characteristics, they can be grouped into three categories: inactivity detection, moving human body shape change analysis (silhouette analysis), and 3D head motion analysis.[16] In the vision-based fall detection

domain, Lee and Mihailidis[17] initially used the centroid, perimeter, and principal axis of the silhouette to describe a series of fall human postures and then identified required postures by a preset threshold value. Lee's system was able to discriminate standing, stooped over, and lying down postures. Liu et al.[18] employed the ratio and differences of width and height of the bounding box of a human body silhouette to classify postures into three categories, namely standing postures, temporary postures, and lying down postures. The performance of their system was promising when the camera was placed sideways. A shape-matching technique,[19] such as shape context, Procrustes shape analysis, was proposed to match different postures. A fall event could be determined by analyzing the human shape deformation during a video sequence using a Gaussian mixture model. The occlusion problem was also partly solved. As for Auvinet et al.,[20] to detect fall incidents of seniors, they utilized an occlusion-resistant method based on multi-camera networks, which reconstructs the three-dimensional shapes of people. Fall events were also determined by analyzing the volume distribution along the vertical axis. An alarm was triggered when there was an anomaly in the distribution during a predefined period. Increasing the number of cameras used, we can enhance the complexity of the final detection result. Yu et al.[21] applied the background subtraction technique to extract the foreground. The moving object was located in the image plane by an ellipse, the parameters of which were obtained by computing several spatial moments of the foreground image. In describing human postures, the ellipse fitting was more accurate than the rectangle box. A projection histogram along the axis of the ellipse (the local features) and the ratio between the major axis and the minor axis (the global feature) could evidently distinguish the postures of a fall incident. Olivieri et al.[22] presented a spatio-temporal motion representation, called motion vector flow instance (MVFI) template, which captured relevant velocity information by extracting the dense optical flow from a video sequence of human actions. As both the magnitude and direction of the velocity were fed into MVFI, falls could be distinguished from daily human motions with high accuracy and computational efficiency. Mirmahboub et al.[23] proposed the use of variations in the silhouette area obtained from only one camera. View-invariant features of the human body region were fed into a classifier of different events, mainly focused on the detection of fall accidents. Chua et al.[24] also proposed a new simple vision-based posture description technique that was based on human body shape variation analysis. Only three key points were used instead of the traditional ellipse and rectangle box. The fall was detected by analyzing the shape changes of the human silhouette through the centroids of three different regions of the foreground. The proposed three points of the human posture representation technique improved the fall detection rate without increasing the computational complexity. A novel posture representation approach, histogram of maximal optical flow projection (HMOFP), was presented in Li et al.[25] HMOFP mainly concerns the motion features of abnormal events, such as falling, running, and crouching, that occur in the crowded scenes. Curvature scale space (CSS) features and the bag-of-words (BOW) method[26] were combined to detect a fall in a depth video. An improved extreme learning machine (ELM) classifier was adopted to distinguish falls and non-falls. In a later work,[27] instead of representing an action as a bag of CSS words, the description of an action was provided using Fisher Vector (FV) encoding on the basis of CSS features. The Microsoft Kinect was also applied to fall detection;[28] a person's vertical states in the depth images were characterized in the first stage, and then an ensemble of decision trees was used in the second stage to compute the confidence that a fall had preceded a ground event.

Our work is slightly different from the above-mentioned vision-based detection features used for fall detection. We present a novel vision–based approach for monitoring elderly people that is focused on detecting falls using different human postures. The representation of a human posture is based on the NDH. It not only considers the spatial information of a human posture but also concerns the motion information by analyzing the change rate of the NDH. Then, a multi-class SVM is used to classify postures into four types (standing, crouching, lying, and sitting) that are closely related to fall event detection, especially the lying postures. Figure 1 shows the flow chart of the proposed fall detection approach. Briefly speaking, the approach involves three main blocks. First, the human body is extracted from each frame using a background model. Second, the NDH is built as an information source to represent a human posture. Third, postures are classified by the DAGSVM. After all the postures are
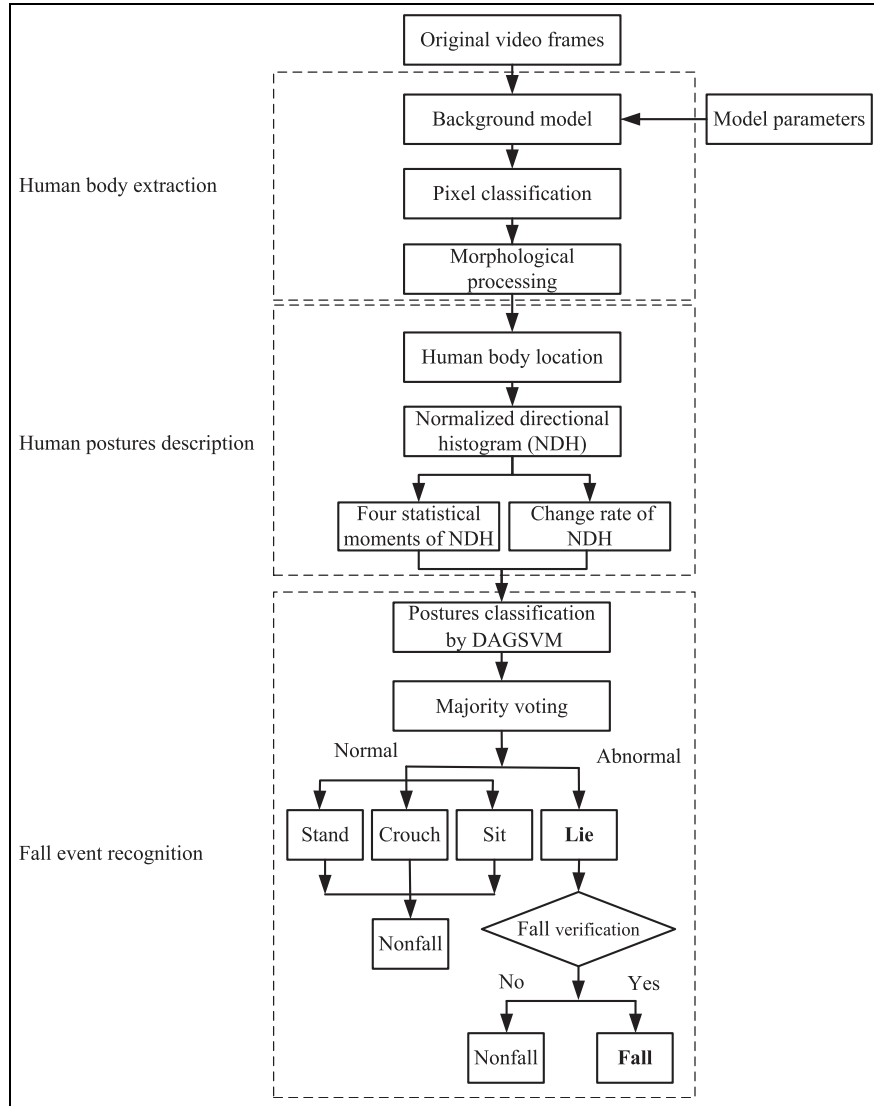
**Figure 1.** Flow chart of the proposed fall detection approach.

classified, the fall accident is verified by a majority voting strategy and immobility detection. In the following sections, we will describe the details of the three blocks.

Human body extraction

Extracting a moving person from an image sequence is one of the most challenging tasks in the computer vision field. Human bodies are highly non-rigid objects with a high degree of variability in size and shape. When people walk toward or away from a video camera, both the shape and size of the human body change greatly. Sometimes, even the color and texture are affected significantly by the shadow or ambient light in a living room. The approach to human body extraction has to cope with such complex situations.

Our initial assumption is that there is only one moving object, the walker, in the video sequence. The camera and the background are always static. Under this assumption, the background model is obtained by computing several model parameters over a number of static background frames. We employ the color distortion model proposed by Horprasert et al.,[29] which can be used to address the problem of slight illumination changes, such as shadows and highlights. The color distortion model is adopted to separate the brightness from the chromaticity component. Figure 2 shows the color distortion model in three-dimensional RGB color space.

Considering a pixel i in the frame, let $E(i) = [E_R(i), E_G(i), E_B(i)]$ represent the pixel's expected RGB color value in the background model, and let $I(i) = [I_R(i), I_G(i), I_B(i)]$ denote the pixel's RGB color
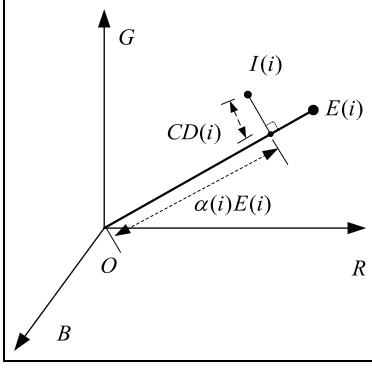
Figure 2. Color distortion model.

value in the current image that needs to be subtracted from the background. The distortion of $I(i)$ from $E(i)$ is decomposed into two parts, brightness distortion $BD(\alpha(i))$ and color distortion $CD(i)$. The brightness distortion $BD(\alpha(i))$ is a scalar value that brings the observed color value close to the expected chromaticity line. It is obtained by minimizing

$$BD(\alpha(i)) = \|I(i) - \alpha(i)E(i)\|_2^2 \qquad (1)$$

$\alpha(i)$ represents the pixel's brightness strength with respect to the expected pixel value. $\|\cdot\|_2$ stands for the two-norm. The color distortion of pixel $I(i)$ is defined as the distance between the observed color and the expected chromaticity line, which is given by

$$CD(i) = \|I(i) - \alpha(i)E(i)\|_2 \qquad (2)$$

There are three main steps in human body extraction. The first step is to construct a reference background image using the background model. Second, threshold selection determines the appropriate threshold values for pixel classification. The final step is to classify the pixels into the background mask, moving object mask, and shadow mask.

These three steps yield a binary foreground image. However, the foreground image may be corrupted by bad noise both inside and outside of the object. The noise can even create small holes in the object. Morphological operations are implemented to remove the noise. The crucial steps of human body extraction are demonstrated in Figure 3.
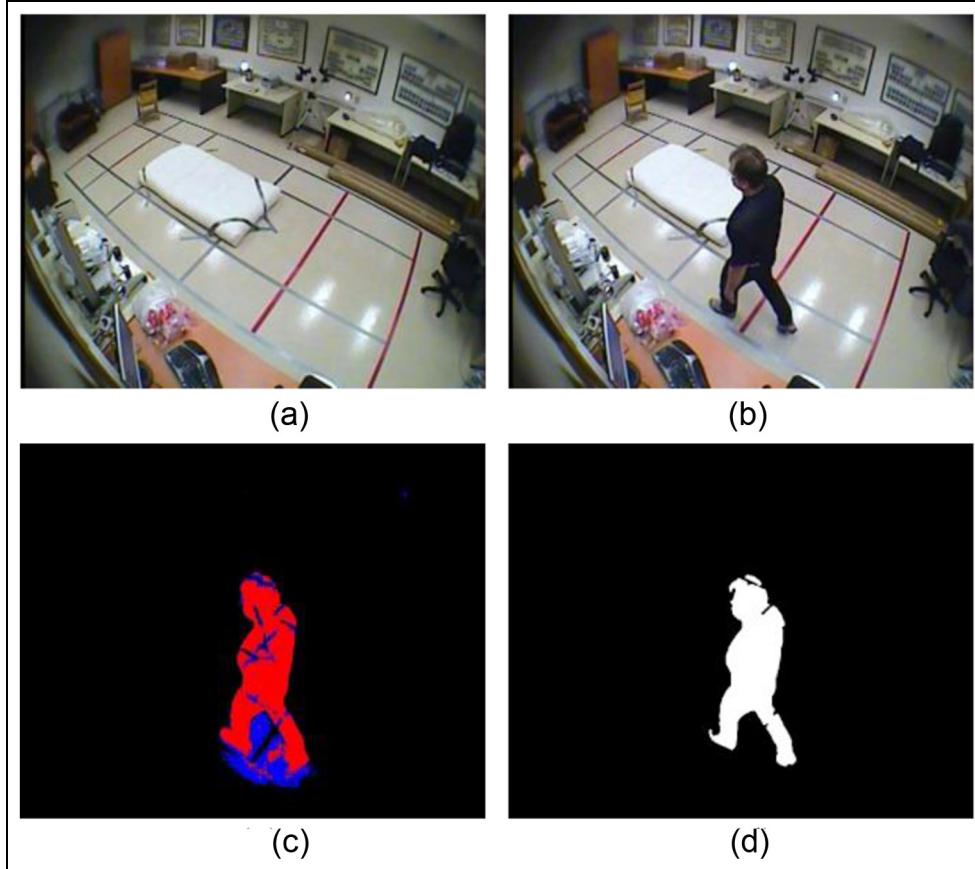
Figure 3. Human body extraction: (a) an original background image, (b) the current image with a human body, (c) the extracted human body in the foreground, including shadows and holes, and (d) the final extracted human object after the morphological operations.
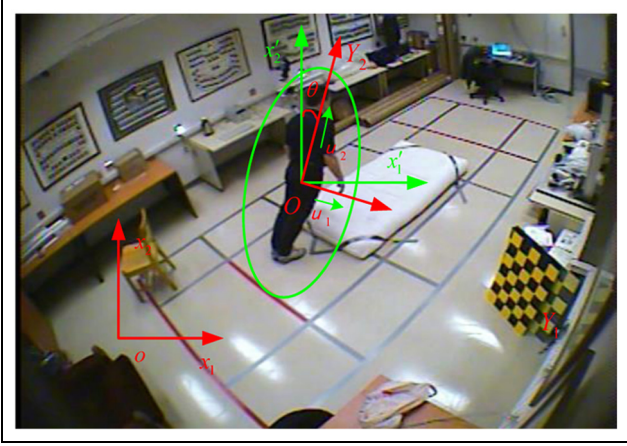
Figure 4. Region of interest.

Human posture description

To conduct a thorough analysis of human movement, the highly non-rigid moving human body should be identified within frames. For an object separated into several small blocks, since it moves rapidly on a similar background along with the human body, we gather all the extracted pixels together into a point set, and identification of the human body in the image plane can be carried out by optimizing the minimum area-enclosing ellipse of these points.

Human body location

Consider the equation of an ellipse in the image plane $px_1^2 + 2qx_1x_2 + rx_2^2 = 1$ and transform the x_1ox_2 coordinate into the centroid of the human body. This equation can be formulated as the new coordinate $x_1'o'x_2'$ as shown in Figure 4

$$[x_1' \quad x_2']E\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = 1 \tag{3}$$

where

$$E = \begin{bmatrix} p & q \\ q & r \end{bmatrix}$$

which can be solved by the Khachiyan first-order algorithm in Kumar and Yildirim.[30] Equation (3) can be rewritten as

$$[x_1' \quad x_2']\begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix}\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = 1 \tag{4}$$

where the u terms are the components of the unit eigenvectors of the symmetric matrix, E, and $E = U\Lambda U^{-1}$

$$U = [u_1 \quad u_2] = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

Since the unit eigenvectors of a real symmetric matrix are orthogonal, let the direction of u_1 parallel the x_1' axis and the direction of u_2 parallel the x_2' axis. Based on this observation, rewrite the matrix of the eigenvectors in equation (4) as

$$\begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix}\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

This means that the matrix of unit eigenvectors for a symmetric 2×2 matrix E can be interpreted as a rotation matrix that relates the coordinate in one orthogonal reference frame (here the x_1', x_2' reference frame) to the coordinate in an orthogonal reference frame along axes defined by the eigenvectors (here, the Y_1, Y_2 reference frame), where

$$\begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{bmatrix}\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

This allows the equation of the ellipse to be expressed in the Y_1, Y_2 reference frame as

$$\lambda_1 Y_1^2 + \lambda_2 Y_2^2 = 1 \tag{5}$$

The reciprocal values of square roots of the eigenvalues for the two unit eigenvectors are the lengths of the semi-major and semi-minor axes of the ellipse. This is a rather important result for the representation of a human body.

NDH of human postures

An example of the ellipse of interest determined by the locating algorithm on the foreground image is given in Figure 5.

We set the center of the ellipse as the origin of the coordinate $x_1'o'x_2'$ and divide the coordinate plane into N different sub-regions according to the directions around the center of the ellipse, $R = \{R_1, R_2, \ldots, R_N\}$. Multi-directional statistical analysis is conducted in every sub-area of the posture sequence.

Given a T-frame sequence of the foreground pixel's position $X = \{X^1, X^2, \ldots, X^T\}$, where X^t is the two-dimensional coordinate of the pixel in the t frame, we can introduce the following cumulative frequency occurrences Bin_i^t of the pixel's position in region R_i ($i = 1, \ldots, N$; in this article, $N = 8$) from a frame sequence

$$Bin_i^t = \frac{\#\{X^t \cap R_i\}}{\sum_{i=1}^{N} \#\{X^t \cap R_i\}} \tag{6}$$
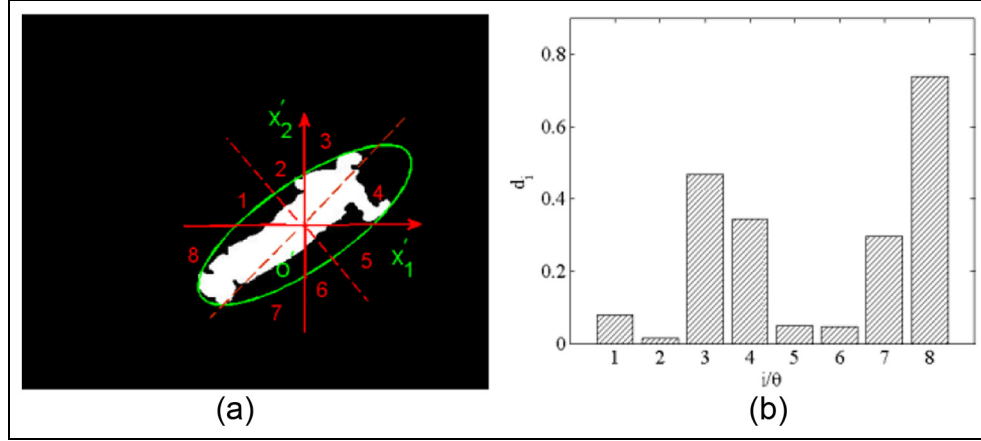
Figure 5. The divided region and normalized directional histogram: (a) the divided region of the human body and (b) the normalized directional histogram of the posture built according to the coding (numbers 1–8).

where # is the cardinal of the sub-region. All of the Bin_i^t pooled together can formulate a posture description basis vector as $D = [Bin_1, Bin_2, \ldots, Bin_8]$ in the tth frame. There is another formation called the NDH in equation (7)

$$\hat{D} = [d_1, d_2, \ldots, d_8] = \frac{D}{\|D\|_2} \tag{7}$$

It can be used for a description of the human posture in the region covered by the ellipse. It is obvious that a human posture has a close relationship to the distribution of the components of $\hat{D}$. We regard $\hat{D}$ as an important source that reflects different human postures.

To give a better understanding of our method, we visualize the NDH of a human posture in Figure 5(b). We also illustrate other concerned NDHs of four closely related postures of a fall incident in Figure 6. Concerning the distribution of the components of $\hat{D}$, we observe that there are some valuable advantages of the NDH. First, the NDH $\hat{D}$ is obtained by analyzing the pixels in every sub-region of an ellipse of interest. It avoids the deviations of posture descriptions caused by the fragmentariness of the extracted object. This approach is effective in determining a human posture notwithstanding the difference in size of the extracted object caused by the variation in distance between the camera and object and by fine changes of visual angle. Second, because of the particular characteristics of the silhouettes of walking or standing people, there are two groups of high bars in the NDH $\hat{D}$. The phase between the two groups of high bars is approximately 180°, while the first group of high bars is approximately 90°. There are similar characteristics when people lie down, but in this case, the phase of the first group of high bars is not 90°. The variance of NDH $\hat{D}$ in the two cases is definitely larger than that of the crouching and sitting postures. Third, by analyzing the NDH $\hat{D}$, we can find that there is a significant difference not only in variance but also in three other aspects, namely mean, skewness, and kurtosis. In the next subsection, we will carry out a deep statistical analysis of NDH $\hat{D}$ and perform an efficiency analysis of human posture classification.

Statistical characteristics of NDHs

Obviously, the mean of NDH $mean(\hat{D})$ expresses the central tendency, while the standard variance $std(\hat{D})$ demonstrates the measures of dispersion of the components. Skewness $skewness(\hat{D})$ is a measure for assessing the asymmetry of the histogram. Kurtosis $kurtosis(\hat{D})$ reflects the shape of the histogram, that is, whether the histogram is peaked or flat compared to a normal distribution. Skewness and kurtosis of the NDH are defined as follows

$$skewness(\hat{D}) = \frac{\sum_{i=1}^{N}(d_i - \mu)^3}{(N-1)\sigma^3},$$

$$kurtosis(\hat{D}) = \frac{\sum_{i=1}^{N}(d_i - \mu)^4}{(N-1)\sigma^4} \tag{8}$$

where μ and σ are the mean and standard variance of $\hat{D}$, respectively. If the left tail is more pronounced than the right tail, the histogram is said to have a negative skewness. If the reverse is true, it has a positive skewness. If the two are equal, it has zero skewness.[31] A negative value for the skewness indicates that the posture is skewed left, and a positive value for the skewness indicates that the posture is skewed right. Any symmetric posture should have a skewness near zero. For kurtosis, the histogram with high kurtosis tends to have a
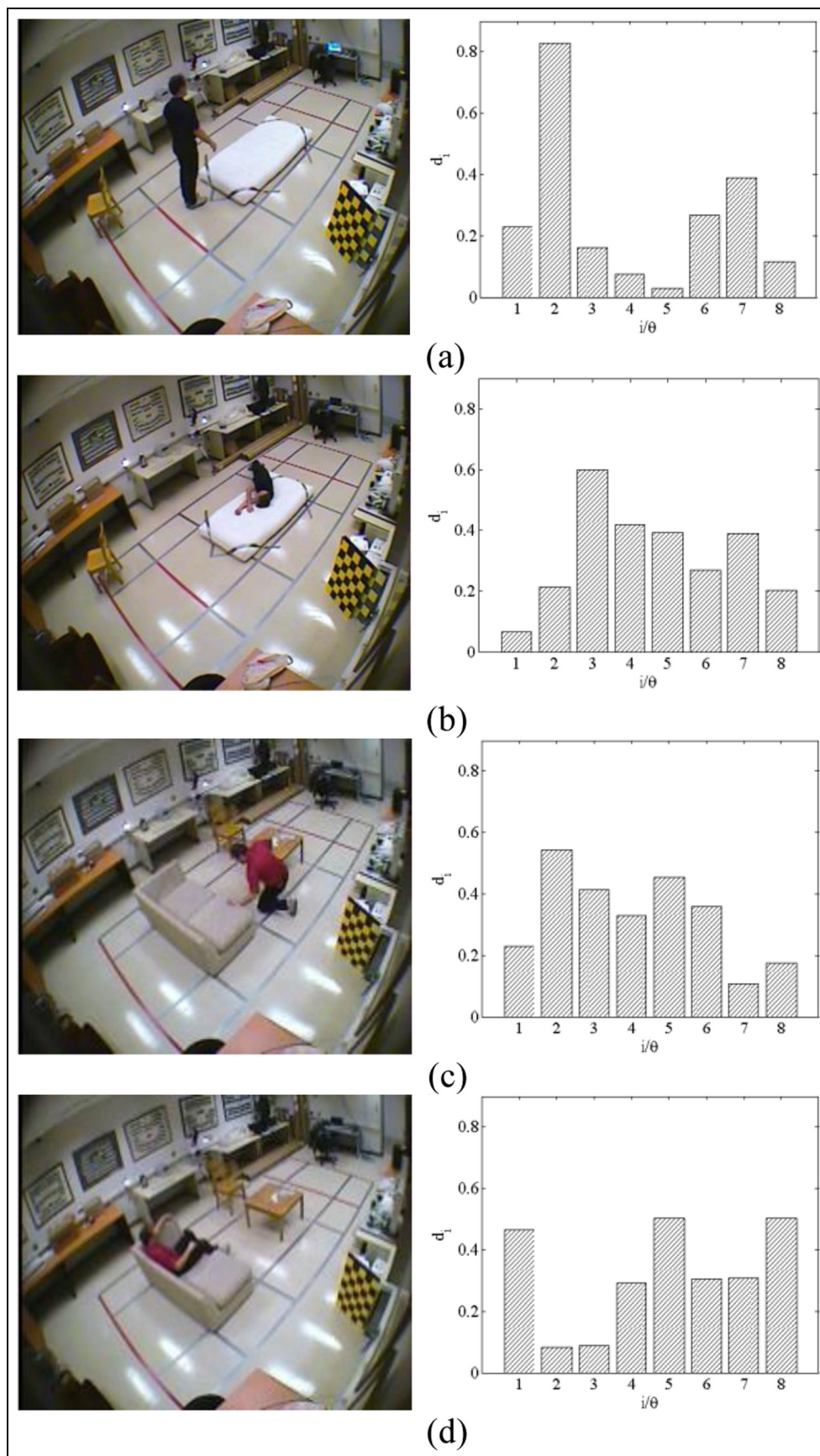
Figure 6. Normalized directional histograms of four postures: (a) standing, (b) lying, (c) crouching, and (d) sitting.

distinct peak near the mean, decline rather rapidly, and have thick tails. A histogram that has low kurtosis tends to have a flat top near the mean rather than a sharp peak. The value of kurtosis can indicate whether the shape of the histogram matches the Gaussian distribution. An even distribution has a kurtosis less than three, a peaked distribution has a kurtosis more than three,[31] and the Gaussian distribution has a kurtosis of three.

To evaluate four statistical moments of $\hat{D}$ (mean, standard variance, skewness, and kurtosis) and discriminate the four postures (standing, lying, crouching, and sitting), we manually choose 4000 postures (including 1000 standing postures, 1000 lying postures, 1000 crouching postures, and 1000 sitting postures) from the multi-view fall dataset.[32] The histogram is a useful graphical technique for obtaining a visual representation of the postures' characteristics. It shows clearly the significant difference among the mean $d_m = mean(\hat{D})$, standard variance $d_v = std(\hat{D})$, skewness $d_s = skewness(\hat{D})$, and kurtosis $d_k = kurtosis(\hat{D})$ of the posture dataset.

The histograms of the mean d_m, stand variance d_v, skewness d_s, and kurtosis d_k of the four postures are given in Figure 7. To investigate more thoroughly the discrimination of one feature for different postures, we take the hypothesis testing approach as a theoretical tool for analyzing the six combinations of the four postures. Generally, in the case of a large number of samples (1000 in our article), we assume that the feature of the two postures' samples comes from two normal distribution models. The statistic T_s in equation (9) for the t distribution is adopted for hypothesis testing at a $(1 - \alpha)$ confidence level

$$T_s = \frac{m_1 - m_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (9)$$

where m_1 and m_2 are the means of the features of two posture samples, S_1^2 and S_2^2 are the variances, and n_1 and n_2 are the sample numbers.

Statistical hypothesis testing is conducted, and the null hypothesis H_O and alternative hypothesis H_A are described as follows:

H_O: There is no significant difference between the two postures with this feature.
H_A: There is a significant difference between the two postures with this feature.

For only one instance, the feature mean d_m is given to verify the ability to discriminate the standing posture and the crouching posture. According to equation (9) $T_s = 43.64$, we set $\alpha = 0.01$, and then $t_{\alpha/2}(n_1 + n_2 - 2) = t_{0.005}(1998) = 2.58$. H_O should be

rejected at the $\alpha = 0.01$ significance level since T_s is larger than $t_{0.005}$. In other words, the alternative hypothesis H_A will be accepted, and that means that the feature d_m offers significant discrimination between the standing posture and the crouching posture. The larger the value of the statistic T_s is, the more significance the difference for posture discrimination the feature has. The statistic T_s can be treated as an indicator of the feature's discrimination level between different postures.

Table 1 gives the statistics T_s of the mean feature d_m, standard variance feature d_v, skewness feature d_s, and kurtosis feature d_k for different combinations of the four postures. From Table 1, we can see that feature d_m is effective for discrimination of all the postures. It has the best effect for discriminating lying postures from crouching postures but the lowest effect for differentiating standing postures and lying postures. The feature d_v has characteristics similar to those of the feature d_m, but it is much more useful than the feature d_m for discriminating standing postures from crouching postures and crouching postures from sitting postures. The feature d_s shows the best results for distinguishing lying postures from sitting postures. It is less efficient for discriminating crouching postures from sitting postures and is completely useless for differentiating standing postures from crouching postures. The feature d_k has the same performance for discriminating standing postures from crouching postures and crouching postures from lying postures. It is not so efficient for discriminating crouching postures from sitting postures, and it is not useful in differentiating standing postures from lying postures.

The NDH $\hat{D}$ mainly focuses on the spatial distribution of an entire human posture (global features) within the located ellipse, that is, different postures have different values in the bins of NDH $\hat{D}$. The eight components of $\hat{D}$ should be used together. However, the four statistical moments of $\hat{D}$ have different emphasis on the measurement of the NDH, which reflects the local characteristics of the human posture (local features). They can be used independently or combined with each other. These 12 elements are static features because they have no relation to human motion. It takes a certain amount of time to execute a movement, such as walking, falling, crouching, and sitting. Some postures may have similar spatial distributions in the ellipse of interest, but the change rate is quite different during this period. For example, there are two groups of high bars in the NDH $\hat{D}$ for walking and lying. The phase differences between the two groups of high bars are approximately 180°. The four statistical features of the two postures have the lowest significance level for classification (in Table 1). The phase of the first group of high bars of every walking or standing posture is approximately 90° and will not fluctuate significantly. The phases of the first group of high bars of lying
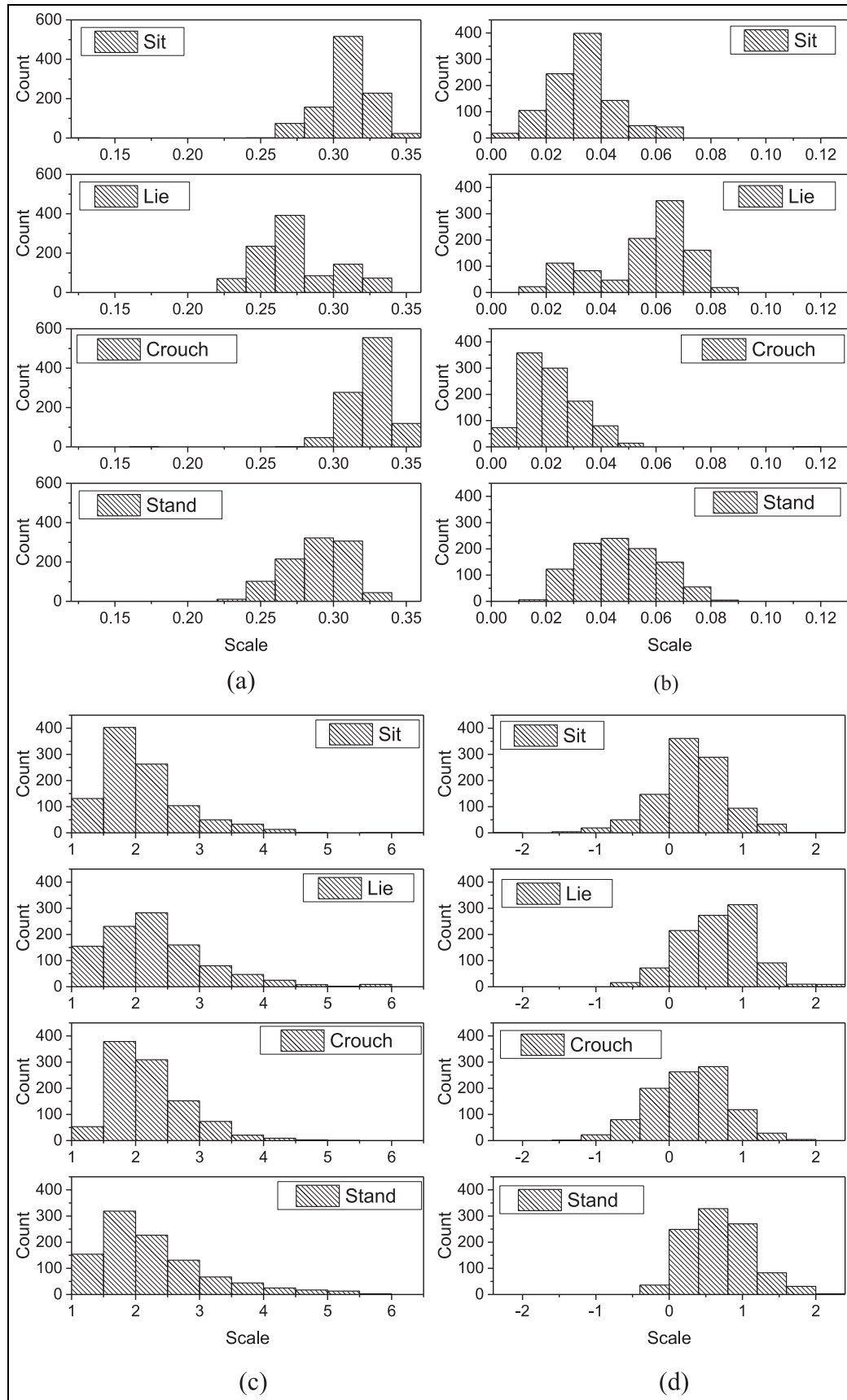
Figure 7. Histograms of four statistical features: (a) mean, (b) standard variance, (c) skewness, and (d) kurtosis.

Table 1. Significance test.

Posture	d_m	d_v	d_s	d_k
Stand/crouch	43.64	45.03	1.52	19.01
Stand/lie	13.59	13.14	1.48	0.70
Stand/sit	21.92	22.29	4.88	18.27
Crouch/lie	53.44	54.39	3.35	17.52
Crouch/sit	22.69	23.48	4.26	1.23
Lie/sit	33.99	33.86	6.77	16.74

postures will increase or decline from 90°. If the first frame of a fall is shown in Figure 6(a) and the last frame is shown in Figure 6(b), there are at least seven components of NDH $\hat{D}$ that changed drastically by more than 50%. Figure 8 shows the variations of the four statistical features of different postures in short video sequences.

Due to these observations, eight change rates of the components of NDH $\hat{D}$ are added as the dynamic features to improve the performance. In our implementation, the classification feature vector of a human posture can be obtained by fusing the 12 static features (eight global features and four local features) and 8 dynamic features

$$F = [d_1, d_2, \ldots, d_8, d_m, d_v, d_s, d_k, \dot{d}_1, \dot{d}_2, \ldots, \dot{d}_8] \quad (10)$$

where $\dot{d}_i$ is the change rate of d_i, and $\dot{d}_i = d_i(t) - d(t-1), i = 1, \ldots, 8$.

Fall event detection

At the beginning of this section, we discuss the posture characteristics of fall incidents, and then a tree structure
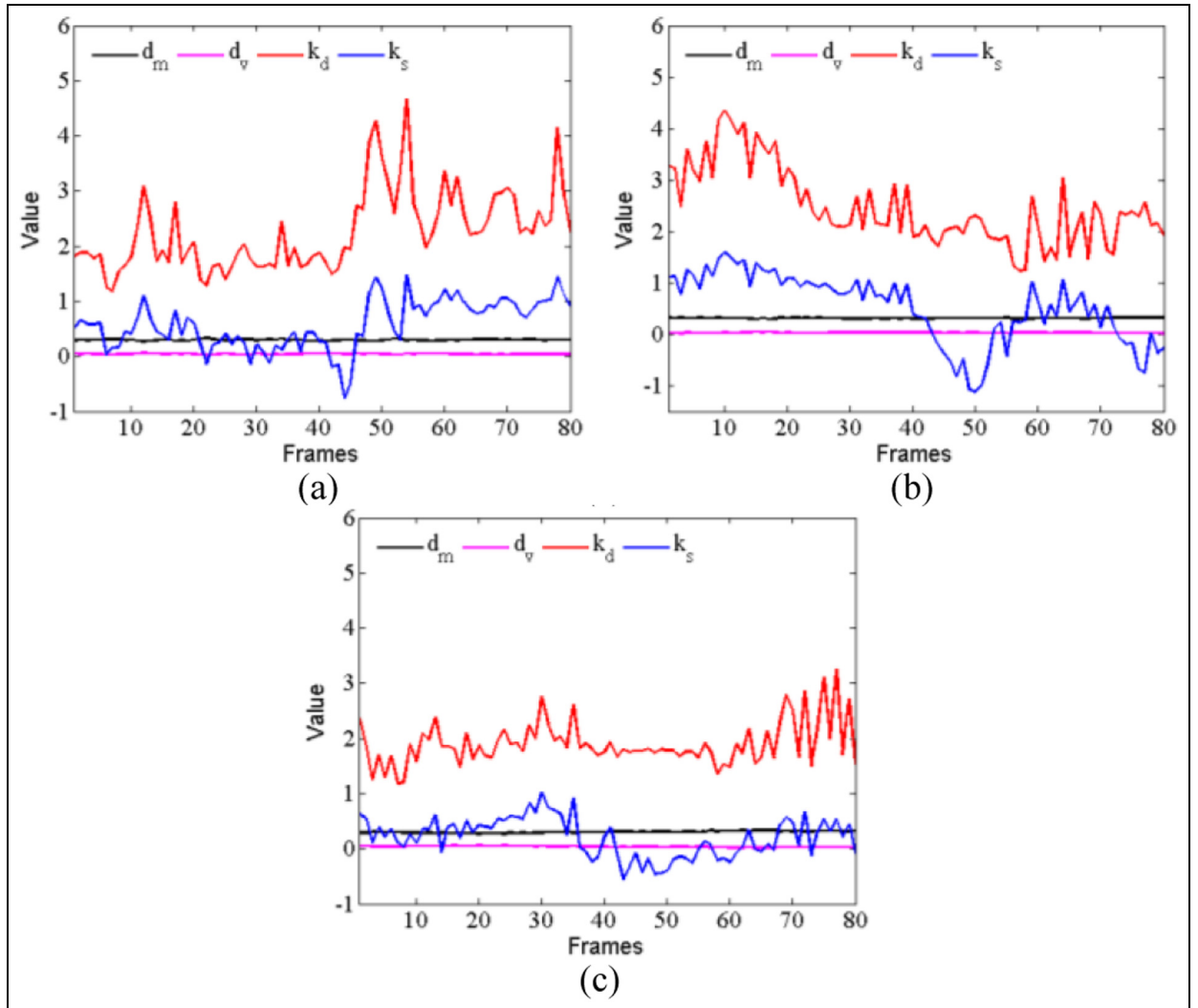
Figure 8. Variations of the four statistical features for video sequences: (a) standing to lying, (b) standing to sitting, and (c) standing to crouching.

classifier for separating lying postures from fall events is elaborated in detail.

The related postures of fall incidents

There are a variety of postures that occur during a fall incident. It is similar to a walking or standing posture in the pre-fall phase of a fall incident. In the following step, the fall speed gradually increases, and different body parts (such as the head and feet) have different velocities. Meanwhile, the body shape also has some distortions. During this period, the NDH $\hat{D}$ changes continually. There are three main reasons for this. First, there is an angle between the fall direction of the head and the ground when a fall incident occurs. Some of the components $\dot{d}_1, \dot{d}_2, \ldots, \dot{d}_8$ will increase, and some of them will definitely decline. Second, the dispersion degree, fluctuations, degree of skewness, and position of the highest peak of NDH $\hat{D}$ will also change along with the variations of NDH $\hat{D}$. Third, the velocities of different body parts are another important factor in a fall incident. This can be reflected by the change rates $\dot{d}_1, \dot{d}_2, \ldots, \dot{d}_8$ of NDH $\hat{D}$. In summary, the feature vector can describe, from different perspectives, the spatial and dynamic characteristics of the human postures that occur in various stages of a fall incident.

Moreover, considering the characteristics of a fall incident that are similar to those of crouching and sitting motions, we can observe several such traits, for example, the speed of the head is much faster than that of the feet, and the body shape has some distortions. However, what is peculiar about a fall incident is that the direction of the head's motion is entirely different from the directions of sitting and crouching movements, and the distortion of body shape is much smaller. To increase the robustness and reduce the error rate of our fall detection system, we also introduce crouching and sitting postures into the experiment.

Posture recognition by DAGSVM

The fall-related postures are represented by a feature vector F, which can be attributed to one of the four categories (standing, crouching, lying, and sitting) by a classifier. SVM is a binary classifier that constructs an optimal hyperplane according to the minimum structural risk principle[31,33] in the feature space or a transformed high-dimensional feature space, which can be used for classification, regression, or other tasks. Naturally, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (the so-called functional margin) since, in general, the larger the margin, the lower the generalization error of the classifier. It was originally designed for binary classification, so how to make it suitable for solving multi-class problems is still

an open issue. Currently, there are two types of strategies for multi-class SVM. One is the construction of several binary classifiers and the combination of the classifiers by various rules, such as one-versus-all and one-versus-one strategies. The other is direct consideration of all of the multi-class data in one optimization formulation. However, because this method solves the multi-class SVM in only one step and the parameters of the SVM are too complicated to solve, a much larger optimization is required.

To take full advantage of the binary classifier SVM, the directed acyclic graph scheme is employed to combine several SVMs to solve multi-class classification problems. The adopted DAGSVM provides an excellent balance between generalization error and time efficiency at both training and computing time compared with other multi-class SVM schemes.[34,35] Figure 9 shows the structure of the DAGSVM designed according to the related information of NDH $\hat{D}$ for solving the four-posture (standing, crouching, lying, and sitting) classification problem. It has a tree-like structure, and each node in this structure corresponds to a simple binary SVM. The overlap ratios of the four statistical features of the histograms in Figure 7 are given in Table 2. The higher the overlap ratio, the smaller the difference in characteristics of the two postures under consideration. It also means that the feature has lower effect for discriminating the two postures. If the overlap ratio reaches a certain threshold, we argue that this feature does not have the ability to differentiate the two postures. The overlap ratios of the four statistical features in the histograms provide the basis for coarse-to-fine classification of human body postures. First, in Table 2, two-thirds of the maximal values in d_m, d_v, and d_k indicate the posture "Crouch/Sit." The two second-largest values in d_m and d_v show that the posture belongs to the "Stand/Lie" class. For this reason, the four postures should first be classified into two categories in the root node, that is, "Stand Lie" and "Crouch Sit." According to the overlap ratios of the four statistical features in Table 2, the standing and lying postures are considered as one category, and the crouching and sitting postures are the other category in the coarse classification step. The classification accuracy should not be too low or too high in this step. All 20 features of F are used to train the SVM1 in the root node I-0. Second, according to the hypothesis test results in Table 1, in the first step of the fine classification, feature d_k is not utilized to train the classifier SVM2 in node I-1, and features d_s and d_k are not taken to train SVM3 in node I-2. Not all of the features were used to train classifiers in nodes I-1 and I-2. Third, in the second fine classification step, three-fourths of the values in $\min\{d_m(\overline{Crouch}/Sit)\}$, $\min\{d_v(\overline{Crouch}/Sit)\}$, $\min\{d_s(\overline{Crouch}/Sit)\}$, and $\min\{d_k(\overline{Crouch}/Sit)\}$ in Table 2 illustrate the posture "Lie/Sit." The classifier
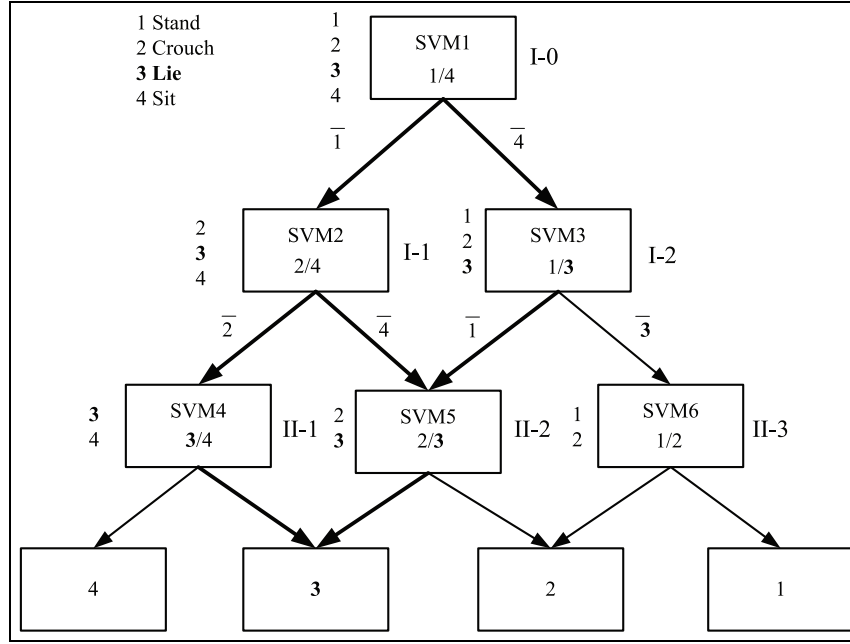
Figure 9. Classification structure of DAGSVM.

Table 2. Overlap ratio of the histogram.

| Posture	d_m	d_v	d_s	d_k
Stand/crouch	32.20%	38.60%	NULL	78.40%
Stand/lie	61.00%	61.50%	NULL	NULL
Stand/sit	35.10%	44.60%	81.40%	55.00%
Crouch/lie	22.50%	24.60%	88.60%	92.70%
Crouch/sit	75.10%	74.20%	85.40%	NULL
Lie/sit	25.20%	29.50%	78.90%	74.80%

SVM4 in node II-1 is employed to classify the posture "Lie/Sit" on the basis of the results obtained in node I-1. Two-thirds of the values in $\min\{d_m(\overline{Sit}/Crouch)\}$, $\min\{d_v(\overline{Sit}/Crouch)\}$, and $\min\{d_k(\overline{Sit}/Crouch)\}$ indicate the posture "Crouch/Lie." The classifier SVM5 in node II-2 is applied to classify the posture "Crouch/Lie." Two-thirds of the values in $\min\{d_m(\overline{Stand}/Lie)\}$, $\min\{d_v(\overline{Stand}/Lie)\}$, and $\min\{d_k(\overline{Stand}/Lie)\}$ in Table 2 indicate the posture "Crouch/Lie." Thus, the classified lying postures in node I-2 are introduced to node II-2. Two-thirds of the values in $\min\{d_m(\overline{Lie}/Stand)\}$, $\min\{d_v(\overline{Lie}/Stand)\}$, and $\min\{d_k(\overline{Lie}/Stand)\}$ indicate the posture "Stand/Crouch." The classifier SVM6 in node II-3 is employed to classify the posture "Stand/Crouch." According to Table 1, all the features are used to train SVM4 and SVM5 in node II-1 and node II-2. However, the feature d_s is not employed to train SVM6 in node II-3. The decision process just follows the structure, based on a sequence of binary classification operations, from the root node to the leaf node. A final decision is made when the bottom node is reached.

Fall incident detection by a majority voting strategy

A fall incident usually occurs in a short period of time. The typical duration of an event is approximately 0.4–0.8 s. That is approximately 10–20 frames. According to the characteristics of fall events, in most of the cases, a fall accident starts with a standing posture and ends with a lying posture. Meanwhile, there are a variety of other similar postures between the start and end postures. At the end of the fall accident, lying postures usually remain unchanged for some time because the fallen elderly person is lying immobile. As explained in the previous section, all the frames are classified into four different types of postures. Once the postures' classes have been found, we filter this per-frame posture solution of the fall event detection by counting the occurrences of lying postures within a short temporal window. For this purpose, the majority voting strategy is adopted as shown in Figure 10.

Having considered the typical duration of a fall event, we decided to set $T = 15$ for the temporal window as described in section "NDH of human
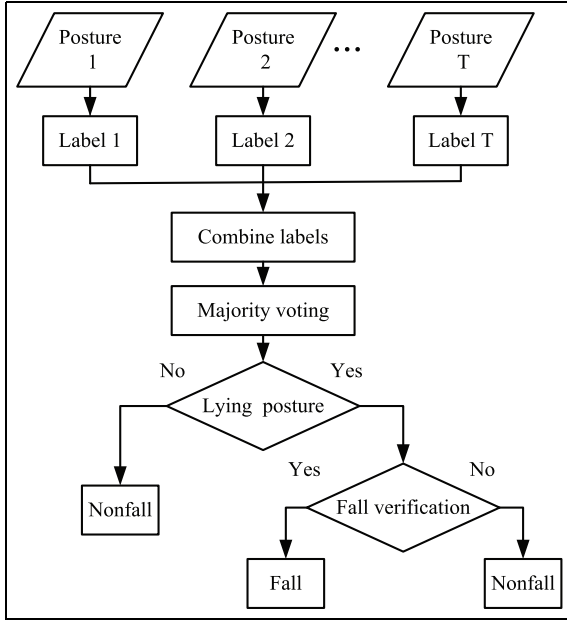
Figure 10. An overview of the voting strategy for fall detection.

postures." The classified postures in the temporal window are $P = \{p_1, p_2, \ldots, p_T\}$, where $p_j \in L = \{1, 2, 3, 4\}, j = 1, 2, \ldots, T$, and the numbers in the label set L represent the postures: standing, crouching, lying, and sitting, respectively. The strategy implemented here for filtering these 15 frames is simply a majority voting of the postures that is conducted within the temporal window. Each posture has one vote; if more than half of these postures are considered to contain falls, then the event as a whole may be regarded as a fall incident. The voting scheme can be formulated as in equation (11)

$$I_T = 1_{\left\{ \#\{p_j = = 3\} \geq \#\{p_j\}/2 \right\}} \tag{11}$$

A fall-like event can be determined by considering whether the indictor $I_T = 1$ or not. However, even if a fall-like event is detected, that is, $I_T = 1$, it does not mean a real fall incident has transpired. Other similar actions, such as lying to standing, lying to sitting, may also trigger this indicator. Therefore, the last verification of a fall accident is very necessary. It is accomplished by checking if the person is lying immobile on the floor. After a fall is detected, we carefully observe the ellipse of the human body for the next few seconds. If an unmoving ellipse is detected, we confirm the fall. An alarm is triggered in this case. If the ellipse still continues to move during the period, this cannot be a fall.

Experimental results

In this section, we present the performance of our fall detection system. The experiments are carried out on a desktop with an Intel(R) Core(TM) i3-2120 CPU and 2.00 GB of RAM. We test it intensively on a public multiple-camera fall dataset.[32]

This video dataset contains simulated falls and normal daily activities recorded in 24 realistic situations. In each scenario, an actor engages in many activities, such as falling, sitting on a sofa, walking. Every scenario is shot simultaneously with eight different cameras mounted around the room where the fall incident happens. All of the actions are performed by the same person with differently colored garments. Some of them include partial occlusion. We used all of the 736 samples of the dataset, which are shot from different viewpoints. Figure 11 illustrates the eight directions of a fallen person with its NDH. It can be seen as eight different lying postures of one person. Figure 6 shows the extracted human body and the NDH of the four typical postures. We should choose samples to train the classifier. Therefore, the dataset is split into two parts. The first part contains 80% of the samples (589). It was used for training, and the other samples were used for testing (148). The commonly used cross-validation technique was applied to evaluate the performance of the classification system. Three different types of evaluations are made. Then, we compare our system with other related fall detection methods using the same data.

Posture classification results

Based on this dataset, we first apply the classification accuracy to evaluate the performance of the DAGSVM in fulfilling the task of posture classification. The first assessment is associated with different types of features used for human posture classification, which is summarized in Table 3. The classification results are compared by regarding the values of the global features, local features, and dynamic features alone and using a combination of these three types of features. The DAGSVM is employed for this purpose. From the table, we can see that the classification result is promising only regarding the global features (93.6%). It is much higher than the performance obtained within the local features (60.4%) and dynamic features (45.8%). However, the accuracy achieved with the combined features improves the overall classification rates significantly. According to the classification rate, a combined group of three types of features achieves the best performance, with a much better classification rate (97.1%) than that of one-type or two-type feature groups.

The second estimation is of different feature extraction methods for human posture classification. The comparison of the proposed method and some other existing approaches conducted on the same data is shown in Table 4. We see that the accuracy values achieved by our method are higher than those of the
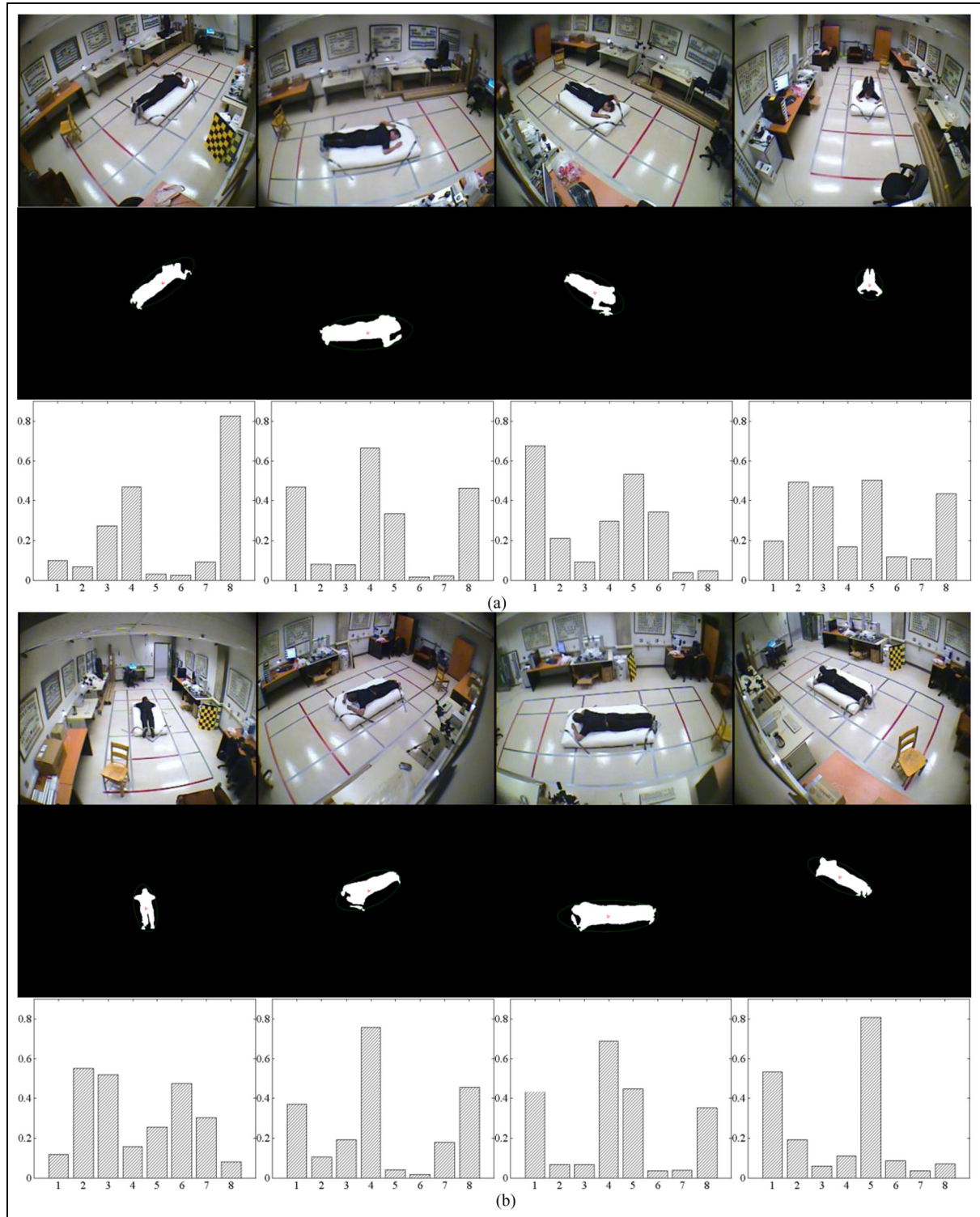
Figure 11. Eight postures of a fallen person: (a) directions 1–4 and (b) directions 5–8.

three other compared techniques. Our approach yields more favorable results compared to those achieved by the complicated projection histograms.[21] The projection histograms are obtained by analyzing the projection on the major and minor axes of the ellipse. They only provide static information without the motion information of the human posture. The method of extracting features by analyzing the variations of the silhouette area[23] just utilizes the dynamic information of a person's movements. The variations of the

Table 3. Comparison of different types of features.

Features types	Accuracy (%)
Global	93.6
Local	60.4
Dynamic	45.8
Global and local	96.9
Global and dynamic	95.8
Global, local, and dynamic	97.1

Table 5. Comparison of three classifiers for the four postures.

Classifiers	Accuracy (%)
ELM	74.7
MLPNN	92.9
DAGSVM	97.1

ELM: extreme learning machine; MLPNN: multilayer perceptions neural network; DAGSVM: directed acyclic graphic support vector machine.

Table 4. Comparison of different feature extraction methods.

Method	Accuracy (%)
Projection histogram[21]	96.1
Silhouette area[23]	95.2
Bounding box ratio[18]	82.2
Our method	97.1

silhouette area are view-invariant. The approach is widely dependent on the background update strategy. The bounding box ratio method[18] is easy to implement only when the camera's view direction is perpendicular to the moving person.

In the third assessment, the performance of DAGSVM in distinguishing the falling posture from the other three postures is compared with the performances of two state-of-the-art classifiers: the ELM and multilayer perceptions neural network (MLPNN)[36]. The results of the comparison are given in Table 5. To conduct a fair and representative comparison, the parameters of the compared classifier are tuned to be optimal using a tenfold cross-validation technique. DAGSVM achieves the best performance with a much better classification rate (97.1%) than that of MLPNN (only 92.9%) and that of ELM (only 74.7%). The overall classification rate of DAGSVM is 4.2% higher than the accuracy of MLPNN. The receiver operating characteristic curves of the three algorithms are shown in Figure 12. It can be found that the proposed DAGSVM outperforms both of the compared classifiers.

It is not a reliable metric to estimate the performance of the classifiers only using classification accuracy for the four-posture classification since classification accuracy is mainly concerned with the percentage of correctly classified instances out of the total number of samples and neglects the rate of misclassified samples. The classification rate cannot accurately reflect the performance of the multi-class classifiers in the case of one class accuracy being very high while the other is very low. To adequately evaluate the effectiveness of classifiers, the confusion matrix, also known as the error matrix, is the best choice. A confusion matrix is a
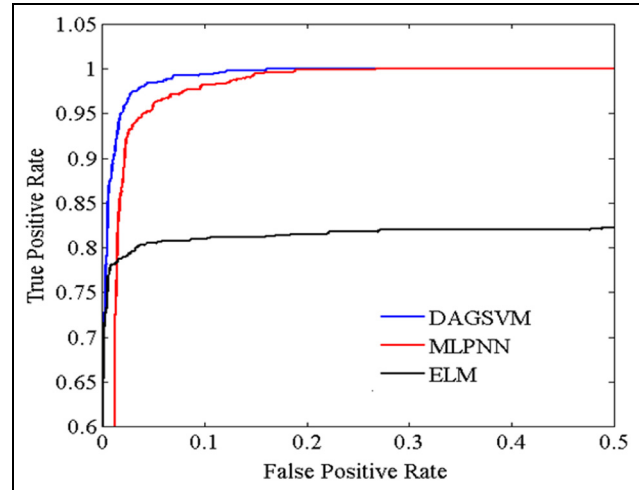

Figure 12. The receiver operating characteristic curves of the three classifiers.

specific table that allows for visualization of the performance of a classifier on a set of test data for which true values are already known. Each column of the matrix represents the instances (postures) in a predicted class, while each row represents the instances in an actual class (or vice versa). All correct predicted instances are located in the diagonal of the table, so it is easy to inspect the table visually and find the errors, which are represented by the values outside the diagonal. It is easy to see whether the classifier confuses two classes (i.e. perceives the sample of one class as belonging to the other). Figure 13 shows the confusion matrixes of the classifiers conducted on the dataset. The differences among the diagonal entries of ELM are very significant, especially that between the first posture (standing) and the fourth posture (sitting). The accuracy of distinguishing standing postures can exceed 95.0%, while the classification rate of sitting postures is only 58.0%. Almost 23.0% of sitting postures are misclassified as crouching postures. The confusion matrix indicates that the ELM algorithm has some trouble distinguishing both falling postures and crouching postures, but it can make the distinction between standing postures and other types of postures pretty well. There are two

reasons for this poor result of ELM. One reason concerns the feature distribution. The input features are suitable to distinguish standing postures. This reflects that the feature's distribution of standing postures is obviously different from that of the other three postures (crouching, lying, and sitting). Furthermore, the feature's distributions of crouching, lying, and sitting are very similar, especially lying and sitting. Thus, the accuracies of the two postures are not so satisfying. The other reason is the ELM classifier. There is only one single hidden layer in the extreme learning machine, which can simply be considered a linear system. The classification performance of ELM is highly dependent on the number of hidden nodes, although it has been proven that N hidden nodes can correctly learn N distinct observations. If the number of hidden nodes is equal to the number of distinct training samples, ELM can approximate these samples with zero error. In fact, in most cases, the number of hidden nodes is much less than the number of distinct training samples. However, considering our small-sized classification task, we only choose 24 hidden nodes, which is equal to the number of features. The four-posture classification rate of MLPNN is almost the same. However, it is not suitable for lying posture recognition since the accuracy is only 84.0%. There is still room for improvement in the accuracy of falling posture discrimination. DAGSVM has almost the same performance for the four postures. The discrimination accuracy of all posture types exceeds 95.0% and is higher than that of the compared algorithms. From the confusion matrix in Figure 13(c), we can see that all standing and crouching postures are not misclassified as lying postures. Only the sitting postures are misclassified. There are only 1.0% of postures misclassified as lying postures. The lying posture classification accuracy of our approach is approximately 96.0%. It is very beneficial for fall-like motion recognition. The rate of lying postures that are not correctly detected is acceptable at only 4% of the total number.

Fall detection results

Posture classification constitutes a critical step in our fall detection system. The labels of the postures in the temporal window are used to detect falls according to the majority vote strategy explained in the last section. If more than half of the labels in the temporal window indicate lying postures, it testifies that a fall-like incident has occurred. After the incident has occurred, we observe the characteristics of the fallen person, that is, the located minimum area-enclosing ellipse of the fallen person. If the person has been lying immobile on the ground for a while, the fall event is verified. To illustrate this, we show eight different cases of a fall incident in Figure 11. The monitored person falls on the floor, and a lying posture is detected on the ground. This lying
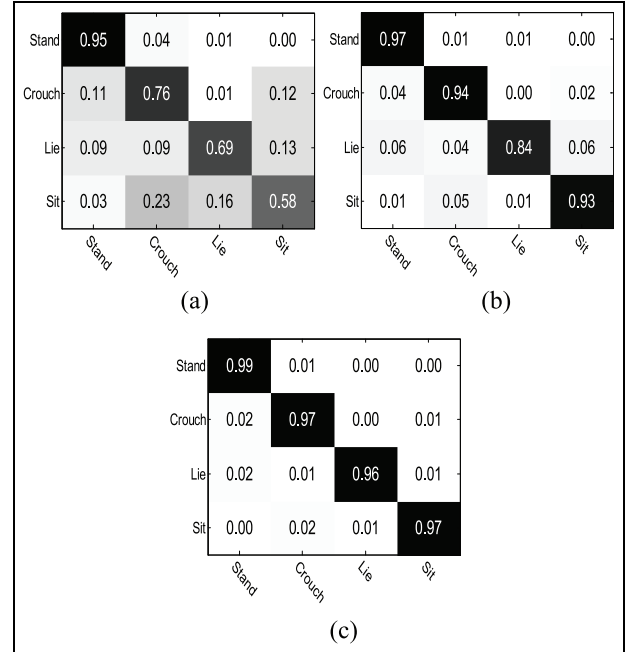
Figure 13. Confusion matrixes of the three classifiers: (a) extreme learning machine, (b) multilayer perceptions neural network, and (c) the directed acyclic graph support vector machine.

Table 6. Evaluation of DAGSVM for fall posture recognition.

Actions	Count	Detected as falls	Detected as non-falls
Walk	38	0	38
Crouch	32	0	32
Fall	42	40	2
Sit	34	1	33

posture is maintained for a while, so a fall event is verified. The system issues an alarm to call for the assistant.

To evaluate our fall detection system, we used this free-fall detection dataset as testing samples. The samples used in this subsection contain only one subject in each video sequence. A total of 148 instances were recorded, including 42 fall incidents and 106 non-fall activities. Table 6 depicts the samples and the detection results. The length of the temporal window T is considered in our approach. It may have significant effects on the overall performance. To investigate the effect of this parameter, we perform fall detection experiments on the public dataset with different parameter settings. Figure 14 shows the accuracies of various parameters of the temporal window T. It can be found that the accuracy of our system increases with the increase in length of the temporal window until $T = 15$. After that, the performance of our approach slightly degrades with the
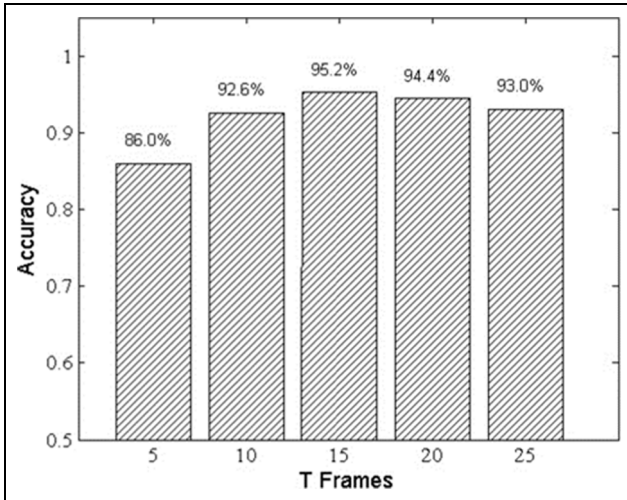
Figure 14. Accuracies of different temporal windows.

Table 7. Comparison of different fall detection methods.

Method	Accuracy (%)
Shape variation[24]	90.5
Projection histogram[21]	95.1
Our method	95.2

increase in length of the temporal window. Obviously, there are too few postures in the short temporal window; even a few misclassified postures may lead to false detection. The system is very sensitive in this case. However, an overly long temporal window also introduces other redundant postures, and it has to contain many more lying postures to trigger the indicator I_T. If the number of detected lying postures of a fall incident cannot satisfy the detection condition, the fall will not be detected. The specificity of this parameter setting is slightly higher. For a compromise between sensitivity and specificity, we choose $T = 15$ for the temporal window.

To demonstrate the effectiveness of our system, two state-of-the-art methods, namely shape variation[24] and projection histogram,[21] were implemented, and their results were compared with those of the proposed methods. To assess the performance of the three methods, the fall detection accuracy was calculated for each one. From Table 7, we can note that the proposed system of fall detection methods based on the posture classification outperforms the other two approaches. The fall detection accuracy of our method can exceed 95.2%, while the accuracies of the two compared methods were only 90.5% and 95.1%, respectively. This implies that the proposed method is more effective than the other two methods.

Discussion

The proposed fall detection system is based on the strategy of detecting the required posture through majority voting, compared to the traditional posture classification system, it has the following merits:

(1) It is conveniently not as greatly affected by background noise in the environment as the methods based on acoustic sensors[12] and floor vibration sensors.[7] A reference background image was first constructed by our background model. Model parameters were used to classify the current image pixels into three categories: the background mask, moving object mask, and shadow mask. The thick noise around the object in the foreground image was removed by morphological operations.

(2) Even if the training dataset is much larger and contains more types of actions of daily life, the well-trained DAGSVM tree structure classifier can still distinguish many types of postures that are used for fall detection. The classified postures are not only used for fall detection but can also be employed for the detection of particular types of fall events by analyzing the labels of postures in the filtered temporal window at the stage of majority voting, for example, slip-backward, trip-forward, and left/right lateral types. In prosthetics research, the adjustment of parts of a knee joint is often a tradeoff because of the different types of falls that may occur. Different types of falls can lead to different types of injuries. For the elderly, a slip may be likely to result in an injury more serious than a trip. It is more likely a lateral fall when the person loses his or her consciousness. Knowing the type of fall incident might be necessary for the fall assistant to choose the more appropriate response. It is also valuable for adjusting the treatment of monitored seniors and for preventing similar types of falls in the future.

(3) Our posture-based approach is implemented at the frame level; there is no need to segment a video clip. Compared with the video clip-based method,[26,27] our approach is computationally more efficient since it extracts classification features only from the foreground rather than from the entire video clip.

The proposed fall detection system yields favorable results compared to those reported in the literature. However, some problems still occur in our system. One core problem is that it is designed to monitor one person. In some special cases, such as the presence of many

people or a large-sized pet along with the elderly in the monitored environment, it does not cope well with the situation. In this case, other object classification techniques should be applied to determine whether the extracted silhouette is a human body or that of a pet and whether this silhouette belongs to the monitored person.

Another problem is the occlusion of the living room of elderly people. The real room environment is sometimes much cluttered which causes occlusions and hampers the extraction of features. A severe occlusion will degrade the final performance of the fall detection system. Before the fall detection phase, the ground is determined initially. It may be helpful for partly solving the occlusion problem. However, if we use only one camera, it is impossible to address this problem completely. More than one camera can be used to make sure that the human body is visible in at least one camera's view. We can find that the view of the occluded part of the human body may be obtained from the other installed cameras, making it much easier to accomplish the task of extraction of human body features. For instance, Auvinet[20] reconstructed the three-dimensional shape of people by multiple camera networks. Each three-dimensional position was independently reconstructed. Hence, the algorithm was able to address several occlusions.

As for the possibility of the background changing drastically, for example, a light being turned on or the sun suddenly illuminating the room, the background subtraction technique is very sensitive to such situations. This is another shortcoming of the posture-based approaches. To suit the varied environment, the system has to undergo a background update procedure and will recover to the normal performance after the procedure is finished. If a fall incident occurs during the period of the upgrade process, most of the fall detection systems based on the RGB video camera fail to detect it. This is an open issue with ongoing research. Fusing multiple sensors in a proper manner may serve as a satisfactory solution to this extreme situation.

Last but not least, although the risk for falls increases dramatically with age, falls are not an inevitable result of aging. There is evidence that falls of the elderly can be prevented with clear guidance for effective interventions. Practically, few old people are offered interventions to prevent falls. General practitioners (GPs) are usually relied on to manage the needs of our system. These GPs are invited to simulate types of fall prevention activities in the environment. Our fall detection system detects trials of these activities. Then, clinical guidelines, such as environmental modification, medication review, and exercise/rehabilitation, are provided according to these results. Furthermore, the development of care plans in fall prevention has been determined based on the resident characteristics. Through practical lifestyle adjustments, the number of falls among seniors can be substantially reduced.

Conclusion

In this article, we have proposed a novel vision–based fall detection approach characterized by the utilization of human postures. There are three main elements in our approach. First, the human body that was extracted from frames by the background model technique is located using a minimum area-enclosing ellipse. This ellipse represents an excellent tool for solving the problem of locating a separate human object. Second, human postures are described by the NDH constructed around the center of the ellipse. The method not only avoids the deviations of posture description caused by the fragmentariness of the extracted object but also eliminates the effects produced by different sizes of the same posture. We obtained classification features by pooling the spatial and dynamic information got from the NDH of a human silhouette. The 12 static features mainly reflect the spatial information, while the eight motion features express the dynamic information. The analysis of four statistical features out of the 20 lays a good foundation for feature unitization and multi-level classification of human postures. Third, we propose DAGSVM, which utilizes the directed acyclic graph strategy to combine several binary classification SVMs to sort human postures. It classifies the postures first into two categories (stand–lie, crouch–sit) and then into four categories, namely standing, crouching, sitting, and lying. An overall accuracy of 97.1% is obtained for the four-posture recognition. These classified postures are then filtered frame by frame in a short temporal window. A fall-like accident is detected by counting the occurrences of lying postures. After conducting the majority voting, the fall event is determined by immobility verification. From the experimental results, our fall detection system achieves up to 95.2% accuracy on a public fall dataset. It is worth noting that the accuracy of our system is greater than that of two other compared methods.

Despite all the advantages listed above, this study is still very limited since we did not use the records of real fall accidents for our research. In real settings, falls are relatively rare events compared to movements such as sitting down, crouching down, and walking. The occlusion and cluttered background problems still need to be solved in our system. In the future, a probable solution to these problems may be found using multiple types of sensors.

Declaration of conflicting interests

References

1. World Health Organization (WHO) and Ageing and Life Course Unit. WHO global report on falls prevention in older age. Report, WHO, Geneva, June 2008.
2. Chaudhuri S, Thompson H and Demiris G. Fall detection devices and their use with older adults: a systematic review. *J Geriatr Phys Ther* 2014; 37: 178–196.
3. Igual R, Medrano C and Plaza I. Challenges, issues and trends in fall detection systems. *Biomed Eng Online* 2013; 12: 1–66.
4. Mubashir M, Shao L and Seed L. A survey on fall detection: principles and approaches. *Neurocomputing* 2013; 100: 144–152.
5. Zhang Z, Conly C and Athitsos V. A survey on vision-based fall detection. In: *Proceedings of the ACM conference pervasive technologies related to assistive environments*, Corfu, 1–3 July 2015, pp.1–7. New York: ACM.
6. Koshmak G, Loutfi A and Linden M. Challenges and issues in multisensor fusion approach for fall detection: review paper. *J Sensor* 2016; 2016: 6931789-1–6931789-12.
7. Madarshahian R, Caicedo J and Arocha Zambrana D. Benchmark problem for human activity identification using floor vibrations. *Expert Syst Appl* 2016; 62: 263–272.
8. He J, Hu C and Wang X. A smart device enabled system for autonomous fall detection and alert. *Int J Distrib Sens N* 2016; 2016: 1–10.
9. Gao X, Chen Z, Tang S, et al. Adaptive weighted imbalance learning with application to abnormal activity recognition. *Neurocomputing* 2016; 173: 1927–1935.
10. Jung S, Hong S, Kim J, et al. Wearable fall detector using integrated sensors and energy devices. *Sci Rep* 2015; 5: 17081.
11. Gibson R, Amira A, Ramzan N, et al. Multiple comparator classifier framework for accelerometer-based fall detection and diagnostic. *Appl Soft Comput* 2016; 39: 94–103.
12. Salman Khan M, Yu M, Feng P, et al. An unsupervised acoustic fall detection system using source separation for sound interference suppression. *Signal Process* 2015; 110: 199–210.
13. Su B, Ho K, Rantz M, et al. Doppler radar fall activity detection using the wavelet transform. *IEEE T Biomed Eng* 2015; 62: 865–875.
14. Lopes I, Vaidya B and Rodrigues J. Towards an autonomous fall detection and alerting system on a mobile and pervasive environment. *Telecommun Syst* 2013; 52: 2299–2310.
15. Costa S, Rodrigues J, Silva B, et al. Integration of wearable solutions in aal environments with mobility support. *J Med Syst* 2015; 39: 1–8.
16. Yu X. Approaches and principles of fall detection for elderly and patient. In: *Proceedings of the IEEE conference e-health networking applications and services*, Singapore, 7–9 July 2008, pp.42–47. New York: IEEE.
17. Lee T and Mihailidis A. An intelligent emergency response system: preliminary development and testing of automated fall detection. *J Telemed Telecare* 2005; 11: 194–198.
18. Liu C, Lee C and Lin P. A fall detection system using *k*-nearest neighbor classifier. *Expert Syst Appl* 2010; 37: 7174–7181.
19. Rougier C, Meunier J, St-Arnaud A, et al. Robust video surveillance for fall detection based on human shape deformation. *IEEE T Circ Syst Vid* 2011; 21: 611–622.
20. Auvinet E, Multon F, Saint-Arnaud A, et al. Fall detection with multiple cameras: an occlusion-resistant method based on 3-D silhouette vertical distribution. *IEEE T Inf Technol B* 2011; 15: 290–300.
21. Yu M, Rhuma A, Naqvi S, et al. A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment. *IEEE T Inf Technol B* 2012; 16: 1274–1286.
22. Olivieri D, Gómez Conde I and Sobrino X. Eigenspace-based fall detection and activity recognition from motion templates and machine learning. *Expert Syst Appl* 2012; 39: 5935–5945.
23. Mirmahboub B, Samavi S, Karimi N, et al. Automatic monocular system for human fall detection based on variations in silhouette area. *IEEE T Biomed Eng* 2013; 60: 427–436.
24. Chua J, Chang Y and Lim W. A simple vision-based fall detection technique for indoor video surveillance. *Signal Image Video P* 2015; 9: 623–633.
25. Li A, Miao Z, Cen Y, et al. Histogram of maximal optical flow projection for abnormal events detection in crowded scenes. *Int J Distrib Sens N* 2015; 11: 406941.
26. Ma X, Wang H, Xue B, et al. Depth-based human fall detection via shape features and improved extreme learning machine. *IEEE J Biomed Health Inform* 2014; 18: 1915–1922.
27. Aslan M, Sengur A, Xiao Y, et al. Shape feature encoding via Fisher Vector for efficient fall detection in depth-videos. *Appl Soft Comput* 2015; 37: 1023–1028.
28. Stone E and Skubic M. Fall detection in homes of older adults using the Microsoft Kinect. *IEEE J Biomed Health Inform* 2015; 19: 290–301.
29. Horprasert T, Harwood D and Davis L. A statistical approach for real-time robust background subtraction and shadow detection. In: *Proceedings of the IEEE conference computer vision*, Kerkira, 20–27 September 1999, pp.1–19. New York: IEEE.
30. Kumar P and Yildirim E. Minimum-volume enclosing ellipsoids and core sets. *J Optimiz Theory App* 2005; 126: 1–21.
31. Bishop C. *Pattern recognition and machine learning*. Berlin: Springer, 2006, pp.100–156.
32. Auvinet E, Rougier C, Meunier J, et al. Multiple cameras fall dataset. Report, Université de Montréal, Montreal, QC, Canada, July 2010.
33. Vapnik V. *The nature of statistical learning theory*. New York: Springer, 1996, pp.267–290.

34. Platt JC, Cristianini N and Shawe-Taylor J. Large margin DAGs for multiclass classification. In: *Proceedings of the IEEE conference neural information processing systems*, Denver, CO, 29 November–4 December 1999, pp.547–553. New York: IEEE.

35. Hsu CW and Lin CJ. A comparison of methods for multiclass support vector machines. *IEEE T Neural Networ* 2002; 13: 415–425.

36. Chang CC and Lin CJ. LIBSVM: a library for support vector machines. *ACM Trans Intel Syst Tec* 2011; 2: 27.