CrossMark

Fall detection in dusky environment

Ying-Nong Chen[1], Chi-Hung Chuang[2*], Hsin-Min Lee[1], Chih-Chang Yu[3] and Kuo-Chin Fan[1]

Abstract

Accidental fall is the most prominent factor that causes the accidental death of elder people due to their slow body reaction. Automatic fall detection technology integrated in a health-care system can assist human monitoring the occurrence of fall, especially in dusky environments. In this paper, a novel fall detection system focusing mainly on dusky environments is proposed. In dusky environments, the silhouette images of human bodies extracted from conventional CCD cameras are usually imperfect due to the abrupt change of illumination. Thus, our work adopts a thermal imager to detect human bodies. The proposed approach adopts a coarse-to-fine strategy. Firstly, the downward optical flow features are extracted from the thermal images to identify fall-like actions in the coarse stage. The horizontal projection of motion history images (MHI) extracted from fall-like actions are then designed to verify the incident by the proposed nearest neighbor feature line embedding (NNFLE) in the fine stage. Experimental results demonstrate that the proposed method can distinguish the fall incidents with high accuracy even in dusky environments and overlapping situations.

Keywords: Fall detection, Optical flow, Motion history image, Nearest neighbor feature line

1 Introduction

Accidental fall is the most prominent factor that causes the accidental death of elder people due to their slow body reaction. Fall accidents usually occur at night with nobody except the elder people if they live alone. It is usually too late to remedy the tragedy when the body is discovered hours or days after with the occurrence of accidental fall. In the occurrence of fall incident, humans usually lie flat on the ground. However, we cannot merely use the images to perceive whether this person is lying on the ground. Hence, we have to detect and avoid the risk caused by fall action. According to the survey, a sudden fainting or body imbalance is the main reason to cause a fall. No matter what reasons, fall is a warning that the subject may be in danger. Moreover, the silhouette images of human bodies are hard to be extracted from conventional CCD cameras in dusky environments due to the illumination constraint. If the incidents occur in a dusky and unattended environment, people usually miss the prime time for rescue. To remedy this problem, a fall detection system using a thermal imager (see Fig. 1) to capture the images of human bodies is proposed in this paper. By using the thermal imager, the human

bodies can be accurately located even in a dusky environment. For comparison, Fig. 2a shows the images obtained by a CCD camera in a dusky environment, whereas Fig. 2b shows the images obtained by a thermal imager in the same environment. It is obvious that the thermal imagers can extract more clear and intact human bodies in the dusky environments than CCD cameras.

Moylan [1] illustrated the gravity of falls as a health risk with abundant statistics. Larson [2] described the importance of falls in elderly. The National Center for Health Statistics showed that more than one third of ages 65 or older fall each year. Moreover, 60 % of lethal falls occur at home, 30 % occur in public region, and 10 % happen in health-care institutions for ages 65 or older [3]. In the literatures of fall detection, Tao et al. [4] applied the aspect ratio of the foreground object to detect fall incidents. Their system firstly tracks the foreground objects and then analyzes the sequences of features for fall incident detection. Anderson et al. [5] also applied the aspect ratio of the silhouette to detect fall incidents. The rationale based mainly on the fact that the aspect ratio of the silhouette is usually very large when the fall incidents occur. On the contrary, the aspect ratio is much smaller when the fall incidents do not occur. Juang [6] proposed a neural fuzzy network method to classify the human body postures, such as

* Correspondence: chchuang@mail.fgu.edu.tw
[2]Department of Applied Informatics, Fo Guang University, Yilan, Taiwan
Full list of author information is available at the end of the article

Chen et al. EURASIP Journal on Image and Video Processing (2016) 2016:16

Page 2 of 14

Fig. 1 The thermal imager

standing, bending, sitting, and lying down. In [7], Foroughi et al. proposed a fall detection method using an approximated eclipse of human body silhouette and head pose as features for multi-class support vector machine (SVM). Rougier et al. [8] applied the motion history image (MHI) and variations of human body shape to detect falls. In [9], Foroughi et al. proposed a modified MHI integrating the time motion image (ITMI) as the motion feature. Then, the eigenspace technique was used for motion feature reduction and fed into individual neural network for each activity. Liu et al. [10] proposed a nearest neighbor classification method to classify the ratio of human body silhouette of fall incidents. In order to differentiate between the fall and lying, the time difference between fall and lying was used as a key feature. Liao et al. [11] proposed a slip and fall

detection system based on Bayesian Belief Network (BBN). They used the integrated spatiotemporal energy (ISTE) map to obtain the motion measure. Then, the BBN model of the causality of the slip and fall was constructed for fall prevention. Olivieri et al. [12] proposed a spatiotemporal motion feature to represent activities termed motion vector flow instance (MVFI) templates. Then, a canonical eigenspace technique was used for MVFI template reduction and template matching.

In this paper, a novel fall detection mechanism based on coarse-to-fine strategy which is workable in dusky environments is proposed. In the coarse stage, the downward optical flow features are extracted from the thermal images to identify fall-like actions. Then, the horizontal projected motion history image (MHI) features of fall-like actions are used in the fine stage to verify the fall by the nearest neighbor feature line embedding.

The contributions of this work are listed as follows: (1) using the thermal imager instead of CCD camera to capture intact human body silhouettes; (2) proposing a coarse-to-fine strategy to detect fall incidents; (3) proposing a nearest neighbor feature line embedding method for fall detection which improves the original nearest feature line embedding method; (4) proposing a scheme to detect fall incidents even though occlusion occurs.

The rest of this paper is organized as follows. In Section 2, the concept of nearest feature line embedding (NFLE) algorithm presented in our previous work [13] will be briefly reviewed. Then, the fall detection based on coarse-to-fine strategy and the nearest neighbor feature line embedding (NNFLE) algorithm are presented in Section 3. Experimental results are illustrated in Section 4 to demonstrate the soundness and effectiveness of the proposed fall detection method. Finally, conclusions are given in Section 5.

2 Nearest feature line embedding (NFLE)

The NFLE transformation is a linear transformation method based on a nearest feature space (NFS) strategy [13] originating from an nearest linear combination (NLC) methodology [14]. Since the points on the nearest feature line (NFL) are linearly interpolated or extrapolated from each pair of feature points, the performance is better than those of point-based methods. In addition, the NFL metric is embedded into the transformation through the discriminant analysis phase instead of in the matching phase.

Consider Nd-dimensional samples $X = [x_1, x_2, \ldots x_N]$ constituting N_c classes, the corresponding class label of x_i is denoted as $l_{x_i} \in \{1, 2, 3, \ldots N_c\}$ and a specified point $y_i = w^T x_i$ in the transformed space. The distance from point y_i to the feature line is defined as $\|y_i - f^{(2)}(y_i)\|$, in which $f^{(2)}$ is a function generated by two points, and
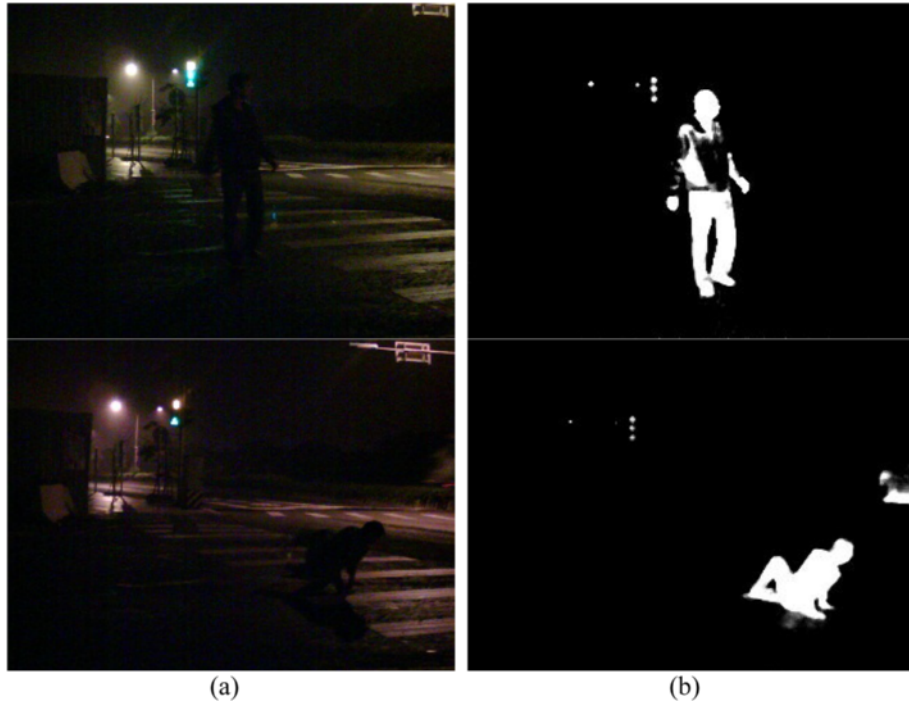
Chen *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:16

Page 3 of 14

Fig. 2 Image extraction results captured by (**a**) CCD camera and (**b**) thermal imager

$f^{(2)}(y_i)$ is the projected point of the line. A number of C_2^{N-1} possible lines for point y_i will be generated. The scatter computation of feature points to feature lines can be obtained and embedded in the discriminant analysis. In consequence, this approach is termed as NFLE.

In NFLE, the objective function in the equation

$$F = \sum_i \left(\sum \left\| \mathbf{y}_i - f^{(2)}(y_i) \right\|^2 w^{(2)}(y_i) \right) \tag{1}$$

is minimized.

The weight values $w^{(2)}(y_i)$ (being 1 or 0) constitute a connected relationship matrix of size $N \times C_2^{N-1}$ for N feature points to their corresponding projection points $f^{(2)}(y_i)$. Consider the distance $\left\| y_i - f_{m,n}^{(2)}(y_i) \right\|$ for point y_i to a feature line $L_{m,n}$ that passes through two points y_m and y_n; the projection point $f_{m,n}^{(2)}(y_i)$ can be represented as a linear combination of points y_m and y_n by $f_{m,n}^{(2)}(y_i) = y_m + t_{m,n}(y_n - y_m)$, in which $t_{m,n} = (y_i - y_m)^T(y_m - y_n)/(y_m - y_n)^T(y_m - y_n)$. The mean square distance for all training samples to their corresponding NFLs is minimized and its representation is given by the following lemma.

Lemma 2.1: The mean square distance from the training points to the NFLs can be represented in the form of a Laplacian matrix.

See Fig. 3 for illustration. For a specified point y_i, the vector from point y_i to the projection point $nf_{(m,n)}^{(2)}(y_i)$ of

the NFL $NL_{m,n}$ which passes through points y_m and y_n can be obtained as follows:

$$
\begin{aligned}
y_i - nf_{(m,n)}^{(2)}(y_i) &= y_i - y_m + t_{m,n}(y_m - y_n) \\
&= y_i - (1 - t_{m,n})y_m - t_{m,n}y_n \\
&= y_i - t_{n,m}y_m - t_{m,n}y_n \\
&= y_i - \sum_j M_{i,j}y_j
\end{aligned}
\tag{2}
$$

Here, $t_{n,m} = 1 - t_{m,n}$ and $i \neq m \neq n$. Two values in the ith row in matrix M are set as $M_{i,m} = t_{n,m}$ and $M_{i,n} = t_{m,n}$. The other values in the ith row are set as $M_{i,j} = 0$ if $j \neq m \neq n$.
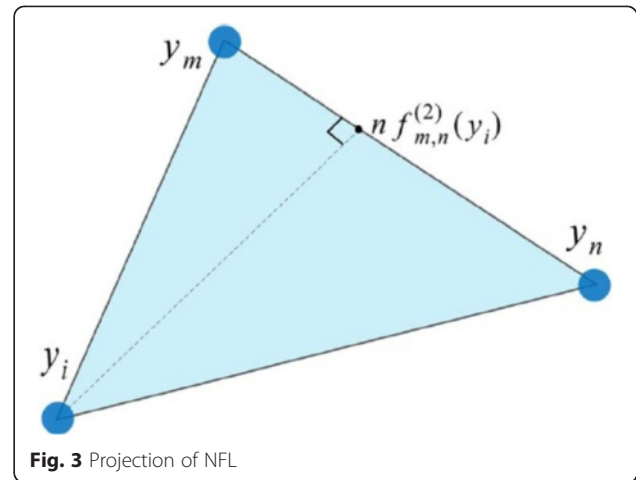

Fig. 3 Projection of NFL

In general, the mean square distance for all training points to their NFLs is obtained as follows:

$$\left(\sum\left\|\mathbf{y}_i - nf^{(2)}(y_i)\right\|^2\right) = \sum_i\left(\mathbf{y}_i - \sum_j \mathbf{M}_{i,j}\mathbf{y}_j\right)^2$$
$$= tr\left(Y^T(I-M)^T(I-M)Y\right)$$
$$= tr\left(Y^T(D-W)Y\right)$$
$$= tr\left(w^T XLX^T w\right)$$

(3)

in which $\sum_j M_{i,j} = 1$ and $L = D - W$. From the conclusions of [15], matrix W is defined as $W_{i,j} = (M + M^T - M^T M)_{i,j}$ when $i \neq j$ and zero otherwise. The function in (3) can thus be represented by a Laplacian matrix.

Moreover, when the K NFLs are chosen from C_2^{N-1} possible combinations, the objective function in (1) is also represented as a Laplacian matrix as stated in the following theorem.

Theorem 2.1: The objective function in (1) can be represented as a Laplacian matrix that preserves the locality among samples.

The objective function F in (1) is first decomposed into K components. Each component denotes the mean square distances for point y_i to the kth NFL. The first component matrix $M_{i,j}(1)$ denotes the connectivity relationship matrix between point x_i and the NFL $L_{m,n}$ for $i, m, n = 1, …, N$ and $i \neq m \neq n$. Two non-zero terms, $M_{i,n}(1) = t_{m,n}$ and $M_{i,m}(1) = t_{n,m}$, exist at each row of matrix $M_{i,j}(1)$ and satisfy $\sum_j M_{i,j}(1) = 1$. According to Lemma 2.1, it is represented as a Laplacian matrix $w^T XL(1)X^T w$. In general, $M_{i,n}(k) = t_{m,n}$ and $M_{i,m}(k) = t_{n,m}$ for $i \neq m \neq n$ if line $L_{m,n}$ is the kth NFL of point x_i and zero otherwise. All components are derived in a Laplacian matrix representation, $w^T XL(k)X^T w$, for $k = 1, 2, …, K$. Therefore, function F in (1) becomes

$$F = \sum_i \sum_{m \neq n}\left\|y_i - f_{m,n}^{(2)}(y_i)\right\|^2 w_{m,n}^{(2)}(y_i)$$
$$= \sum_i \sum_{m \neq n}(y_i - t_{n,m}y_m - t_{m,n}y_n)^2 w_{m,n}^{(2)}(y_i)$$
$$= \sum_i\left(y_i - \sum_j M_{i,j}(1)y_j\right)^2 + \sum_i\left(y_i - \sum_j M_{i,j}(2)y_j\right)^2$$
$$+ … + \sum_i\left(y_i - \sum_j M_{i,j}(K)y_j\right)^2$$
$$= tr(Y^T(I-M(1))^T(I-M(1))Y$$
$$+ Y^T(I-M(2))^T(I-M(2))Y$$
$$+ … + Y^T(I-M(K))^T(I-M(K))Y)$$
$$= tr(Y^T(D-W(1))Y) + Y^T(D-W(2))Y$$
$$+ … + Y^T(D-W(K))Y)$$
$$= tr(Y^T(L(1))Y + Y^T(L(2))Y + …Y^T(L(L))Y)$$
$$= tr(Y^T LY) = tr(w^T XLX^T w)$$

(4)

where $W_{i,j}(k) = (M(k) + M(k)^T - M(k)^T M(k))_{i,j}$ and $L(k) = D(k) - W(k)$, for $k = 1, 2, …, K$, and $L = L(1) + L(2) + … + L(K)$. Since the objective function in (4) can be represented as a Laplacian matrix, the locality of the samples is also preserved in the low-dimensional space. More details are given in [13].

Consider the class labels in supervised classification, the two parameters, K_1 and K_2 are manually determined for the computation of the within-class scatter $\mathbf{S}_w$ and the between-class scatter $\mathbf{S}_b$, respectively

$$\mathbf{S}_w = \sum_{p=1}^{N_c}\left(\sum_{\substack{x_i \in C_p \\ f^{(2)} \in F_{K_1}^{(2)}(x_i, C_p)}}\left(x_i - f^{(2)}(x_i)\right)\left(x_i - f^{(2)}(x_i)\right)^T\right)^2$$

(5)

$$\mathbf{S}_b = \sum_{p=1}^{N_c}\left(\sum_{\substack{l=1 \\ l \neq p}}^{N_c}\sum_{\substack{x_i \in C_p \\ f^{(2)} \in F_{K_2}^{(2)}(x_i, C_l)}}\left(x_i - f^{(2)}(x_i)\right)\left(x_i - f^{(2)}(x_i)\right)^T\right)$$

(6)

in which $F_{K_1}^{(2)}(x_i, C_p)$ indicates the K_1 NFLs within the same class C_p of point x_i and $F_{K_2}^{(2)}(x_i, C_l)$ is a set of the K_2 NFLs belonging to the different classes of point x_i. The Fisher criterion $tr(\mathbf{S}_B / \mathbf{S}_W)$ is then maximized to find the projection matrix w which is composed of the eigenvectors with the corresponding largest eigenvalues. A new sample in the low-dimensional space can be obtained by the linear projection $y = w^T x$. After that, the NN (one-NN) matching rule is applied to classify the samples. The training algorithm for the NFLE transformation proposed in our previous work [13] is described in Fig. 4.

Although the point-to-line strategy is successfully adopted in the training phase instead of the classification phase for the nearest feature line-based transformation, some drawbacks still remained and limited its performance. The problems are as follows: (1) extrapolation/interpolation inaccuracy: NFLE may not preserve the locality precisely when prototypes are far away from the probes (the probes are the training samples that would be projected on the NFL, and the prototypes are the training samples that generate the NFL); (2) high computation complexity: a large number of feature lines are generated when there are too many training samples; and (3) singular problem: the NFLE needs the inverse procedure to find the final transformation matrix w, which is troubled by the problem of singularity especially when the sample size is small. Motivated from the three problems of NFLE, we propose a modified NFLE algorithm to avoid the above three problems. Meanwhile, the algorithm is optimized for detecting fall incidents. The

Training algorithm for the NFLE transformation

Input: A training set $X = [x_1, x_2, ..., x_N] \in R^{d \times N}$ and two parameters K_1 and K_2.

Output: The projection transformation w.

Step 1. PCA projection: Sample data are transformed from space R^d into a low-dimensional subspace R^n. Let w_{PCA} denote the transformation matrix of PCA.

Step 2. Computation of the within-class scatter: The possible feature lines are first generated from the samples within the same class of a specified point x. Find the projection points and calculate the corresponding distances. Select K_1 vectors with the smallest distances and compute the within-class scatter $\mathbf{S}_w$ by Equation (5).

Step 3. Computation of the between-class scatter: Generate the possible feature lines from the samples that belong to the different class of point x. Find the projection points and calculate the corresponding distances. Select K_2 discriminant vectors with the smallest distances from point x to the feature lines. The between-class scatter $\mathbf{S}_b$ is obtained from Equation (6).

Step 4. Fisher criterion maximization: The Fisher criterion $$w^* = \arg \max_{\mathbf{w}} \mathbf{S}_b / \mathbf{S}_w$$ is maximized to obtain the best transformation matrix, which is composed of r eigenvectors with the largest eigenvalues.

Step 5. Output the final transformation matrix: $w = w_{PCA} w^*$.

Fig. 4 Training algorithm for the NFLE transformation

reason why we apply the modified NFLE is stated as follows: NFLE generates virtual training samples by linearly interpolating or extrapolating each pair of feature points. By doing so, the generalization and data diversity are increased. However, three drawbacks are shown as well. For the completeness and no repetition, the details of the three problems of NFLE and the proposed modified NFLE (NNFLE) algorithm are elaborated in Section 3.4.
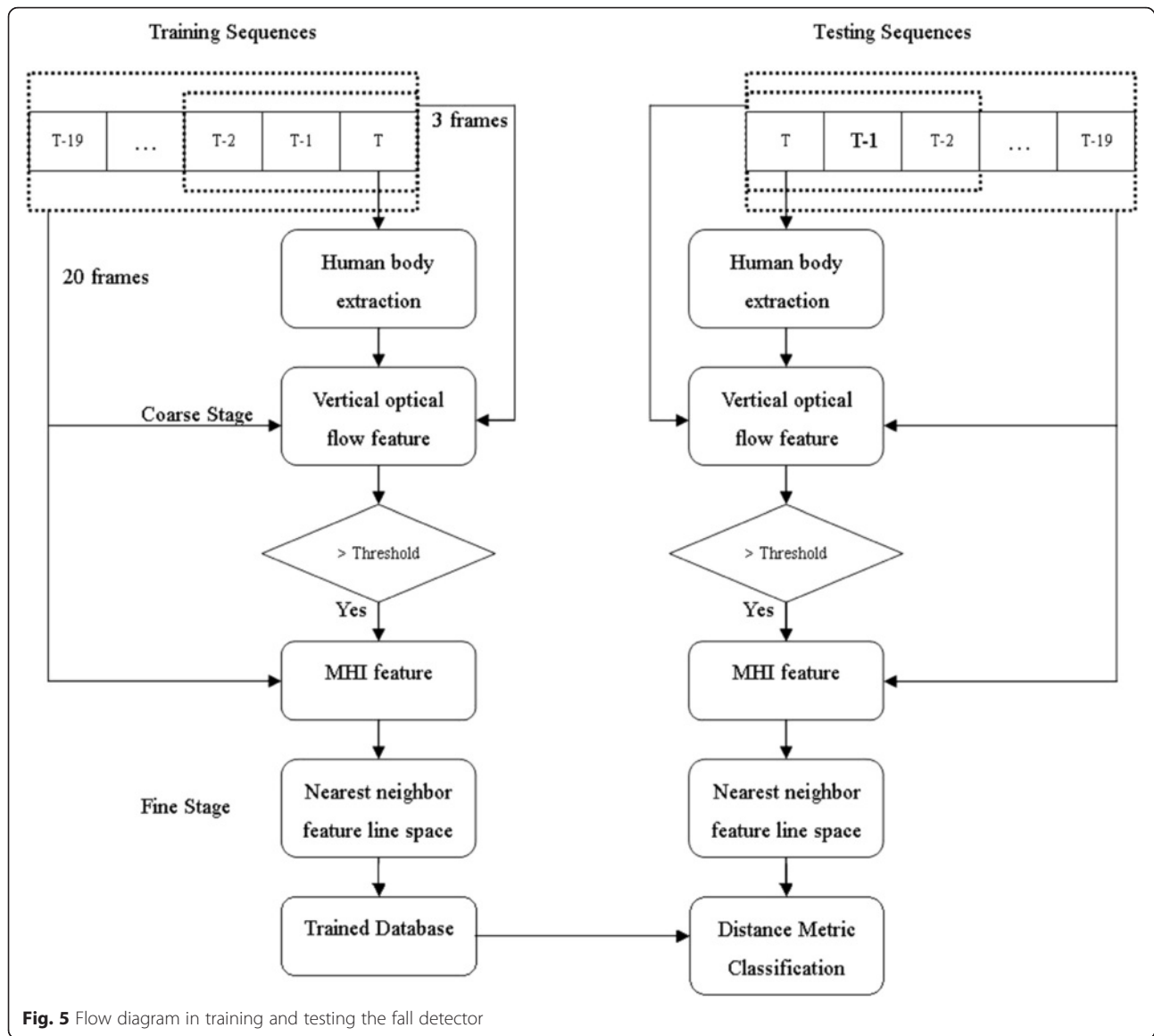
3 The proposed fall detection mechanism

The proposed fall detection mechanism consists of two modules including human body extraction and fall detection. In human body extraction module, temperature frames obtained from the thermal imager are processed with image processing techniques to obtain intact human body contours and silhouettes. In fall detection module, a coarse-to-fine strategy is devised to verify fall incidents. In the coarse stage, the downward optical flow features are extracted from the temperature images to identify possible fall down actions. Then, the 50-dimensional temporal-based motion history image (MHI) feature vectors are projected into the nearest neighbor feature line space to verify the fall down incident in the fine stage. Figure 5 depicts the proposed system flow diagram. The details associated with each step including the human body extraction, the analysis of optical flows in the coarse stage, the extraction of MHIs in the fine stage, and the nearest neighbor feature line embedding for fall verification are described in the following contexts.

3.1 Human body extraction

To improve fall detection accuracy, complete silhouettes of human body must be extracted to obtain accurate bounding box of human body. To this end, the temperature images captured from a thermal imager are binarized by Otsu's method firstly. Then, the morphological closing operation is employed to obtain a complete human silhouette. Finally, a labeling process is

Chen *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:16

Page 6 of 14

Fig. 5 Flow diagram in training and testing the fall detector

performed to locate each human body in the image and filter out background noises. The process of human body extraction is depicted in Fig. 6. Figure 6a shows the temperature images captured from the thermal imager, Fig. 6b shows the Otsu's binarization results, and Fig. 6c shows the results of morphological closing operation. The bounding box of the human silhouette can be successfully generated after the morphological closing operations.

3.2 Optical flow in the coarse stage
After the bounding box of human body has been determined, a coarse-to-fine strategy is utilized to verify fall incidents. The purpose of the coarse stage is to identify possible fall actions. Wu [16] had shown that a fall could be described by the increase in horizontal and vertical velocities. Moreover, this work observes that the histogram of vertical optical flows has also demonstrated the significant difference between walking and falling (see Fig. 7). In our work, a multi-frame optical flow method proposed by Wang [17] is adopted to extract the downward optical flow features inside the extracted bounding box (see Fig. 8) in this stage. A possible fall action can be identified by two heuristic rules:

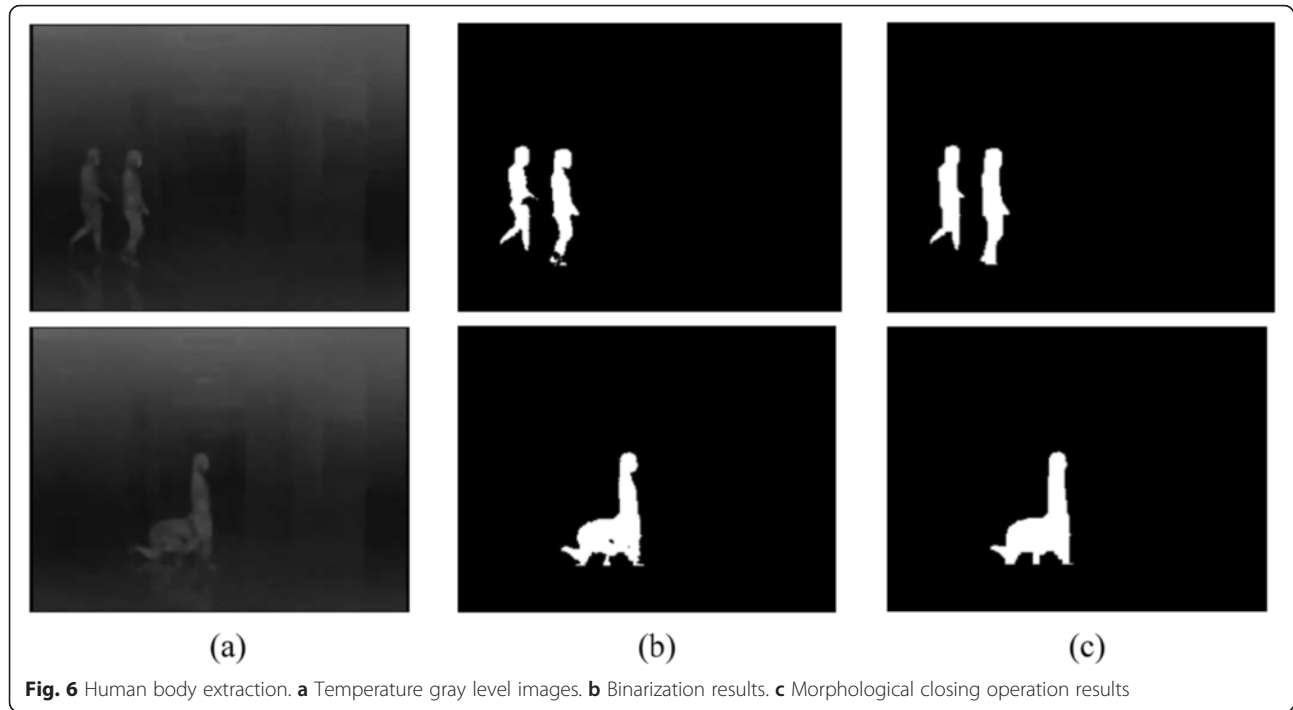(1) Rule 1: Given 20 consecutive frames, the average vertical optical flows exhibit downward more than 75 % of frames.

Chen *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:16

Page 7 of 14

Fig. 6 Human body extraction. **a** Temperature gray level images. **b** Binarization results. **c** Morphological closing operation results

(2) Rule 2: The sum of the average vertical optical flows in 20 consecutive frames is larger than a threshold, say 10 in this study.

As shown in Fig. 8a, a fall incident may not be identified if the subject is overlapped by the other. To solve this problem, the bounding box is divided into two equal boxes if overlapping occurs. The width of the silhouette is used to identify whether the overlapping occurs or not. The optical flow features are then extracted in each divided box. The one which has larger average downward optical flow is used to identify possible fall action. As a result, the fall incidents can be extracted correctly as shown in Fig. 8b, and Fig. 8a demonstrates the result without using the bounding box division strategy.

3.3 Motion history image in the fine stage

In the coarse stage, most non-fall actions can be filtered out via the downward optical flow features. However, some fall-like actions are identified as fall incidents due to the swing of arms. To solve this problem, we devise a feature vectors which are formed by projecting the MHI horizontally to verify fall incidents in the fine stage. MHI proposed by Bobick [18] is a template which condenses a determined number of silhouette sequences into a gray scale image (as shown in Fig. 9a) which is capable of preserving dominant motion information. Since the main difference between fall and other actions is the vertical component

changes, our work projects the MHI horizontally to obtain a 50-dimensional feature vectors using equation (7):

$$Q(i) = \frac{1}{U_w} \sum_{j=1}^{n} g\left(\left\lfloor \frac{U_h}{50} \times i \right\rfloor, j\right), \quad i = 1, 2, ..., 50 \quad (7)$$

where U_h, U_w, and $g(i, j)$ are the height, the width, and the pixel value of the motion energy in row i and column j, respectively. $Q(i)$ is the obtained 50-dimensional feature vectors. Figure 9c, f illustrates the comparison between the feature vectors of walk and fall in this study. The distributions of the walk action and the fall action are significantly different. As can be seen, the vertical motion information of the fall action is encoded directly with the horizontal projections, which can be viewed as extracting from MHI but not the silhouette. Therefore, the MHI features of fall-like actions will be fed into the constructed NNFLE verifier to identify fall incidents after the coarse stage.

3.4 Nearest neighbor feature line embedding (NNFLE)

Because the projection of MHI is a high-dimensional feature vector, a dimensional reduction scheme is employed to extract more salient features for fall detection. In our previous work [13], NFLE has demonstrated its effectiveness in pattern recognition. However, three problems of the NFLE have also been indicated in Section 2. To mitigate the three problems of NFLE, a
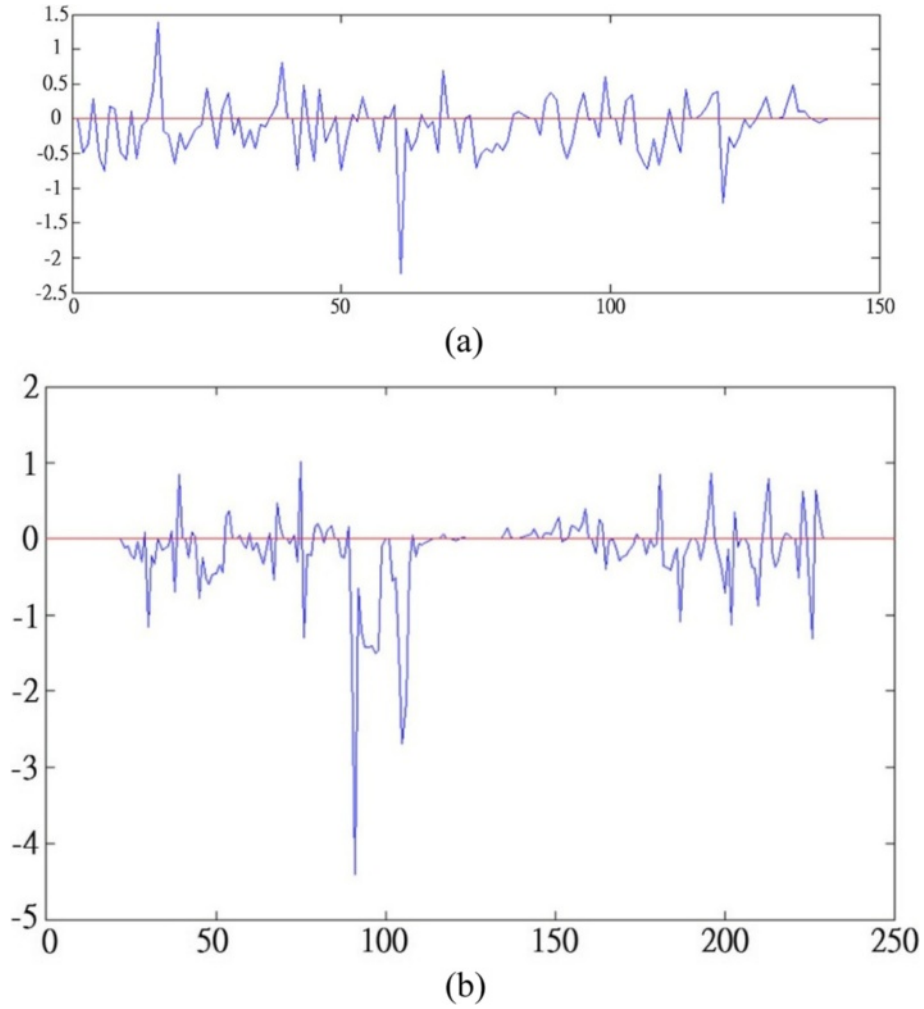
Chen *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:16

Page 8 of 14

Fig. 7 The histogram of vertical optical flow of (**a**) walking and (**b**) falling down

modified NFLE termed Nearest Neighbor Feature Line Embedding (NNFLE) is proposed as a fall verifier in the fine stage. Here, given a feature vector x_i, which is extracted from MHI, the proposed NNFLE method is formulated as the following optimization problem:

$$\max_{w} J(w) = \sum_{i=1}^{N} \left\| w^T x_i - w^T x_i^{\text{between}} \right\|^2 - \sum_{i=1}^{N} \left\| w^T x_i - w^T x_i^{\text{within}} \right\|^2 \tag{8}$$

where x_i^{within} indicates the projected point of x_i on the nearest neighbor feature lines (NNFLs) formed by the samples with the same labels, and x_i^{between} indicates the projected point of x_i on NNFLs formed by the samples with different labels from x_i. Here, it has to be mentioned that in the NFLE, each NFL is formed by the samples with the same class. However, in the proposed NNFLE, the NNFLs on which the projected point x_i^{between} of the x_i could be formed by the samples with

different labels from each other. In other words, all the other classes are treated as one class while calculating the projected point x_i^{between}.

With some algebraic operation, the $J(w)$ can be simplified to the following form:

$$\begin{aligned}
J(w) &= \sum_{i=1}^{N} \left\| w^T x_i - w^T x_i^{\text{between}} \right\|^2 - \sum_{i=1}^{N} \left\| w^T x_i - w^T x_i^{\text{within}} \right\|^2 \\
&= \sum_{i=1}^{N} tr\left[w^T \left(x_i - x_i^{\text{between}} \right) \left(x_i - x_i^{\text{between}} \right)^T \right] \\
&\quad - \sum_{i=1}^{N} tr\left[w^T \left(x_i - x_i^{\text{within}} \right) \left(x_i - x_i^{\text{within}} \right)^T \right] \\
&= tr[w^T (\mathbf{S}_B - \mathbf{S}_W) w]
\end{aligned} \tag{9}$$

Then, we impose a constraint $w^T w = 1$ on the proposed NNFLE. The transformation matrix w can thereby be obtained by solving the eigenvalue problem:

Chen *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:16
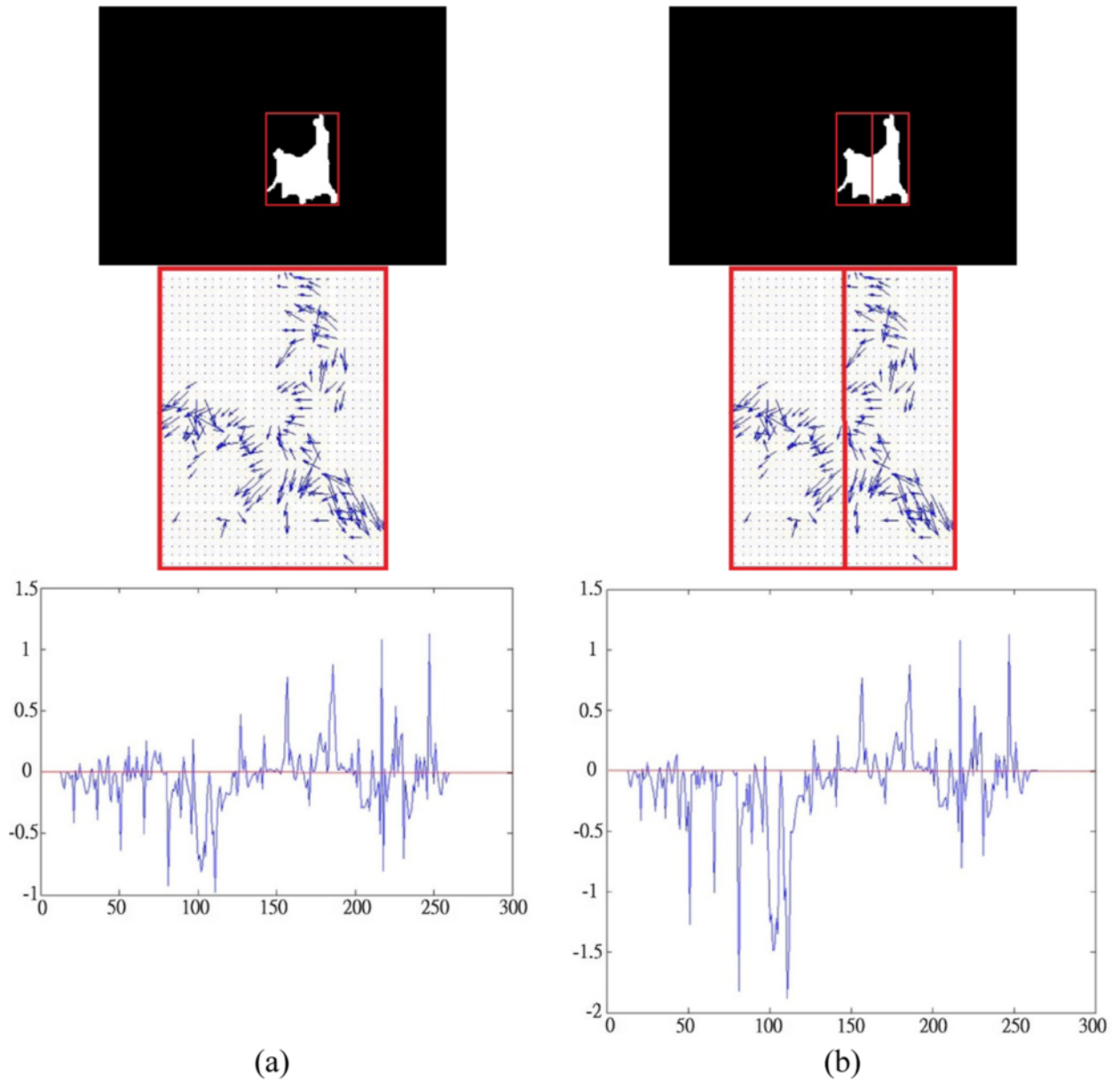
Page 9 of 14

Fig. 8 Fall incident in overlapping situation. The *first row* is the silhouettes, the *second row* is the corresponding optical flow results, and the *third row* is the histograms of vertical optical flows. **a** The results generated by original method. **b** The results generated by using dividing method

$$(\mathbf{S}_B - \mathbf{S}_W)w = \lambda w \qquad (10)$$

Since the proposed NNFLE method does not need the inverse of any matrix, it can solve the singular problem of NFLE. However, the extrapolation and interpolation errors existing in NFLE may decrease the performance of locality preserving as shown in Fig. 10. Let us consider two feature line points $L_{2,3}$ and $L_{4,5}$ generated from two prototype pairs (x_2, x_3) and (x_4, x_5), respectively. Points $f_{2,3}(x_1)$ and $f_{4,5}(x_1)$ are two projection points $L_{2,3}$ of lines $L_{2,3}$ and $L_{4,5}$ for a query point x_1. From Fig. 11, it is clear that point x_1 is close

to points x_2 and x_3 but far away from points x_4 and x_5. However, the distance $\|x_1 - f_{4,5}(x_1)\|$ for line $L_{4,5}$ is smaller than that for line $L_{2,3}$, i.e., $\|x_1 - f_{2,3}(x_1)\|$. The discriminant vector for line $L_{4,5}$ to point x_1 is hence selected instead of the other one. In addition, a great deal of computational time is needed due to the vast number of feature lines in the classification phase, e.g., C_2^{N-1} possible lines.

To overcome the inaccuracy problem resulted from extrapolation and interpolation, feature lines for a query point are generated from the k nearest neighborhood
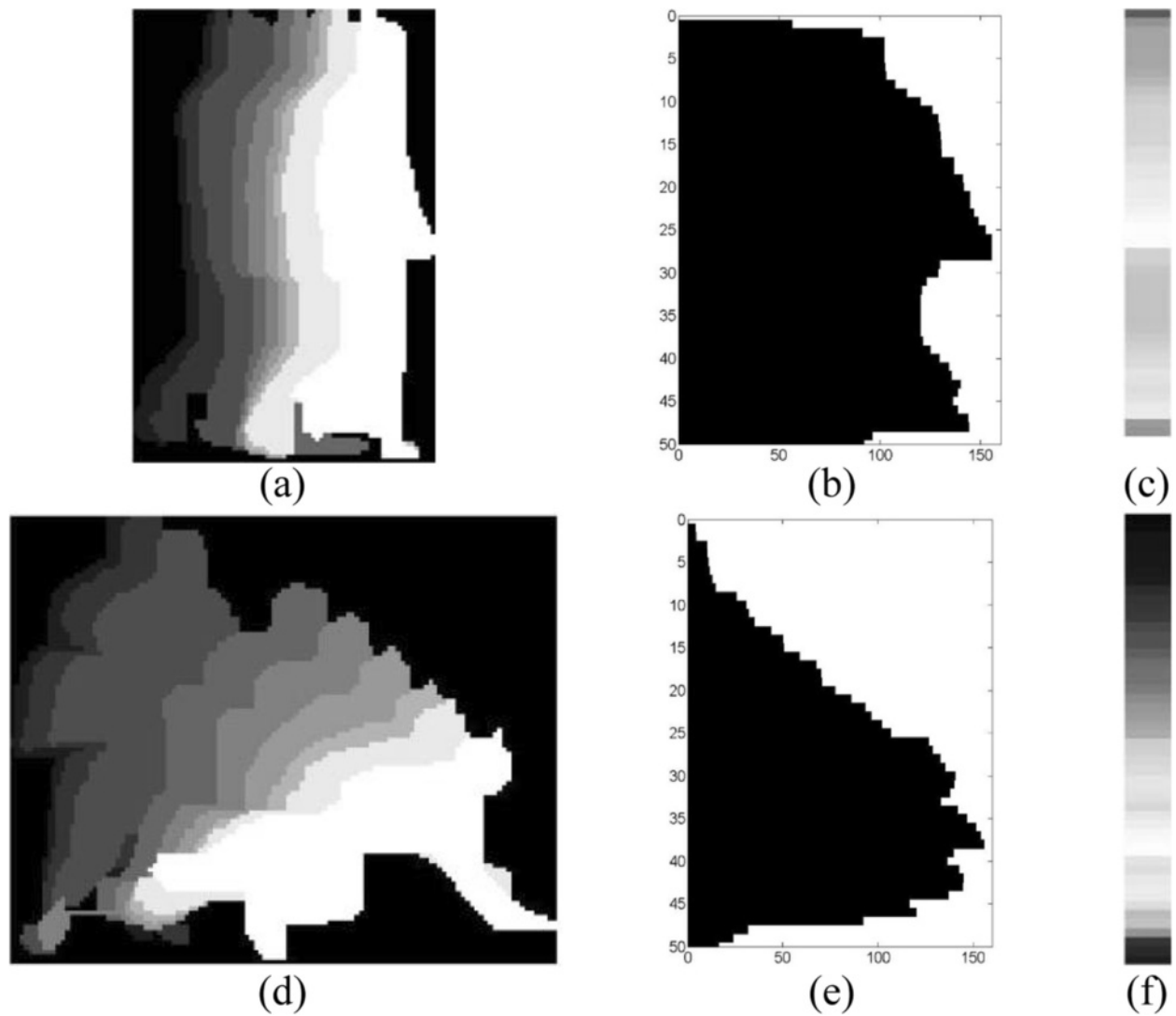
Chen *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:16

Page 10 of 14

Fig. 9 Fine stage feature vector extraction. **a** MHI of walk. **b** Horizontal projection of walk MHI. **c** The obtained fine stage feature vector from walk MHI. **d** MHI of fall. **e** Horizontal projection of fall MHI. **f** The obtained fine stage feature vector from fall MHI
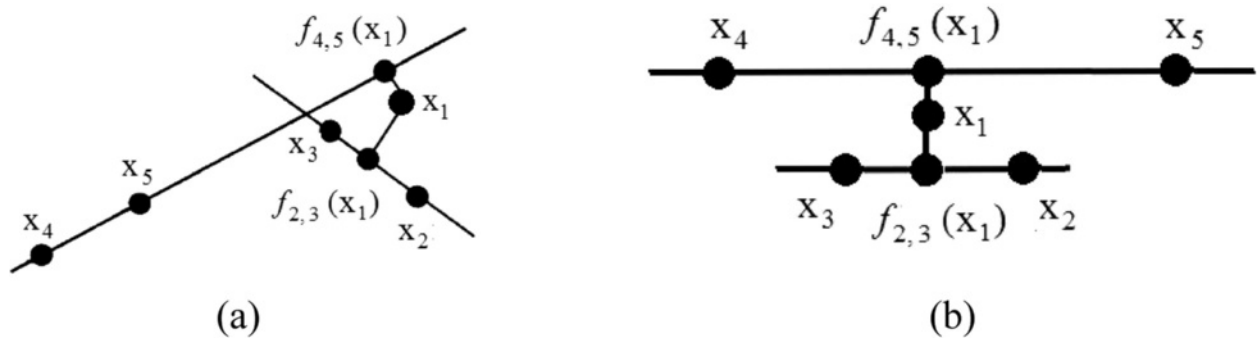

Fig. 10 a An extrapolation error. **b** An interpolation error

---

Training algorithm for the NNFLE transformation

Input: A training set $X = [x_1, x_2, \ldots x_N] \in R^{d \times N}$ and four parameters k_1, k_2, K_1, and K_2.

Output: The projection transformation W

Step 1.   PCA projection: Sample data are transformed from space R^d into a low-dimensional subspace R^n. Let w_{PCA} denote the transformation matrix of PCA.

Step 2.   Computation of the within-class scatter: The possible NNFLs are generated from the k_1 nearest neighbor samples within the same class of a specified point x. Find the projection points and calculate the corresponding distances. Select K_1 vectors with the smallest distances and compute the within-class scatter S_W by Equation (11).

Step 3.   Computation of the between-class scatter: Select k_2 nearest neighbor samples in different classes from a specified point x. Generate the possible NNFLs, find the projection points and calculate the corresponding distances. Select K_2 discriminant vectors with the smallest distances from point x to the feature lines. The between-class scatter S_B is obtained from Equation (12).

Step 4.   Objective function maximization: $w^* = \arg \max\limits_{w} (S_B - S_W)$ is maximized to obtain the best transformation matrix, which is composed of r eigenvectors with the largest eigenvalues.

Step 5.   Output the final transformation matrix: $w = w_{PCA} w^*$.

Fig. 11 Training algorithm for the NNFLE transformation

prototypes. More specifically, when two points x_m and x_n belong to the nearest neighbors of a query point x_i, a straight line passing through points x_m and x_n is NNFL. The discriminant vector $x_i - f_{m,n}(x_i)$ is chosen for the scatter computation. The selection strategy for discriminant vectors in NNFLE is designed as follows:

(1) The within-class scatter S_W: The NNFLs are generated from the k_1 nearest neighbor samples within the same class for the computation of the within-class scatter matrix, i.e., a set $F_{k_1}^+(x_i)$.

(2) The between-class scatter S_B: Select k_2 nearest neighbor samples in different classes from a specified point x_i, i.e., a set $F_{k_2}^-(x_i)$, to generate the NNFLs and calculate the between-class scatter matrix.

$$\mathbf{S}_W = \sum_{p=1}^{N_c} \left(\sum_{\substack{\mathbf{x}_i \in C_p \\ f^{(2)} \in F_{k_1}^+(x_i)}} (x_i - f(x_i))(x_i - f(x_i))^T \right) \quad (11)$$

Chen *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:16

Page 12 of 14

$$\mathbf{S}_B = \sum_{p=1}^{N_c} \left(\sum_{\substack{l=1 \\ l \neq p}}^{N_p} \sum_{\substack{x_i \in C_p \\ f \in F_{k_2}^-(\mathbf{x}_i)}} (x_i - f(x_i))(x_i - f(x_i))^T \right)$$

(12)

The training algorithm for the NNFLE transformation proposed in this study is described in Fig. 11.

The proposed NNFLE method is a simple and effective method to alleviate the extrapolation and interpolation errors. In addition, the scatter matrices are also generated based on the Fisher's criterion and represented as a Laplacian matrix. Moreover, the complexity of NNFLE is more efficient than that of NFLE. Consider N training samples, C_2^{N-1} possible feature lines will be generated and C_2^{N-1} distances have to be calculated for a specified point. The K_1 nearest feature lines are chosen from all possible lines to calculate the class scatter. The time complexity is $O(N^2)$ for line generation and $O(2N^2 \log N)$ for distance sorting. At the same time, the time complexity for selecting the K_1 nearest feature lines is $O(k^2)$ + $O(2k^2 \log k)$ when nearest prototypes are chosen for line generation. Extra overhead $O(N \log N)$ is needed for finding the k nearest prototypes. When N is large, traditional method needs longer time to calculate the class scatter.

4 Experimental results

In this section, experimental results conducted on fall incident detection are illustrated to demonstrate the effectiveness of the proposed method. This work compares the proposed method with two state-of-the-art methods. Results are evaluated by using the simulated video data set captured from outdoor scenes. The data set is formed by 320 videos. In each video, the environment is in the dusky environments as shown in Fig. 2. Only the thermal imager can effectively capture the human silhouette under the environments. Table 1 tabulates the data sets used in the experiments. In this study, videos used for training are different from that used for testing. More specifically, training videos and testing videos were captured under different conditions (different places at different time). Among these data sets, video sequences which contain only one subject are utilized to compare

the performance of the proposed method with other state-of-art fall detection methods and the results will be illustrated in Section 4.1. The identification capability of coarse-to-fine verifier is evaluated and illustrated in Section 4.2. In Section 4.3, the performance of the proposed method is evaluated by using video sequences which contain multiple subjects. Different from the other researches, the experimental results in Section 4.3 demonstrate that the proposed method can effectively detect fall incidents even when multiple persons overlap.

4.1 Performance comparisons of various fall detection algorithms

The data sets used in this subsection contain only one subject in each video sequence. Two state-of-the-art methods, BBN [11] and CPL [12], are implemented for comparison. The CPL takes a sequence as a sample, whereas the BBN and our proposed method take a frame as a sample. Therefore, the performance comparison of these three methods is based on each video sequence. In the experiments, 60 video sequences of one person are used as training sets and 100 video sequences of one person are used for testing. In addition, the projection matrix w of the proposed NNFLE is constructed from the eigenvectors of $\mathbf{S}_b - \mathbf{S}_w$ with the largest corresponding eigenvalues when the objective function J is maximized. In our work, the dimensionality of feature vectors is reduced by the PCA transformation to remove noises. More than 99 % of the feature information is kept in the PCA process. After the PCA transformation, the optimal projection transformations are obtained for the proposed NNFLE method. All of the testing frames are matched with the trained prototypes using the NN matching rule. The performance comparisons of these three methods are tabulated in Table 2. From Table 2, we can notice that the proposed coarse-to-fine strategy of fall detection outperforms the other two methods. It implies that the proposed method is much more effective than the other two methods.

Table 1 The data sets used in the experiments

Actions	Number of training videos	Number of testing videos
Walk (one person)	30 (5135 frames)	50 (17,125 frames)
Fall (one person)	30 (545 frames)	50 (1822 frames)
Walk (multiple persons)	30 (5069 frames)	50 (16,130 frames)
Fall (multiple persons)	30 (460 frames)	50 (1810 frames)

Table 2 The fall detection performance on the data set (%)

Method	Classification action (videos)	Reference action (videos)	
		Fall	Walk
CPL	Fall	92.00 (46/50)	8.00 (4/50)
	Walk	10.00 (5/50)	90.00 (45/50)
BBN	Fall	80.00 (40/50)	20.00 (10/50)
	Walk	12.00 (6/50)	88.00 (44/50)
NFLE	Fall	94.00 (47/50)	6.00 (3/50)
	Walk	6.00 (3/50)	94.00 (47/50)
Ours	Fall	98.00 (49/50)	2.00 (1/50)
	Walk	0.00 (0/50)	100.00 (50/50)

Chen *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:16

Page 13 of 14

Table 3 The identification capability of coarse stage and fine stage of the proposed method

Classification actions	Reference actions in coarse stage (frames)		Reference actions in fine stage (frames)	
	Walk	Fall	Walk	Fall
Walk	17,079/17,125	46/17,125	46/46	0/46
Fall	68/1822	1754/1822	153/1754	1601/1754

4.2 The identification capability of coarse-to-fine verifier

In this subsection, the discriminability of the proposed coarse-to-fine strategy is analyzed as tabulated in Table 3. The identification capability is evaluated by the total number of frames of those video sequences which contains only one person. Among these 18,947 frames, there are 1822 and 17,125 frames of "fall" and "walk" actions predefined, respectively. As depicted in Table 3, the proposed method can identify most of the walk actions in the coarse stage. Only a small amount of fall-like actions are needed to be verified in the fine stage. In other words, almost all of the fall actions can pass through the coarse stage filter. Hence, the proposed coarse stage is very useful for pre-filtering non-fall actions so that the performance of the NNFLE classifier in the fine stage is less affected by the noisy data in both training and testing phases.

4.3 Performance evaluation of fall detection under overlapping situations

In this subsection, the performance evaluation of fall detection in overlapping situations is illustrated. Video sequences which contain multiple persons are used for evaluation. Similar to the comparison described in Section 4.1, the NN matching rule is adopted to identify each testing frames in the fine stage. In the experiments, the performance evaluation of fall detection under overlapping situations is conducted based on each video sequence. Here, 30 video sequences are used for training and 100 video sequences are used for testing. The detection results are tabulated in Table 4. The proposed method utilizing coarse-to-fine strategy can effectively detect fall incidents while two persons are overlapping each other, and the performance is almost the same as that of the "one person fall" data sets described in Section 4.1.

Table 4 The performance evaluation of fall detection under overlapping situations (%)

Method	Classification action (videos)	Reference action (videos)	
		Fall	Walk
NFLE	Fall	92.00 (46/50)	10.00 (5/50)
	Walk	6.00 (3/50)	90.00 (45/50)
Ours	Fall	96.00 (48/50)	4.00 (2/50)
	Walk	0.00 (0/50)	100.00 (50/50)

5 Conclusions

In this paper, a novel fall detection mechanism based on a coarse-to-fine strategy in dusky environment is proposed. The human body in dusky environment can be successfully extracted using the thermal imager, and fragments inside the human body silhouette can also be significantly reduced as well. In the coarse stage, the optical flow algorithm is applied on thermal images. Most of walk actions are filtered out by analyzing the downward flow features. In the fine stage, the projected MHI is used as the features followed by the NNFLE method to verify fall incidents. The proposed NNFLE method, which adopts a nearest neighbor selection strategy, is capable of alleviating extrapolation/interpolation inaccuracies, singular problem, and high computation complexity. Experimental results demonstrate that the proposed method outperforms the other state-of-the-art methods and can effectively detect fall incidents even when multiple subjects are moving together.

Author details
[1]Department of Computer Science and Information Engineering, National Central University, Taoyuan, Taiwan. [2]Department of Applied Informatics, Fo Guang University, Yilan, Taiwan. [3]Digital Education Institute, Institute for Information Industry, Taipei, Taiwan.

References
1. KC Moylan, EF Binder, Falls in older adults: risk assessment, management and prevention. Am. J. Med. **120**(6), 493–497 (2007)
2. L Larson, TF Bergmann, Taking on the fall: the etiology and prevention of falls in the elderly. Clin. Chiropr. **11**(3), 148–154 (2008)
3. S Gs, Falls among the elderly: epidemiology and prevention. Am. J. Prev. Med. **4**(5), 282–288 (1988)
4. J Tao, M Turjo, M-F Wong, M Wang, Y-P Tan, Fall incidents detection for intelligent video surveillance, in *Proceedings of the 15th international conference on communications and signal processing*, 2005, pp. 1590–1594
5. D Anderson, JM Keller, M Skubic, X Chen, Z He, Recognizing falls from silhouettes, in *Proceedings of the 28th IEEE EMBS annual international conference*, 2006
6. CF Juang, CM Chang, Human body posture classification by neural fuzzy network and home care system applications. IEEE Trans. SMC, Part A **37**(6), 984–994 (2007)
7. H Foroughi, N Aabed, A Saberi, HS Yazdi, An eigenspace-based approach for human fall detection using integrated time motion image and neural networks, in *Proceedings of the IEEE International Conference on Signal Processing (ICSP)*, 2008
8. C Rougier, J Meunier, AST Arnaud, J Rousseau, Fall detection from human shape and motion history using video surveillance, in *Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops*, vol. 2, 2007, pp. 875–880
9. H Foroughi, A Rezvanian, A Pazir023e, Robust fall detection using human shape and multi-class support vector machine, in *Proceedings of the Sixth Indian Conference on CVGIP*, 2008
10. CL Liu, CH Lee, P Lin, A fall detection system using k-nearest neighbor classifier. Expert Syst. Appl. **37**(10), 7174–7181 (2010)
11. YT Liao, CL Huang, SC Hsu, Slip and fall event detection using Bayesian Belief Network. Pattern Recogn. **45**, 24–32 (2012)

Chen *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:16

Page 14 of 14

12. DN Olivieri, IG Conde, XAV Sobrino, Eigenspace-based fall detection and activity recognition from motion templates and machine learning. Expert Syst. Appl. **39**(5), 5935–5945 (2012)
13. YN Chen, CC Han, CT Wang, KC Fan, Face recognition using nearest feature space embedding. IEEE Trans. Pattern Anal. Mach. Intell. **33**(6), 1073–1086 (2012)
14. SZ Li, J Lu, Face recognition using the nearest feature line method. IEEE Trans. Neural Netw. **10**(2), 439 (1999). -433, 1999
15. S Yan, D Xu, B Zhang, HJ Zhang, S Lin, Graph embedding and extensions: general framework for dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell. **29**(1), 40–51 (2007)
16. G Wu, Distinguishing fall activities from normal activities by velocity characteristics. J. Biomech. **33**(11), 1497–1500 (2000)
17. CM Wang, KC Fan, CT Wang, Estimating optical flow by integrating multi-frame information. J. Inf. Sci. Eng. **24**(6), 1719–1731 (2008)
18. AF Bobick, JW Davis, The recognition of human movement using temporal templates. IEEE Trans. Pattern Anal. Mach. Intell. **23**(3), 257–267 (2001)