Fall Detection from Human Shape and Motion History using Video Surveillance

Caroline Rougier and Jean Meunier
Département d'Informatique
et de Recherche Opérationnelle
Université de Montréal
Montréal, Canada
{rougierc, meunier}@iro.umontreal.ca

Alain St-Arnaud
Centre de santé
et de services sociaux
Lucille-Teasdale
Montréal, Canada
astarnaud@ssss.gouv.qc.ca

Jacqueline Rousseau
Centre de Recherche de
l'Institut Universitaire
de Gériatrie de Montréal
Montréal, Canada
jacqueline.rousseau@umontreal.ca

Abstract

Nowadays, Western countries have to face the growing population of seniors. New technologies can help people stay at home by providing a secure environment and improving their quality of life. The use of computer vision systems offers a new promising solution to analyze people behavior and detect some unusual events. In this paper, we propose a new method to detect falls, which are one of the greatest risk for seniors living alone. Our approach is based on a combination of motion history and human shape variation. Our algorithm provides promising results on video sequences of daily activities and simulated falls.

1. Introduction

1.1. Context

Canada, as others Western countries, faces the growing population of seniors. The statistics provided by the Public Health Agency of Canada [2] give us an idea of the problem: in 2001, one Canadian out of eight was older than 65 years old. In 2026, this proportion will be one out of five. It is important to notice that a majority of seniors, 93%, reside in private house, and among them, 29% live alone [2]. One of the greatest danger for old people living alone are the falls. Almost 62% of injury-related hospitalizations for seniors are the result of falls [3]. And the gravity of the situation can increase if the person can not call for help.

Nowadays, the usual solution to detect falls is to use some wearable sensors like accelerometers or help buttons. However, the problem of such detectors is that older people often forget to wear them. Moreover, in the case of a help button, it can be useless if the person is unconscious or immobilized.

To overcome these limitations, we use a computer vision system which doesn't require that the person wears anything. Another advantage of such a system is that a camera gives more information on the motion of a person and his/her actions than an accelerometer. Thus, we can imagine a computer vision system providing information on falls, but also, checking some daily behaviors (medication intake, meal and sleep time and duration, etc.).

Moreover, when we talk about vision system, we must ensure that the system is entirely automated to take into account the private life of the person. Another study is currently done by our team about the acceptability by older people of such vision systems.

1.2. Related Work

Recently, some research has been done to detect falls using image sensors. For instance, a simple method consists to analyze the bounding box representing the person in a single image [13]. This can only be done if the camera is placed sideways, and can fail because of occluding objects. Therefore, in realistic situation, the camera will be placed higher in the room to not suffer of occluding objects and to have a larger field of view. In this case, depending on the relative position of the person, the field of view of the camera, a bounding box will not be sufficient to discriminate a fall from a person sitting down.

To overcome this problem, some researchers [8] [9] have mounted the camera on the ceiling. Lee and Mihailidis [8] detect a fall by analyzing the shape and the 2D velocity of the person, and define inactivity zones like the bed. Nait-Charif and McKenna [9] track the person using an ellipse, and analyze the resulting trajectory to detect inactivity outside the normal zones of inactivity like chairs or sofas.

Infrared sensor can also be used to detect falls [12]. In their work, Sixsmith and Johnson classify falls with a neural network using the 2D vertical velocity of the person. But, 2D vertical velocity could not be sufficient to discriminate a real fall from a person sitting down abruptly.

21st International Conference on
Advanced Information Networking and Applications Workshops (AINAW'07)
0-7695-2847-3/07 $20.00 © 2007 IEEE

In our case, to cover large areas, we use wall-mounted cameras. In our previous work on fall detection [11], a fall was detected using the vertical and horizontal 3D velocities of the head of the person extracted from a monocular camera video sequence. 3D information is really helpful to analyze the actions of a person in a room. In this current work, we show the results of another method using simple 2D motion and shape information, which could be combined later on to 3D information.

1.3. System Overview

We present here a new method based on the Motion History Image and some changes in the shape of the person.

Motion History Image (MHI) Our method is based on the fact that the motion is large when a fall occurs. So, the first step of our system is to detect large motion of the person on the video sequence using the Motion History Image. This method is described in section 2.

Change in the Human Shape When a motion is detected, we analyze the shape of the person in the video sequence. During a fall, the human shape changes and, at the end of the fall, the person is generally on the ground with few and small body movements. A change in the human shape can discriminate if the large motion detected is normal (e.g.: the person walks or sits) or abnormal (e.g.: the person falls). The extraction of the human shape is described in section 3.

The different steps of our fall recognition system combining MHI and human shape are described in section 4.

2. Motion History Image

Motion gives a crucial information about fall, because no serious fall occurs without a large movement. Based on this observation, we decided to extract some motion information from the video sequence.

Optical flow [4] is commonly used to detect motion in a video sequence. But, optical flow is not well-suited for real-time application, and can generate errors in case of large movement as it happens during a fall. Another attempt to extract motion is the "Motion History Image" (MHI), first introduced by Bobick and Davis [5]. The MHI is an image where the pixel intensity represents the recency of motion in an image sequence, and therefore gives the most recent movement of a person during an action. The MHI is commonly used for activity recognition [5].

To define a MHI, we first extract a binary sequence of motion regions $D(x, y, t)$ from the original image sequence $I(x, y, t)$ using a image-differencing method. Then, each pixel of the Motion History Image H_τ is a function of the temporal history of motion at that point, occurring during a fixed duration τ (with $1 \leq \tau \leq N$ for a sequence of length N frames) [5] [1].

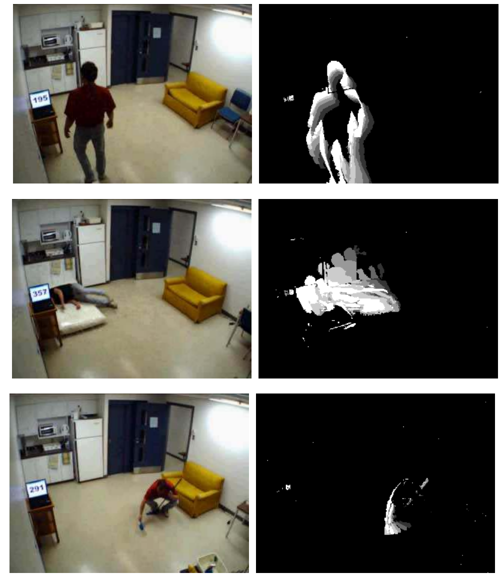

(a)
(b)
(c)

Figure 1. Motion History Images

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

The result is a scalar-valued image where more recently moving pixels are brighter.

MHI is useful in our case, because it is not necessary for us to detect the direction of the movement, we want above all to quantify the motion of the blob of a person. The motion will be high in the case of a fall. Figure 1 shows some examples of Motion History Images in various situations: (a) a walking person, (b) a fall, (c) a daily activity (housekeeping). In (b), a mattress is used to protect the person during the simulated fall.

3. Human Shape

To analyze the human shape, the person is segmented in the video sequence using a background subtraction method, and then, the blob is approximated by an ellipse. The steps of the human shape extraction are explained below.

3.1. Foreground Segmentation

First, we need to extract the moving person in the image. For this purpose, we use a background subtraction method described in the article [7], which gives good results on image sequences with shadows, highlights and high image compression. An example of background subtraction is shown in Fig. 2.

21st International Conference on
Advanced Information Networking and Applications Workshops (AINAW'07)
0-7695-2847-3/07 $20.00 © 2007 IEEE

3.2. Approximated Ellipse

The person is then approximated by an ellipse using moments [10]. An ellipse is defined by its center $(\bar{x}, \bar{y})$, its orientation θ and the length a and b of its major and minor semi-axes.

For a continous image $f(x, y)$, the moments are given by:

$$m_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy \qquad with \;\; p, q = 0, 1, 2, ...$$

The center of the ellipse is obtained by computing the coordinates of the center of mass with the first and zero-order spatial moments: $\bar{x} = m_{10}/m_{00}$, $\bar{y} = m_{01}/m_{00}$.

The centroid $(\bar{x}, \bar{y})$ is used to compute the central moment as follows:

$$\mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy$$

The angle between the major axis of the person and the horizontal axis x gives the orientation of the ellipse, and can be computed with the central moments of second order:

$$\theta = \frac{1}{2} \arctan \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right)$$

To recover the major semi-axis a and the minor semi-axis b of the ellipse, we need to compute I_{min} and I_{max}, respectively the least and greatest moments of inertia. They can be computed by evaluating the eigenvalues of the covariance matrix [10]:

$$J = \begin{pmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{pmatrix}$$

The eigenvalues I_{min} and I_{max} are given by:

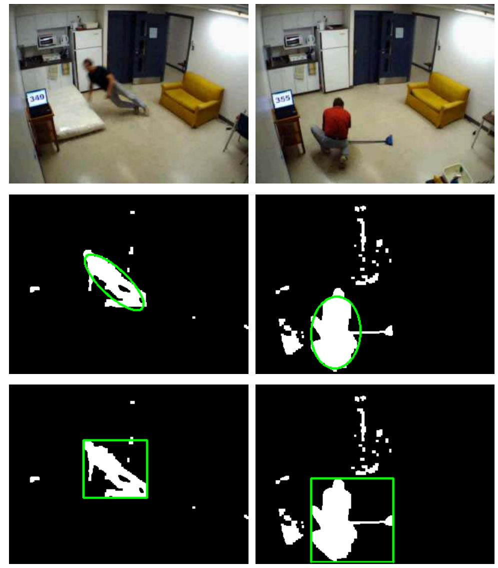
$$I_{min} = \frac{\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2}$$

$$I_{max} = \frac{\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2}$$

Then, the major semi-axis a and the minor semi-axis b of the best fitting ellipse are given by [6]:

$$a = (4/\pi)^{1/4} \left[\frac{(I_{max})^3}{I_{min}} \right]^{1/8} \quad b = (4/\pi)^{1/4} \left[\frac{(I_{min})^3}{I_{max}} \right]^{1/8}$$

With a and b, we can also defined the ratio of the ellipse $\rho = a/b$ (related to its eccentricity).



(a) (b)

Figure 2. Human shape

The approximated ellipse gives us information about the shape and orientation of the person in the image. Figure 2 shows some examples of the approximated ellipse of the person and the corresponding bounding box in special situations like (a) the person falls down and (b) a daily activity. Notice that a mattress is used to protect the person during the simulated fall. This figure demonstrates that the choice of an approximated ellipse is better than a bounding box to analyze the human shape.

4. Fall Recognition System

4.1. Fall Detection System Overview

Our fall detection system is based on the Motion History Image and some changes in the shape of the person. An overview of our fall recognition algorithm is shown in Fig. 3. The three main steps of the algorithm are:

Motion Quantification The quantification of the motion of the person allows to detect large motion like falls. But a large motion can also be a characteristic of a walking person, so we need to analyze further to discriminate a fall from a normal movement.

Analysis of Human Shape An analysis on the moving object is performed to detect a change in the human shape, more precisely in orientation and proportion.

Lack of motion after a fall The second analysis of the moving object is to check if there is a lack of motion just a few seconds after the fall.

In the next subsections, we describe in more details each one of these steps.

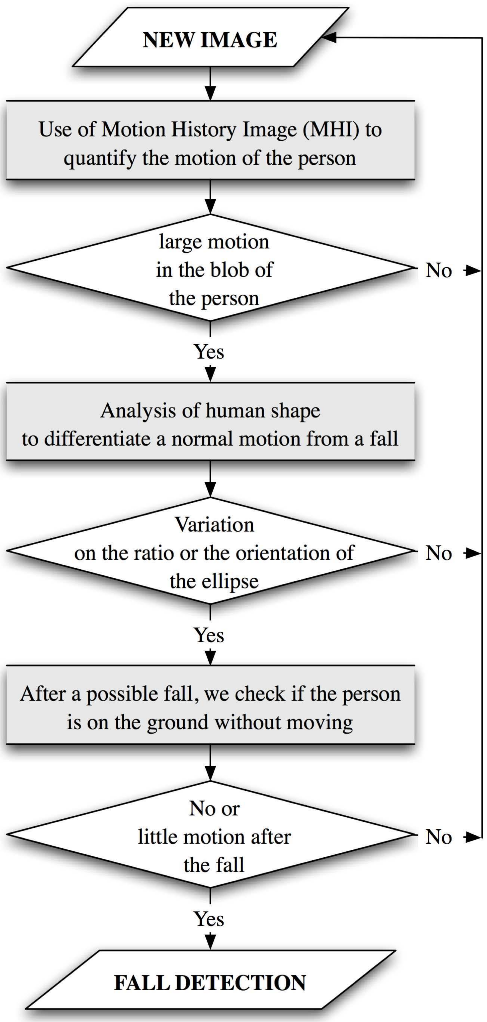


Figure 3. Our fall recognition algorithm

4.2. Motion Quantification

As we want to quantify the motion of the person, we compute a coefficient C_{motion} based on the motion history within the blob representing the person using :

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

With $blob$ being the blob of the person extracted using background subtraction, and H_τ the Motion History Image. Only the largest blob is considered here.

This coefficient is then scaled to a percentage of motion between 0%, no motion, and 100%, full motion.

The duration of a fall is extremely short, typically less than a second. So, we compute the Motion History Image by accumulation of motion during $500ms$. We consider a motion as a possible fall if the coefficient C_{motion} is higher than 65%.

4.3. Analysis of Human Shape

If a large motion is detected ($C_{motion} > 65\%$), we analyze more precisely the change in the human shape to discriminate a fall from another normal activity. For this purpose, we compute two values:

The orientation standard deviation σ_θ of the ellipse If a person falls perpendicularly to the camera optical axis, then the orientation will change significantly and σ_θ will be high. If the person just walks, σ_θ will be low.

The ratio standard deviation σ_ρ of the ellipse If a person falls parallelly to the camera optical axis, then the ratio will change and σ_ρ will be high. If the person just walks, σ_ρ will be low.

σ_θ and σ_ρ are computed for a $1s$ duration. We consider that a large motion is a fall if σ_θ is higher than $15\ degrees$ or if σ_ρ is higher than 0.9. These thresholds on σ_θ and σ_ρ are sufficient to be insensitive to little ellipse variations due to a noisy trajectory (bad image segmentation or motor disorder (variation in human gait, parkinson, etc.)).

4.4. Lack of motion after a fall

A last verification is accomplished by checking if the person is immobile on the ground after a possible fall. As soon as a fall is detected, we look for an unmoving ellipse during the 5 seconds following the fall. If an unmoving ellipse is detected, then we confirm the fall. If the ellipse still continue to move during these 5 seconds, we consider that this can not be a fall.

All the criterions below must be respected to detect an unmoving ellipse :

Few motions in the blob of the person $C_{motion} < 15\%$

An unmoving centroid of the ellipse We compute the standard deviation for $\bar{x}$ and $\bar{y}$, and check for a small deviation. An unmoving centroid is defined by $\sigma_{\bar{x}} < 2\ pixels$ and $\sigma_{\bar{y}} < 2\ pixels$.

An unmoving shape of the ellipse We compute the standard deviation for H_x, H_y and θ (σ_θ is already computed), and check for a small deviation. An unmoving shape is defined by $\sigma_a < 2\ pixels$, $\sigma_b < 2\ pixels$ and $\sigma_\theta < 15\ degrees$.

Some examples of different situations are shown in Fig. 4 with the corresponding coefficients C_{motion}, σ_θ and σ_ρ.

5. Experimental Results

Our system is designed to work with a single uncalibrated camera. As we want a low-cost system, our video sequences were acquired using a USB webcam with a wide

Backward fall: $C_{motion} = 86\%$, $\sigma_\theta = 16.2$, $\sigma_\rho = 0.64$
In this image, the person falls backward without refraining. This type of fall is extremely dangerous, the person can have a serious head injury. In this simulation, a mattress is used to protect the person during the simulated fall. This fall is detected because C_{motion} and σ_θ are high, and no movement occurs after the fall.

Sitting down: $C_{motion} = 67.5\%$, $\sigma_\theta = 11.96$, $\sigma_\rho = 0.59$
In this image, the person sits down brutally, so the motion is large and a possible fall is considered. But σ_θ and σ_ρ are below the fixed thresholds, so no fall is detected.

Normal walk: $C_{motion} = 40.4\%$, $\sigma_\theta = 2.16$, $\sigma_\rho = 0.19$
In this image, the person walks slowly, so no large motion is detected. The algorithm stops at the first step because of the motion lack.

Fast walk: $C_{motion} = 74.2\%$, $\sigma_\theta = 1.29$, $\sigma_\rho = 0.33$
In this image, the person walks quickly, so the motion is high and over the threshold. A possible fall is considered, but σ_θ and σ_ρ are very low which means that it can not be a fall.

Figure 4. Examples on various situations.

angle of more than 70 degrees to see all the room (model Live! Ultra from Creative Technology Ltd). As you will see further down, our method gives good results in spite of the low-quality images (high compression artifacts, noise) of our acquisition system. Our fall detection system is implemented in C++ using the OpenCV library [1].

Our dataset for our experiments is composed of video sequences representing 24 daily normal activities (walking, sitting down, standing up, crouching down) and 17 simulated falls (forward falls, backward falls, falls when inappropriately sitting down, loss of balance).

The table 1 itemizes all the results obtained with our dataset.

	DETECTED	NOT DETECTED
FALLS	True Positive : 15	False Negative : 2
LURES	False Positive : 3	True Negative : 21

Table 1. Recognition results

Figure 5 shows an example of a video sequence of a forward fall. In this sequence, the person walks, sits down in the sofa (frame 93), stands up again (frame 115), walks (see frame 125 above) and falls (see frame 146 above). The confirmation of the fall is done at frame 174 (above) since the ellipse doesn't move anymore.

Our recognition system gives very good results. We get a good rate of fall detection with a sensitivity of 88%, and an acceptable rate of false detection with a specificity of 87.5%.

We choose to detect a maximum number of falls (true positives) by minimizing the threshold for C_{motion}. This is done at the price of a poorer specificity. Indeed, with this low threshold, we sometimes detect a fall when the person brutally sits down because of the high level of motion generated in this case and a changing orientation as shown in Fig. 6. However, to limit this type of false detection, we could define normal inactivity zones [8] [9].

$C_{motion} = 72.95\%$ $\sigma_\theta = 16.13$ $\sigma_\rho = 0.64$

Figure 6. Detection error example

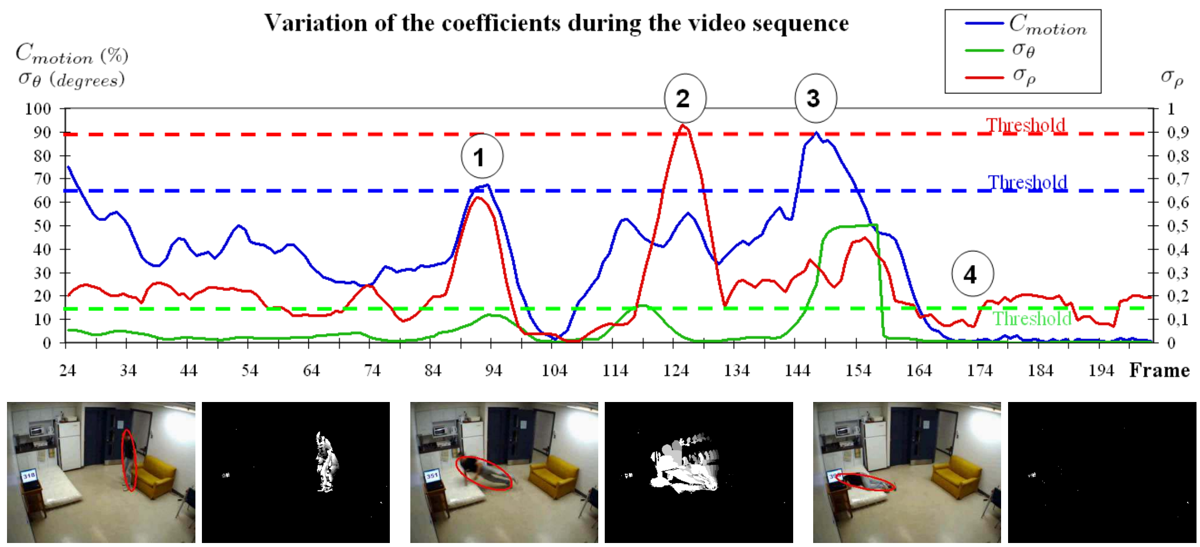
The computational time for an image size of 320x240 pixels is less than $80\ ms$, which is adequate for our application because a frame rate of 10 fps is sufficient to detect a fall. The method is entirely automated and robust in spite of the bad video quality and the fluctuant frame rate of the webcam. Our dataset is also realistic representing various falls in different directions from the camera point of view, but also numerous daily activities occurring in the elderly person's life.

6. Conclusion and Discussion

In this work, a new method to detect elderly person falls is proposed. The combination of motion and change in the human shape gives crucial information on human activities. Our fall detection system has proven its robustness on realistic image sequences of simulated falls and daily activities.

In this work, we make the assumption that the person is on the ground with no or little motion after a fall. This can be argued in some circumstances, for example in the case of injury, where the person could move rapidly because of

21st International Conference on
Advanced Information Networking and Applications Workshops (AINAW'07)
0-7695-2847-3/07 $20.00 © 2007 IEEE



Frame 125: $C_{motion} = 52.9\%$, $\sigma_\theta = 2.78$, $\sigma_\rho = 0.93$

Frame 146: $C_{motion} = 86.54\%$, $\sigma_\theta = 19$, $\sigma_\rho = 0.35$

Frame 174: $C_{motion} = 1.23\%$, $\sigma_\theta = 0.51$, $\sigma_\rho = 0.14$

The person (1) sits down, (2) stands up again and (3) falls. Frame 174, the fall is confirmed because of little motion (4).

Figure 5. Example on a video sequence

the pain. A solution to increase the robustness of our system could be the addition of 3D information to check if the head is near the floor for instance [11]. Another way to improve the system could be the use of the audio information from the microphone of the webcam. A speech recognition algorithm could also be used to identify distress cries like "Help!".

Error detections typically occurs when the person brutally sits down because of large motion and variation in orientation. To limit this type of false detection, we could define normal inactivity zones [8] [9].

The thresholds were chosen manually by logical reasoning on what is a fall and observation of our video sequences. However, an automatic method could be implemented to define the thresholds using a training dataset.

Acknowledgment

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] Intel open source computer vision library http://www.intel.com/research/mrl/research/opencv.

[2] Canada's aging population. Public Health Agency of Canada, Division of Aging and Seniors, 2002.

[3] Report on senior's falls in canada. Public Health Agency of Canada, Division of Aging and Seniors, 2005.

[4] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, February 1994.

[5] A. Bobick and J. Davis. The recognition of human movement using temporal templates. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 23, pages 257–267, March 2001.

[6] A. Jain. *Fundamentals of digital image processing*. Prentice Hall, Englewood Cliffs, New Jersey, 1989.

[7] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*, 11(3):172–185, June 2005.

[8] T. Lee and A. Mihailidis. An intelligent emergency response system: preliminary development and testing of automated fall detection. *Journal of telemedicine and telecare*, 11(4):194–198, 2005.

[9] 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)*, volume 4, pages 323–326, 2004.

[10] W. Pratt. *Digital Image Processing*. 3rd edition, John Wiley & Sons, New York, 2001.

[11] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Monocular 3d head tracking to detect falls of elderly people. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6384–6387, August 2006.

[12] A. Sixsmith and N. Johnson. A smart sensor to detect the falls of the elderly. *IEEE Pervasive Computing*, 3(2):42–47, April-June 2004.

[13] B. Töreyin, Y. Dedeoglu, and A. Çetin. Hmm based falling person detection using both audio and video. In *IEEE International Workshop on Human-Computer Interaction*, Beijing, China, 2005.