

FALL DETECTION IN THE ELDERLY BY HEAD TRACKING

Miao Yu, Syed Mohsen Naqvi and Jonathon Chambers

Advanced Signal Processing Group
Electronic and Electrical Engineering Department
Loughborough University
Loughborough, Leicester, UK
{elmy, s.m.r.naqvi, eljac}@lboro.ac.uk

ABSTRACT

In the paper, we propose a fall detection method based on head tracking within a smart home environment equipped with video cameras. A motion history image and code-book background subtraction are combined to determine whether large movement occurs within the scene. Based on the magnitude of the movement information, particle filters with different state models are used to track the head. The head tracking procedure is performed in two video streams taken by two separate cameras and three-dimensional head position is calculated based on the tracking results. Finally, the three-dimensional horizontal and vertical velocities of the head are used to detect the occurrence of a fall. The success of the method is confirmed on real video sequences.

Index Terms— motion history image, code-book background subtraction, particle filtering, head tracking, fall detection

1. INTRODUCTION

In recent years, the topic of caring for old people has gained increasing public concern. Among the unexpected events which happen in the elderly group, falls are the leading cause of death due to injury and 87 percents of all fractures are caused by fall [1]. Traditional methods to detect falls use some wearable sensors. However, the problem of such detectors is that older people often forget to wear them and they are intrusive. In order to solve this problem, fall detection based on the digital video processing technology is developed.

There are many related works in the field of fall detection basing on the digital video processing techniques [2] [3] [4]. By putting a camera in the ceiling, Lee and Mihailidis [2] detect a fall using the shape of the person's silhouette, and Nait-Charif and McKenna [3] detect inactivity outside the normal zones of inactivity such as chairs or sofas. L.Hazelhoff and P.H.With [4] adopted a system with two uncalibrated cameras. A Gaussian multi-frame classifier helps to recognize fall events using the two features of the direction of the main

axis of the body and the ratio of the variances in the motion in the x and y directions.

In this paper, we propose a fall detection method based on the three-dimensional head velocities in both the horizontal and vertical directions. Unlike the above methods, this method can detect a fall event when it is happening so the alarm signals can be sent immediately. The structure of this paper is as follows: In Section 2, we introduce the concepts of motion history image (MHI) and the code-book background subtraction method, which are then combined to calculate a parameter defined as C_{motion} to determine whether the movement is large or small. In Section 3, a particle filter head tracking method based on gradient and colour information is proposed to track the head in two video streams recorded by the two cameras. The two-dimensional head tracking results are converted to three-D real head positions to obtain the head's horizontal and vertical velocities. Some experimental results are shown in Section 4 and in Section 5, conclusions and future work are given.

2. MOVEMENT DETERMINATION

Before head tracking, we must determine whether the movement is large or small so that a proper state model can be used for particle filter head tracking. In this paper, we determine the types of movements by a parameter C_{motion} , which is calculated from the motion history image and the result of background subtraction.

2.1. Motion History Image

The MHI, first introduced by Bobick and Davis [5] is an image where the pixel intensity represents the recency of motion in an image sequence. Motion history image shows the tendency of a person's movement by the magnitudes of its pixels so that it is commonly used for activity recognition.

The definition of the MHI is given as follow:

$$H_\tau(x, y, t) = \begin{cases} \tau & \text{if } D(x,y,t)=1 \\ \max(0, H_\tau(x, y, t-1) - 1) & \text{otherwise} \end{cases} \quad (1)$$

where $D(x,y,t)$ represents a pixel at the position (x,y) of a binary image D at the time t with ‘1’ representing the motion regions. It is extracted from the original image sequence $I(x, y, t)$ using an image-differencing method [5].

The resulting $H_\tau(x, y, t)$ is a scalar-valued image in which more recently moving pixels are brighter. We can see if large movement occurs, more brighter pixels will be in the area of a person.

2.2. Background Subtraction

To remove the unnecessary objects within the video scene, we use the code-book background subtraction method proposed in [6]. Compared with other background subtraction methods, it gives good results on image sequences with shadows, highlights and high image compression. Besides, it can cope efficiently with non-stationary background.

Initially, a code-book is trained as the background model. The procedure of code-book training is shown in detail in [6].

After obtaining the code-book, we carry out the background subtraction step for every pixel $P(x,y)$ as follow:

Step I. For each pixel $P(x,y)=(R,G,B)$, calculating the intensity from the (R,G,B) value of a colour image by

$$I \leftarrow \sqrt{R^2 + G^2 + B^2}$$

Step II. Find the first codeword c_m from the code-book at the position (x,y) matching to $P(x,y)$ based on two conditions:

1. $\text{colordist}(P(x,y), c_m) \leq \varepsilon_2$
2. $\text{brightness}(I, \langle \hat{I}_m, \check{I}_m \rangle) = \text{true}$

Update the matched codeword

Step III. If there is no match, then the pixel $P(x,y)$ is categorized as foreground, otherwise, it is regarded as a background pixel.

The $\text{colordist}(x, c_m)$ represents the color distortion between x and c_m and the $\text{brightness}(I, \langle \hat{I}_m, \check{I}_m \rangle)$ is true if I is in the range of $[\hat{I}_m, \check{I}_m]$.

2.3. Movement Classification

After obtaining the MHI and the background subtraction result, we can determine whether the movement is large or small by calculating the parameter C_{motion} . The formula to calculate C_{motion} is shown as follow [7]

$$C_{motion} = \frac{\sum_{\text{pixel}(x,y) \in \text{blob}} H_\tau(x, y, t)}{\#\text{pixels} \in \text{blob}} \quad (2)$$

where blob represents the region of the person extracted using the code-book background subtraction, and $H_\tau(x, y, t)$

is the MHI. The denominator $\#\text{pixels} \in \text{blob}$ can be obtained by multiplying the number of pixels in the extracted person’s region with 255. $C_{motion} = 0$ represents no motion and $C_{motion} = 1$ represents full motion. We assume that if $C_{motion} > 1/2$, then large movement occurs.

3. HEAD TRACKING

For head tracking, we use the conventional particle filter based on the condensation algorithm. The state model of the particle filter will be chosen differently according to the value of C_{motion} , while the measurement model is based on the gradient and colour information.

Particle filtering is an useful tool widely used for tracking, a particle filter consists of two models—state model and the measurement model, which can be represented as follow:

$$\begin{aligned} \mathbf{s}(t) &= f(\mathbf{s}(t-1)) + \mathbf{n}(t) \\ \mathbf{z}(t) &= f(\mathbf{s}(t)) + \mathbf{v}(t) \end{aligned} \quad (3)$$

where $\mathbf{s}(t)$ is the state vector at discrete time t and $\mathbf{z}(t)$ is the measurement vector.

Assume at time $t-1$, we have N particles $\{\mathbf{s}_{t-1}^i | i = 1 : N\}$ and the corresponding normalized weights $\{w_{t-1}^i | i = 1 : N\}$, such that $\sum_{i=1}^N w_{t-1}^i \mathbf{s}_{t-1}^i$ is the minimum mean square estimator (MMSE) of the state $\mathbf{s}(t-1)$. In order to obtain the MMSE at time t , the condensation algorithm is applied. The procedure of the condensation algorithm is as follows [8]:

For $\{\mathbf{s}_{t-1}^i | i = 1 : N\}$ and $\{w_{t-1}^i | i = 1 : N\}$

I. Obtain the new samples \mathbf{s}_t^i from the proposal distribution $p(\mathbf{s}_t | \mathbf{s}_t^i)$ for every i .

II. Calculate the weight w_t^i from $w_t^i = w_{t-1}^i p(\mathbf{z}_t^i | \mathbf{s}_t^i)$

III. Normalizing the w_t^i to make $\sum_{i=1}^N w_t^i = 1$

Finally, we calculate the MMSE $\hat{\mathbf{s}}(t)$ by $\hat{\mathbf{s}}(t) = \sum_{i=1}^N w_t^i \mathbf{s}_t^i$

In order to avoid the degeneracy problem [8], we re-sample the particles after every iteration to make all the weights of particles identical.

In our work, the state variable $\mathbf{s}(t)$ has the form $[x_t, y_t, l_t]$ corresponding to an ellipse, where x_t and y_t are the coordinates of the ellipse center and l_t is the shorter axis (we assume the ratio between the long and short axis is 1.2 for an ellipse representing a head). The state model adopted in this paper is a first-order random walk model with additive Gaussian noise, which has the following form: $\mathbf{s}(t) = \mathbf{s}(t-1) + \mathbf{n}(t)$. The variance of the added noise will be different according to the value of C_{motion} .

The measurement model used is based on two cues—the gradient and colour. Assuming the observations of measurement are independent, the following equation holds: $p(\mathbf{z}_t^i | \mathbf{s}_t^i) =$

$p(\mathbf{z}_{t,gradient}^i | \mathbf{s}_t^i) p(\mathbf{z}_{t,color}^i | \mathbf{s}_t^i)$. The formulas for calculating the $p(\mathbf{z}_{t,gradient}^i | \mathbf{s}_t^i)$ and $p(\mathbf{z}_{t,color}^i | \mathbf{s}_t^i)$ are:

$$p(\mathbf{z}_{t,gradient}^i | \mathbf{s}_t^i) = \frac{1}{\sqrt{2\pi}\sigma_g} e^{-\frac{(\frac{1}{2\sigma_g^2})^2}{2\sigma_g^2}} \quad (4)$$

and

$$p(\mathbf{z}_{t,color}^i | \mathbf{s}_t^i) = \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{(1-\rho[p_{s_t^i}, q])}{2\sigma_c^2}} \quad (5)$$

which are given in detail in [9].

4. EXPERIMENTS AND EVALUATIONS

Figures 1 to 4 show the MHI and background subtraction results for four situations (fast walking, slow walking, sitting down and a fall) and the corresponding C_{motion} is calculated.



Fig. 1. MHI and background subtraction result for the fast walking case, $C_{motion} = 63.55\%$



Fig. 2. MHI and background subtraction result for the slow walking case, $C_{motion} = 26.33\%$



Fig. 3. MHI and background subtraction result for the sitting down case, $C_{motion} = 46.20\%$

Large C_{motion} represents large movement and vice versa.



Fig. 4. MHI and background subtraction result for a fall, $C_{motion} = 61.24\%$

Figure 5 and Figure 6 show some tracking results of the video streams taken by the two cameras.



Frame No.1, Camera 1



Frame No.200, Camera 1



Frame No.400, Camera 1



Frame No.600, Camera 1

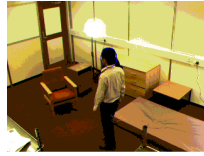
Fig. 5. Some tracking results for camera 1

Based on the tracking results obtained from the two video streams, we obtain the head's three-dimensional positions, then we calculate the head's horizontal and vertical velocities. Figure 7 shows the variations of the two velocities versus time for four different activities during 25s. For detecting falling, we set two 'alert thresholds' for the head's horizontal and vertical velocities respectively (0.4m/s and 0.3m/s). If both velocities exceed the 'alert threshold', then we assume a fall event has occurred.

During periods from 0s-5s, people walk very fast so the horizontal velocity is very high, however, the vertical velocity is low. From 5s-10s, people walks slowly and both velocities are low. During 10s-20s the person is the sitting down and standing up and large horizontal or vertical velocities may be found but at least one of them will be below the corresponding 'alert threshold'. At time 20s, the fall occurs and both velocities exceed the 'alert threshold'.



Frame No.1, Camera 2



Frame No.200, Camera 2



Frame No.400, Camera 2



Frame No.600, Camera 2

Fig. 6. Some tracking results for camera 2

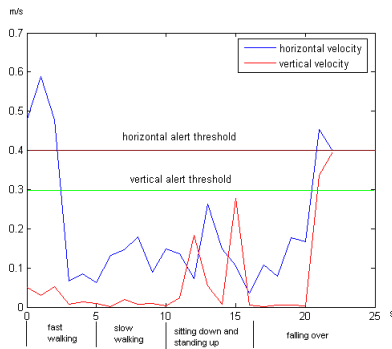


Fig. 7. Variation of velocities for the fast walking, slow walking, sitting down standing up and a fall in 25s

5. CONCLUSION AND FUTURE WORK

In this paper, we propose a new fall detection system based on head tracking in two video frames recorded by two cameras. The head's three-dimensional vertical and horizontal velocities are used as criteria for determining a falling event. A particle filter based on gradient and colour information is used for head tracking and its state model is determined by C_{motion} , a parameter which is calculated from the MHI and background subtraction results.

A more robust fall detection system can be achieved by the combination of audio and video information, which is well known as multimodal processing [1]. Blind source separation technique can be applied to extract the person's voice information, such as the words such as 'help', from the noisy environment. And a speech recognition system could then

be used to analyze the extracted voice to make a decision on whether a fall may have occurred.

6. REFERENCES

- [1] A.P.Lopez, "Multimodal source separation for cognitive assistive technology," *Erasmus Exchange Program Report, Loughborough University*, 2009.
- [2] T. Lee and A. Mihailidis, "An intelligent emergency response system: preliminary development and testing of automated fall detection," *Journal of Telemedicine and Telecare*, vol. 11, pp. 194–198, 2005.
- [3] H. Nait-Charif and S. McKenna, "Activity summarisation and fall detection in a supportive home environment," *In Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, vol. 4, pp. 323–326, 2004.
- [4] L.Hazelhoff, J.Han, and P.H.With, "Video-based fall detection in the home using principal component analysis," *In Proceedings of the 10th International Conference on Advanced Concepts for Intelligent Vision Systems*, vol. 11, pp. 298–309, 2008.
- [5] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 257–267, 2001.
- [6] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using code-book model," *Real-Time Imaging*, vol. 11, pp. 172–185, June 2005.
- [7] C.Rougier, J.Meunier, A.St-Arnaud, and 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, pp. 875–880, 2007.
- [8] B.Ristic, S.Arulampalam, and N.Gordon, "Beyond the Kalman filter: particle filters for tracking applications," *In Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops*, 2001.
- [9] X.Xu and B.Li, "Head tracking using particle filter with intensity gradient and colour histogram," *In IEEE International Conference on Multimedia and Exploration*, pp. 888–891, 2005.