How to detect human fall in video? An overview

Jared Willems[1], Glen Debard[2], Bert Bonroy[2], Bart Vanrumste[2], Toon Goedemé[1]

[1]*De Nayer Instituut, Lessius University College,*
Jan De Nayerlaan 5, 2860 Sint-Katelijne-Waver, Belgium
jared.willems@student.denayer.wenk.be, toon.goedeme@denayer.wenk.be

[2]*MOBILAB, Katholieke Hogeschool Kempen,*
Kleinhoefstraat 4, 2440 Geel, Belgium
{glen.debard, bert.bonroy, bart.vanrumste}@khk.be

***Abstract*— Every year, thousands of elderly people are victim of a fall incident. Sometimes with severe consequences such as hip fractures or even death but, in many cases, the main problem is that an injured elderly may be laying on the ground for several hours or even days after a fall incident has occurred. This makes it important to have a fall detection system. Commercial types of fall detection systems are mostly based on wearable sensors, which the elderly may forget to wear. Although we will give a short presentation of these sensor-based devices, this paper focusses on the existing approaches to detect a fall in video. Therefore we have to deal with the different types of background subtraction. After having studied the practical approaches for background subtraction, we went further to the next step in the algorithm, namely fall detection itself. Beside these specific techniques, we also give an overview in difficulties while implementing a fall detection algorithm. In our conclusion we will see that all systems studied in this paper have their own advantages and disadvantages. To become a good video-based fall detection system, a combination of different techniques will be needed.**

I. Introduction

Falling of elderly people is a major health issue. Every year, thousands of elderly people are victim of a fall incident. Chan et al. say that approximately one third of the home-living adults aged 75 or more even fall each year [1]. This makes falling one of the five most common causes of death amongst the elderly population [2]. Falling is also the most important cause of hip fractures in the aged population and has a very high psychological impact on the victims, even if there isn't any injury. To reduce the fear of laying down on the cold floor for several hours or even days, and to overcome serious injuries related to this (e.g. hypothermia), the need of a fall detection and alarm system is obvious. In this paper, we will describe some fall detection principles based on video processing, focussing on systems that can be used in real-life situations.

The organisation of this paper is as follows. Section II will give an overview of existing non-video fall detection systems and the advantages of a camera system. Section III will discuss the difficulties in implementing a video fall detection algorithm. Section IV will handle background subtraction and moving object detection while in section V we see how to detect a fall. A conclusion is formulated in section VI.

II. Fall detection systems

In this section we will discuss the existing fall detection systems and the advantages of a video-based system.

First of all, we have the wearable sensor based devices such as the Zenio system of vitaltronics. Most of those systems make use of accelerometers [4][5], which detect abnormal accelerations and trigger an alarm. An example is the system proposed by Zhang et al. which uses an non-negative matrix factorization method for feature extraction. These systems are quite accurate in detecting falls, but elderly people may forget wearing the device or to recharge the batteries needed for the power supply. It is clear that if the person forgets the sensor, or is reluctant to wear it, no fall can be detected. Another important disadvantage of this kind of system is comfort. It can be very compact, but even then, the elderly may feel uncomfortable wearing the device which discourages the elderly to use it.

Another sensor based fall detection is described in [3]. The system described here makes use of MEMS (Micro Electro-Mechanical System) gyroscope sensors to detect a fall incident based on angular rate. Also a high speed camera set-up was present and synchronised with the MEMS sensor device to analyse the falls, but not to detect the falls.

The systems presented above are automatic alarm generation types. Simple manual systems exist as well, though we should not call them 'fall detection' systems. These devices are operated by the elderly people themselves and do not really detect a fall. The person wearing it can simply push a button on the device to request assistance. This may be a first step to avoid an elderly lying on the ground for several hours but doesn't cope with situations where the victim is unconscious or when the victim isn't able to push the alarm button for any other reason.

As suggested, wearable devices aren't the best offer for fall detection because the risk is too high that the elderly person just forgets to wear it or feels uncomfortable with the system. This is the main reason why we should search further for an optimal, accurate and comfortable fall incident detection system without the need of physical contact with the elderly. From this point of view, a camera-based system could bring the ultimate solution. The main disadvantages of the wearable sensor-based devices are bypassed in a system consisting of a few cameras. There is no need for electronic sensors, attached to the elderly person, to get the necessary information and to monitor his activities. We do not work with electrical signals from sensors to detect a fall, instead we will use advanced computer-based image processing algorithms to detect suspicious (in)activity or fall incidents. In figure 1, a typical basic camera-based fall detection set-up is illustrated.

As can be seen in figure 1, the first step in the algorithm consists of background subtraction to determine the moving objects (people). When the person is detected, we can extract features of it to detect a fall in the fall detection stage. After a fall has occurred and is detected, the need for alarm generation is obvious and will be done in the alarm generation step.

In the next section we will define the difficulties and problems that can arise in developing a fall detection algorithm.

III. Difficulties in implementing a camera system

As we have seen in the previous section, the disadvantages that are present with wearable devices are considerable, but a camera system has its difficulties as well.

First of all, we have the privacy concerns of a camera system [6]. The elderly should be fully aware of the fact that he is filmed - although no-one will look at the video - and it is very important to have permission to install camera's in their home. The elderly should be well informed that the video data isn't viewed by other persons but is only processed on a computer and, except if necessary during the test period, the data will not be recorded or used for any purpose. Especially because the person may be filmed in private situations, such as the bathroom, which is unfortunately a high fall-risk location. As illustrated in fig. 1, there may be the possibility of sending an alarm message to a remote appliance such as a cell phone or PDA, possibly including an image of the fallen person although with respect to the privacy concerns, it will be necessary to process the original image to a binary image before sending. Beside these privacy issues, there are several technical challenges we have to deal with while implementing a video based fall detection algorithm. We will give a brief summary of these aspects [7] in the next paragraph.

A. *Difficulties while developing an algorithm*

- Multi-source artificial lighting whether or not in combination with natural light sources may affect the moving object detection algorithm. Different light situations may cause large, different shadows to appear in the image and thus makes the use of a shadow detection system indispensable [8]. Beside this, colour changes can occur with changing light conditions.
- Houses may contain a lot of objects and pieces of furniture that can be moved. As a background subtraction is used most of the time, it is important that the background image will be updated fast enough so that the changes in the background will disappear as fast as possible in the background reference image.
- People that should be tracked may change their shape (e.g. bending) . Partially- ,or even fully occluding by other people or objects [9] can occur.

While designing a camera-based fall detection system, we do not only cope with difficulties in the algorithm, let's call them 'software considerations'. Another major topic is what we would call 'hardware considerations' which will be discussed in the following paragraph.

B. *Hardware considerations*

- In a research setting, a perfectly working algorithm can be developed in quasi lab circumstances with a lot of processing power at your disposition. On the other side, when the system has to be made commercial, a real-time system is needed and due to cost reductions, a high amount of processing power may be impossible.
- Another cost-related issue is the type of camera used in the video capturing system. If the algorithm can work with a relative low quality video input, the use of cheap webcam-like cameras may be appropriate. If higher quality video input is necessary, more expensive cameras should be used and consequently the overall cost of the system will increase. Other camera systems like omnidirectional ones, used in [10], or infra-red cameras may do a good job as well, but as stated before, the overall system cost will increase and may become too expensive for the elderly or the home care organisation.
- Not only the legal aspect of privacy is paramount, also the feeling of privacy is very important. We should not forget that we put cameras in people's personal environment. The elderly can feel uncomfortable about this, especially when the cameras are obviously present. Therefore, cameras should be placed as discrete as possible and even hidden in an aesthetical ornament.
- Although algorithms that can handle occlusions are described [11,12], it's better to avoid them. This makes the system easier and more accurate. It is probably not possible to avoid all occlusions in a real environment but when camera positions are well-considered, most of them can be avoided.

If these requirements are considered, we can go to the real work in the next chapter which will handle background subtraction and moving object detection.

IV. Background subtraction and moving object detection

As the title of this chapter already suggests, this section will give an overview of used background subtraction algorithms in combination with moving object detection. In fact, most of the background subtraction algorithms can only detect moving objects, as non-moving objects will disappear in the background model.

We see in fig. 1 that background subtraction will follow directly after the video capturing stage. The aim of this step is to extract the person from the video input by subtracting a background estimation model from the original video input so

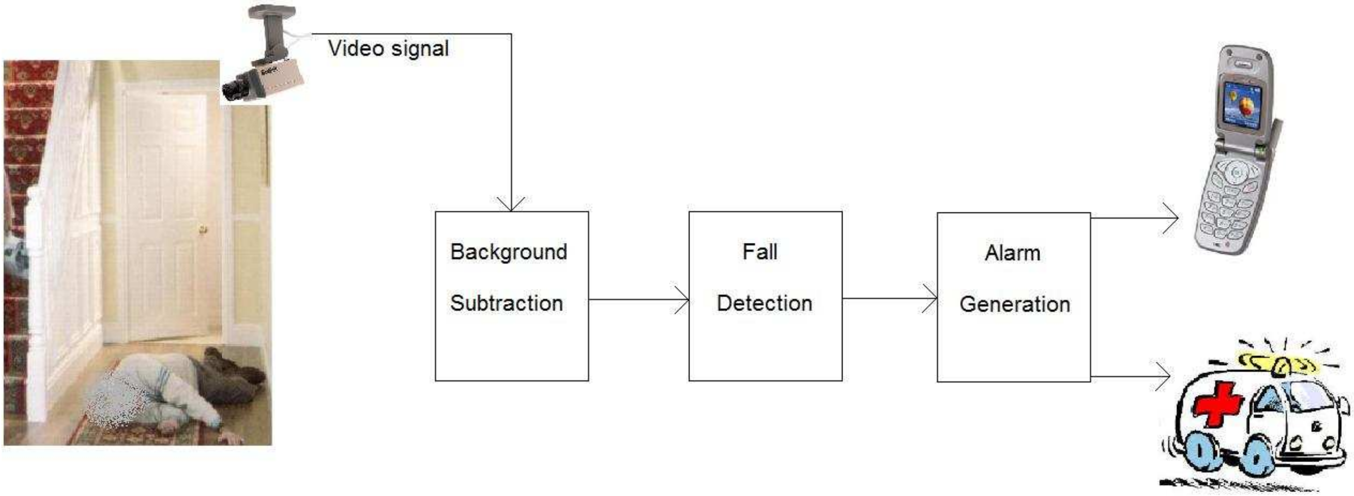

Fig. 1. Camera-based fall detection system

that we can calculate features of this person in the following step, fall detection.

Before we can subtract the estimated background from the current frame, we have to generate this background image in the background modelling stage. We can say that this is the most important stage in the entire background subtraction algorithm. A lot of research in this domain has been established which has led to different new algorithms such as those in [13,14] where they do not use ’simple’ straightforward background subtraction methods. Chen et al. for example make use of a knowledge-based adaptive background update algorithm. Because of the high number of variations in algorithms and the fact that a lot of these new methods are only realistic in labs where you have a lot of processing power to your disposition [15], we will limit us in this chapter by giving a short description of the most practically-used techniques.

We can split background modelling roughly in two main categories, namely recursive and non-recursive techniques. We will first discuss the non recursive techniques and then continue with the recursive methods.

A. Non-recursive techniques

Non-recursive methods use some sort of a sliding window function. It will keep a number (N) of frames in a buffer and depending on the frames in the buffer calculates an estimation for the background model. These techniques are highly adaptive as the model is only determined by the previous N frames and is not affected by frames before these buffered frames. On the other hand, these buffer function will require a significant amount of memory, especially when a large buffer is used [16].

1) Frame differencing: Frame differencing is the most simple background model one can think of. Just take the frame at time t - 1 as background model. This means that the background is simply modelled as the previous frame. To get the foreground (moving object), we should subtract the background model from the actual image and use some thresholding [17].

$$|I_t - I_{t-1}| > T, \tag{1}$$

where I_t is the intensity of frame t and T is a fixed threshold.

The technique is very sensitive to the threshold value which is the only factor that can influence the result.

This type of modelling does not give the most precise result and is very sensitive to noise but can be usable for some situations. An important disadvantage while tracking people is that when someone is not moving for a very short time (one frame period), the person will become part of the background. Main advantages are the small computational load, little memory space needed and its high adaptivity. This means that the background model will be updated very fast after a change in the background.

2) Median filtering: Median filtering Another effective background subtraction approach is the use of median filtering [8]. The estimated background value of each pixel in the background model is calculated as the median of that pixel in all frames in the buffer.

This method can achieve quite good results while not needing too much computational power. Disadvantage is the memory requirement ($N\times$framesize) [17].

3) Linear predictive filter: The current background model estimate is computed by applying a linear predictive filter on the pixels of the frames in the buffer. The filter coefficients are estimated depending on the sample covariances at each frame time [16].

This is one of the techniques that are not usable in real-time systems because of the difficult calculations.

B. Recursive techniques

The difference between recursive and non-recursive techniques is that the recursive types do not use a buffer with previous frames. Instead they update their background image recursively. The advantage is that there should only be one frame stored and this image will be updated everytime a new frame is received. On the other hand, when a fault is introduced in the background image, it will take much longer to disappear. This means that a recursive technique is less adaptive as the non-recursive methods.

1) Running average: A very simple and fast background modelling algorithm without high memory requirements is the running average. This can be computed as:

$$B_{i+1} = \alpha C_i + (1 - \alpha) B_i \tag{2}$$

where B stands for background and C_i is the current frame. α is defined as the learning rate with a typically value of 0.05. As with most of the simple methods, the running average gives not the most accurate results but depending on the application and with some fine-tuning of α, it can be acceptable [17].

As it is a very simple calculation, the running average method is very fast and does not need a lot of memory space. As with most simple background subtraction algorithms, the accuracy gained with this technique is not very high but can be good enough depending on your the application.

2) Approximated median filtering: A quite interesting background modelling technique is approximated median filtering. The algorithm was developed in 1995 by McFarlane and Schofield [18] to track piglets.

It works as follows. There is one background model estimate. When a pixel in the current frame has a grayvalue which is larger then the pixel in the background estimate, than the pixel in the estimate is incremented by one. On the other side, when the value of a pixel in the current frame has a value which is lower than that in the background estimate, the pixel in the background estimate is decremented by one. When applying this function to the background model, the model converges to an estimate where half the input pixels are greater than the background and the other half are less than the background model.

Although the very simple implementation, approximated median filtering may give good results which can even achieve an accuracy near to that of algorithms with a much higher complexity. The memory requirements are low and it is a robust technique. One major drawback is the slow adaptation to big changes in the real background [16].

3) Kalman filtering: This method assumes that the best information of a system state is obtained by an estimation [19]. To make this estimation, several approaches are presented in the literature [16]. Most using the luminance intensity, intensity and its temporal derivative (estimated variety) or intensity and its spatial derivatives. In the most simple variation, we can model the background estimation B(t) as:

$$B(t) = A(t)B(t-1) + K(t)\left[z(t) - H(t)A(t)B(t-1)\right] \tag{3}$$

where $A(t)$ is the system matrix which describes the background dynamics, $H(t)$ is the constant measurement matrix, $z(t)$ is the system input and $K(t)$ is the Kalman gain matrix.

An advantage is that the gain matrix can switch between fast and slow adaptation whether the pixel is a foreground or background pixel. Kalman filtering is known for the disadvantage of leaving long trails behind a moving object.

4) Mixture of Gaussians: Last but not least, the mixture of Gaussians background modelling method. Where Kalman filtering only tracks one Gaussian, mixture of Gaussians tracks usually 3 to 5 Gaussian distributions simultaneously [16]. It is a highly popular technique for background modelling but is considered as computational complex and is very sensitive to sudden changes in illumination Beside this, it scores very well for accuracy and is not very memory consuming. An important difference with a lot of other methods is that mixture of Gaussians do not use one image of values as background model. Instead, each pixel is modelled by a number of Gaussians that will form a probability distribution function F. The main formula for the algorithm is:

$$F(i, \mu) = \sum_{i=1}^{k} \omega_{i,t} \eta(\mu, \sigma) \tag{4}$$

Although the formulas might look quite complicated, the theory behind the method is straightforward. The mean μ of each Gaussian (1 to k), also called components, can be seen as an estimation of the pixel value in the next frame. The weight and the standard deviation σ will give an impression of confidence in the estimation. A comparison between an input pixel and the means of the Gaussians tracking that pixel should be done. The absolute difference between the pixel value and the mean of the Gaussian should be less than the component's standard deviation, scaled by a factor D. If so, the pixel is considered to be part of the background, if not, the pixel will be classified as foreground.

$$|i_i - \mu_{i,t-1}| \leq D\sigma \tag{5}$$

After each frame, the component variables ω, μ and σ have to be updated. Formulas to update these variables are clearly explained in [16].

This is one of the more accurate techniques which can also handle multi-modal backgrounds due to the number of components [15]. Important disadvantages are the computational complexity and the high sensitivity to sudden changes in illumination.

In this chapter we have presented the most commonly-used background subtraction algorithms. Every model has as we can see its own advantages and disadvantages. However the more complex variants may give better results concerning accuracy and robustness, the more simple models require much less processing power and may give satisfying results as well (e.g. approximated median filtering). When background subtraction is dealt with in a fall detection system, the next step is the fall detection itself which is summarized in section five.

V. Fall detection

In this section, we will handle the fall detection stage of the fall detection system. A lot of methods are described in the literature. A detailed discussion covering all these techniques would be beyond the scope of this paper. Instead, we will give a general overview of these techniques.

A first major subdivision can be made between fall detection algorithms which are based on clear immediate visual clues, such as a sudden change in dimensions and on the other hand algorithms that need some sort of specific processing such as hidden Markov models. Another possibility to classify the algorithms is the difference between algorithms which actually detects the fall, where others may detect 'abnormal' behaviour.

One of the most used and most simple techniques to detect a fall is the aspect ratio of the bounding box [21][6]. The bounding box is a rectangular box which can be drawn around the moving object. The aspect ratio is the ratio between the dimensions in x and y direction of the bounding box. When someone falls, the bounding box will change drastically in x and y direction and thus the aspect ratio will change as well [2].

A second method to detect a fall is the use of a fall angle [2]. Fall angle can be defined as the angle between the ground and the person from where it is certain that the person will fall. Although the fall angle may differ from people to people, a good estimate for this angle is $45°$. Note that this method may fail if the person is falling towards the camera.

Some other algorithms make use of the centroid of the falling person. The centroid changes significantly and rapidly during the fall. Some related work even suggest that the fall incident will take approximately $0.4s$ to $0.8s$ [20]. However we believe that a fall incident can not be characterized by a duration interval as there are large differences between falls.

A more reliable detection method is found in [20] where they use vertical projection histograms to detect a fall. The vertical projection histogram of a person will change significantly when a fall occurs [7] and is computed as follows:

$$H(x,y) = \begin{cases} 1 & if \quad (x,y) \quad is \quad an \quad object \quad pixel \\ 0 & otherwise \end{cases} \tag{6}$$

$$V(x) = \sum_y H(x,y) \tag{7}$$

The last simple feature we want to present is the horizontal and vertical gradient [2]. When a person is falling, the vertical gradient will be less than the horizontal gradient.

It is clear that all methods mentioned above do work only in specific circumstances. Therefore, it is necessary to combine a number of these techniques to get a reliable system to detect a fall.

More advanced fall detection algorithms are most of the time based on Hidden Markov Models. The major drawback for HMM's are the need for training data – including data with falling persons – which is necessary for the learning phase of the algorithm. Hidden Markov Models were used for instance in [21] where they first computed the wavelet transform of the one-dimensional aspect ratio and then used this as feature signal for the HMM based classification. Two three-state HMM's were defined to classify the motion of the person, one model for walking and one for falling. Other papers describing the use of Hidden Markov Models are [9] and [6] where they define two models, one with falling sequences and the other with non-falling sequences.

The techniques discussed above do have at least one thing similar, they detect a fall or a specific event. There is also another possibility to generate an alarm when someone has fallen. When someone is laying on the ground for a while [10] or is laying in bed for an abnormal long time, this may be seen as unusual inactivity [22]. This event can trigger an alarm but isn't a real detection of a fall. A disadvantage of these methods is the necessity of declaring inactivity zones. These are zones where the person may lay down for a longer period without activating the alarm system. This creates the need of recalibrating the system every time the furniture is replaced.

In this chapter, we have given an overview of the most interesting fall detection algorithms. As we have seen, there is not one best way in detecting a fall. To become a reliable system, a combination of several methods should be applied.

VI. Conclusion

The goal of this paper was to study the existing fall detection algorithms. Not only the fall detection algorithm on its own but the system set-up was presented. We have seen that the use of low cost cameras is preferable because of cost-related issues and that it should be possible because most background subtraction algorithms don't need high quality video input. The fall detection methods deal only with a number of falls so that it is necessary to implement a combination of several approaches to get a reliable detection system. This combination can exist of a detection step and a confirmation step or only a detection step. Because of the importance of a real-time system, it is better to keep algorithms as simple as possible, while maintaining a sufficient accuracy which is necessary to obtain a reliable result. The systems studied here were tested with simulated videos of falling people and even then none of the systems studied in this paper gave an accuracy of 100% in all circumstances. All described algorithms have their advantages and disadvantages. To get a system that is most performant in real life environments, a combination of these different techniques is needed. This provides a big challenge.

Acknowledgment

The authors greatly acknowledge the financial support of this work by the Institute for the Promotion of Innovation through Science in Flanders (IWT-Vlaanderen), Belgium.

References

[1] B. K. S. Chan, L. M. Marshall, K. M. Winters et al., 'Incident fall risk and physical activity and physical performance among older men', American Journal of Epidemiology, 2006.

[2] V. Vishwakarma, C. Mandal, and S. Sural, 'Automatic detection of human fall in video', Lecture Notes in Computer Science, Pattern Recognition and Machine Intelligence, vol. 4815, pp. 616-623, 2007.

[3] M. N. Nyan, F. E. H. Tay, A. W. Y. Tan et al., 'Distinguishing fall activities from normal activities by angular rate characteristics and high-speed camera characterization', Medical Engineering & Physics, vol. 28, no. 8, pp. 842-849, Oct, 2006.

[4] U. Lindemann, A. Hock, M. Stuber et al., 'Evaluation of a fall detector based on accelerometers: a pilot study', Medical & Biological Engineering & Computing, vol. 43, pp. 548-551, 2005.

[5] T. Zhang, J. Wang, L. Xu et al., 'Using wearable sensor and NMF algorithm to realize ambulatory fall detection', Lecture Notes in Computer Science, Advances in Natural Computation, vol. 4222, pp. 488-491, 2006.
[6] A. Williams, D. Ganesan, and A. Hanson, 'Aging in place: fall detection and localization in a distributed smart camera network', Proceedings of the 15th international conference on Multimedia, pp. 892-901, 2007.
[7] R. Cucchiara, C. Grana, A. Prati et al., 'Computer vision system for in-house video surveillance', IEEE Proceedings Vision Image and Signal Processing, vol. 152, no. 2, pp. 242-249, Apr, 2005.
[8] R. Cucchiara, C. Grana, M. Piccardi et al., 'Detecting moving objects, ghosts, and schadows in video streams', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 10, pp. 1337-1342, Oct, 2003.
[9] R. Cucchiara, A. Prati, and R. Vezzani, 'A multi-camera vision system for fall detection and alarm generation', Expert Systems, vol. 24, no. 5, pp. 334-345, Nov, 2007.
[10] S. J. McKenna, and H. N. Charif, 'Summarising contextual activity and detecting unusual inactivity in a supportive home environment', Pattern Analysis and Applications, vol. 7, no. 4, pp. 386-401, Jun, 2005.
[11] A. Senior, 'Tracking people with probabilistic appearance models', ECCV workshop on Performance Evaluation of Tracking and Surveillance Systems, pp. 48-55, 2002.
[12] T. B. Moeslund, and E. Granum, 'A survey of computer vision-based human motion capture', Computer Vision and Image Understanding, vol. 81, pp. 231-268, 2001.
[13] X. Chen, Z. He, D. Anderson et al., 'Adaptive silhouette extraction and human tracking in complex and dynamic environments', Image Processing, 2006 IEEE International Conference, pp. 561-564, 2006.
[14] A. M. Bagci, Y. Yardimci, and A. E. Cetin, 'Moving object detection using adaptive subband decomposition and fractional lower-order statistics in video sequences', Signal Processing, vol. 82, no. 12, pp. 1941-1947, Dec, 2002.
[15] S. Benton, 'Background subtraction, part 1: Matlab models', http://www.dspdesignline.com/howto/210000460, 2008-10-24.
[16] S.-C. S. Cheung, and C. Kamath, 'Robust techniques for background subtraction in urban traffic video', Proceedings of the SPIE, vol. 5308, pp. 881-892, 2004.
[17] M. Piccardi, 'Background subtraction techniques, a review', Proc. of IEEE SMC 2004 International Conference on Systems, Man and Cybernetics, The Netherlands, Oct, 2004.
[18] N.J.B. McFarlane, and C.P. Schofield, 'Segmentation and tracking of piglets in images', Machine Vision and Applications, vol. 8, no. 3, pp. 187-193, May, 1995.
[19] C. Ridder, O. Munkelt, and H. Kirchner, 'Adaptive background estimation and foreground detection using kalman-filtering', Proceedings of International Conference On Recent Advances in Mechatronics (ICRAM), 193-199, 1995.
[20] C.-W. Lin, and Z.-H. Ling, 'Automatic fall incident detection in compressed video for intelligent homecare', Computer Communications and Networks, ICCCN 2007, pp. 1172-1177, 2007.
[21] B. U. Töreyin, Y. Dedeoğlu, and A. E. Çetin, 'HMM based falling person detection using both audio and video', IEEE international workshop on human-computer interaction 2005, Chine , vol. 3766, pp. 211-220, 2005.
[22] H. Aghajan, J. C. Augusto, C. Wu, et al., 'Distributed vision-based accident management for assisted living', ICOST 2007, pp. 196-205, 2007.