FallCam: Practical Considerations in Implementing a Camera-based Fall Detection System

Glen Debard[1], Jonas Van den Bergh[2],Bert Bonroy[1], Mieke Deschodt[3], Eddy Dejaeger[4], Koen Milisen[3], Toon Goedemé[2], Bart Vanrumste[1]

[1]*K.H. Kempen University College, MOBILAB:Biosciences and Technology Department*
Kleinhoefstraat 4, 2440 Geel, Belgium
{glen.debard, bert.bonroy, bart.vanrumste}@khk.be

[2]*Hogeschool voor Wetenschap & Kunst, Campus De Nayer*
Jan De Nayerlaan 5, 2860 Sint-Katelijne-Waver, Belgium
{jonas.vandenbergh, toon.goedeme}@denayer.wenk.be

[3]*K.U.Leuven, Center for Health Services and Nursing Research*
Kapucijnenvoer 35/4, 3000 Leuven, Belgium
{mieke.deschodt, koen.milisen}@med.kuleuven.be

[4]*U.Z.Leuven*
Herestraat 49, 3000 Leuven
eddy.dejaeger@uz.kuleuven.ac.be

***Abstract*— Many older persons fall and are not able to get up again. The lack of timely aid can lead to severe complications. A camera based fall detection system can provide a solution. This paper gives the current status of our IWT-TETRA project FallCam. The goal of the project is to create a camera based fall detection system. Part of the project is creating an acquisition system for recording simulated as well as real-life falls at the homes of older people. The detection system will be developed and tested using these recordings. Furthermore, we will explain the advantages and disadvantages of different cameras and processing systems. Our choice is a centralized PC platform in combination with two regular IP cameras in each room of the monitored house.**

I. Introduction

Many older persons fall and are not able to get up again. Thirty to forty-five percent of these persons that still live at home and more than half of the elders that are living in a nursing home fall at least once a year. One out of three up to one out of two older persons fall more than once every year [1][2]. Ten to fifteen percent of the ones that fall, suffer severe injuries. The lack of timely aid can even lead to more severe complications. Although not all falls lead to physical injuries, psychological consequences are also important, leading to fear of falling, losing self-confidence and fear of losing independence[1]-[4]. Taking the ongoing aging of the population into account, it is obvious that a manner to detect fall incidents is getting more and more important.

The existing technological detectors are mostly based on wearable sensors. But a market study of SeniorWatch [5] discovered that the sensors are often not worn at all times. This happens for example when a person leaves the house, as the wireless system (connected to the sensor) can't reach the base station. Also during housekeeping tasks the sensors are removed, to prevent false alarms due to the needed sensitivity [6]. The devices are using battery power, so no alarm will be generated if the batteries are depleted. They are also sometimes removed when the person finds them uncomfortable. If a fall occurs at these moments, it will not be detected.

A camera based system can overcome these disadvantages. The purpose of this paper is to describe our ongoing research, via an IWT-TETRA project [7], to the practical implementation of a camera based fall detection system: "FallCam".

The next section describes the goal and the structure of our ongoing research. Section 3 deals with the different kinds of cameras, processing systems and other considerations for building a camera system. It also describes our current approach. Finally section 4 describes the structure of a fall detection system and the changes we plan to introduce.

Fig. 1 FallCam logo

II. Goal and Structure of FallCam

The main goal of the ongoing project is building a prototype of a camera system, that can detect human falls. Also the validation of the system, on the target group of subjects with a higher fall-risk, is one of the goals. The progress of our project can be followed via our homepage: www.fallcam.be [7].

To achieve our main goal, we divided the work into following subdivisions:

1. Build an experimental vision system which can gather video in a real-life environment.

2. Create a dataset with labelled videos, consisting of falls and non-falls (daily handlings and near falls) in a simulated environment and capture images of older people with a high-fall risk. These older people are residence of an elderly home or are accompanied by a partner.
3. Implement an existing algorithm, which will be tested and enhanced using the recorded videos.
4. Implement this enhanced algorithm on a prototype platform to detect falls in real-time. The cost of the system has to be as low as possible without degrading the performance of the detections.
5. To verify this prototype, it will be used for a longer period to monitor the older persons at their home.

The next section will discuss the different approaches to build a camera system that can be used to capture and process the real-life videos as well as the simulated videos.

III. Materials and Methods

Our first sub goal is to build a camera system that is able to monitor a complete home, so that we can gather real-life video in all parts of the home. In case the participating persons have an objection against monitoring all rooms in our acquisition phase, at least rooms with a higher fall-risk should be monitored. To reduce the costs and the development time, we are already focussing on the possibility to use this system also as prototype in the fourth phase of our project. To allow optimizations and error-corrections in the fifth phase, the prototype will have to record the undetected falls and the erroneous alarmed non-falls. To increase the chance to record an actual fall, a minimum of one camera is needed in each room.

This also means that the cost per camera has to be as low as possible without decreasing the quality and detail of the images. This is one of the main considerations that has to be taken into account.

There are a lot of different possibilities to build a camera system. A lot of different types of cameras exist and also several approaches to process the data can be considered. In this section, first we will evaluate the types of cameras, after this we will discuss the different system architectures.

To have an overview of a complete room, we will compare practical choices that can be taken: an omnidirectional camera, a thermal infrared camera or several wide-angle cameras. It is also possible to use regular cameras with a smaller view-angle (around 45 degrees), but to cover a complete room, quite a number of cameras is needed in this case. So this path is not taken into consideration for the moment.

A. *Omnidirectional Cameras*

A camera normally has a field of view that ranges from a few degrees to, at most, 180 degrees. This means that it captures, at most, light falling onto the camera focal point through a semi-sphere. In contrast, an omnidirectional camera captures light from all directions falling onto the focal point, covering a full sphere. There are 3 main categories of omnidirectional cameras: the fish eye lens, the catadioptric camera and the combined spherical lens.

The fish eye lens has a very short focal length that enables the camera to view objects within as much as a hemisphere. The advantage of this type of lens is the rather compact size. The biggest disadvantage is the large distortion of the image. Nayar states in [9] that the acquired image does not permit the construction of a distortion-free perspective image of the scene. The higher complexity of the lens also increases the price.

A Catadioptric system uses a spherical mirror in combination with a camera. The shape of the mirror defines the distortion that is viewed on the image. Yamazawa *et al*. propose in [10] the usage of a hyperboloidal mirror. This approach makes it possible to correctly describe the distortion mathematically. Using this, it is possible to convert the image to a perspective view. The draw-backs of this type of camera are the size, the price and also that there is a death-spot directly below the camera.

Another approach that can be taken for an omnidirectional camera is the combination of several spherical cameras in one. The camera itself calculates the complete image out of the different images of the cameras. This type of cameras provide a good quality image with a low distortion. They are however quite expensive, and the unwrapped image is more difficult to process.

An omnidirectional camera has the advantage that only one is required in each room, however there are also some problems that should be considered. The camera has to be attached to the ceiling in the middle of the room. Often also the chamber lighting is situated there. This can in some cases obstruct the sight, or cause over-illumination in case the light is directed to the camera. In case of a fall right next to an obstruction, like a table, the person can be completely occluded by the obstacle. Fig. 2 shows an image taken with an omnidirectional camera. It also shows pictures of the different kinds of omnidirectional cameras.

Fig. 2 Top: image of an omnidirectional camera (unwarped). Bottom: Three types of omnidirectional cameras: left: fish-eye lens, middle: Catadioptric camera, right: Combined spherical camera

B. Thermal Infrared camera

A thermal infrared (IR) camera, like the Gobi-384 of XenICs [11] uses the infrared spectrum for imaging. Every human produces a certain amount of heat. This heat generates infrared radiation, so the image is not based on reflection of light. With a thermal infrared camera, differences in temperature can be seen, this makes it also possible to detect humans in an image. An infrared camera does not have problems with shadows and different light conditions, but interference of other heat sources is a problem. Also occlusions present a problem with this type of camera. The thermal infrared cameras are quite expensive, but in the scope of the study, it is interesting to test the performance of this type of camera in a simulated environment, to see what we can expect from this in the future.

C. Classic Camera

Another type of camera that we can use is the traditional camera. The biggest advantage of this type of camera, is that there is a whole range of specifications that you can choose from. Some of the most important specifications are the lens, the interface, image quality and video compression.

1) Lens: The most important feature of the lens is the viewing angle. By using a horizontal viewing angle of 90 degrees, it is possible to view almost the complete room, if you install the camera in a corner. A camera with a changeable lens is useful to be able to replace the lens depending of the location.

2) Interface: The interface is the means to transport the video stream to the processing unit. Commercial cameras are equipped with USB2.0, IEEE 1394 (Firewire) or Internet Protocol (IP) connections. For USB and Firewire, an adjacent processing unit is needed An IP cam can easily be connected to a network to stream the image to a central server. Also an analog transmission is possible, but for analog wired connections additional capture devices are needed to convert the images to a digital format. Using an analog wireless connection is not advisable for the security of the video signal.

3) Image Quality: To be able to process the video in real-time, the necessary resolution of the image is quite low, around 320x240 pixels, which all cameras can achieve. It would be better to have a higher resolution, but with a doubling of the resolution, the data that has to be processed quadruples. It is desired that the resolution is configurable, to test the different settings while designing the algorithm, since at this stage the processing is not in real-time. The amount of noise generated should be as low as possible. It is also very useful that the camera is able to work with low noise in an environment with low illumination. An IR light modus could be used in case of complete darkness.

4) Video Compression: Unfortunately the bandwidth needed to transmit raw data is rather high. For video with resolution 320x240, 8 bit for each colour and 15 fps, the needed bandwidth is around 27 Mbit/s. Since the theoretical bandwidth of the network, that we would like to use, is only 150 Mbits/s, a compression is definitely needed to allow transmission of multiple video-streams at the same time. We found few papers in the literature regarding the influence, that video compression has on various image processing algorithms. Delac *et al.* state in [12] that for face recognition the usage of Joint Photographic Experts Group (JPEG) compression only has an influence in case of a high compression rate. In video it is possible to make usage of Motion JPEG (M-JPEG) compression, which is basically a sequence of JPEG images. New compression techniques such as used in Moving Picture Experts Group 4 (MPEG-4) are expected to have only a small impact on various algorithms. To be able to choose freely between M-JPEG or MPEG-4 and the associated compression ratio in the camera software, is a valuable option.

On the prototype, also the actual fall detection algorithm has to be computed. Also for this part, different system architectures can be used: decentralized processing, centralized and a hybrid approach.

D. Decentralized Processing

The approach is to connect each camera with its own embedded processing unit. In this case there is no communication between the different units, it will do the processing of the fall detection completely on its own. No network is needed, which makes it possible to work with uncompressed data. Since the embedded processing unit has to do all the handling of the data on its own, his processing power has to be sufficient. The fall detection can't be verified with the data of another camera, this increases the risk of false alarms. Another problem is that the recorded data is distributed over all the devices. It should be possible to send an alarm to get help via e.g. Short Message Service (SMS), in case of a complete decentralized system architecture, this system has to be available on each unit.

The processing can be done via an embedded PC, or via a Digital Signal Processor (DSP). An embedded pc is mostly based on the X86 platform. This is the same architecture like is used on most desktop PC's.

Since an embedded PC uses the same architecture as a server based solution, it is easier to switch to code from centralized to decentralized processing and the other way around. The existence of a standard development platform for PC, decreases the development time. The problems with this approach is its size and higher energy consumption, certainly if you take into account the number of cameras that are needed to monitor a complete home. Also the price is rather high to be used in a commercial product.

Another option is to use a DSP, optimized for this kind of processing. The circuit board can be small and inexpensive, if bought in large quantities. The needed development boards are however very expensive. The low power consumption is an advantage. Writing optimal code for a DSP is rather difficult, so the development time is longer. The limited amount of memory that is available, has to be taken into account while developing the detection algorithms. The main disadvantage of a DSP is the need to develop code for a specific DSP type. As we want to develop a standard platform that can be used by different suppliers forces us to avoid the use of DSPs even with its advantages.

E. Centralized Processing

In centralized processing, all the processing is done on one central unit, which mostly will be a PC. Every video stream is continuously sent to the central unit. All calculations are done on this unit, which facilitates the usage of data from more than one viewpoint. It also gives the possibility to track the person from one camera to another. This is also useful in detecting and avoiding occlusions, since a comparison between different views can be done. Since the video streams of all cameras are constantly sent to the server, a network with a high bandwidth and a fast computer is needed to handle this amount of data. The export of the recorded videos is easier, since it is stored on one central place.

F. Hybrid solution

This is a combination of decentralized and the centralized system architectures. Basic processing like for example motion detection can be done close to (or on) the camera. This way the video will only be streamed when it is really necessary. It is also possible to perform already a basic fall detection on the decentralized processor, in this case the central processing unit will only be used to combine and verify the results for all cameras. The hybrid solution makes it possible to use a cheaper central processing unit. Furthermore the risk for network congestion is less than with a complete centralized approach.

Most IP cameras have motion detection built in, so a separate processing unit is often not needed for a basic processing.

G. Other Considerations

There are still a lot of other considerations that have to be taken into account while building a camera system. These include for example privacy, data network and power consumption of the whole system.

The privacy of the test persons is a very important part that has to be considered. In the first acquisition phase, a lot of data will be stored. For privacy reasons, only 48 hours of data will be available on the storage device for each camera. User input by e.g. a button to stop the data acquisition in case the monitored person or a visitor wants to, has to be provided. Since only 48 hours of data is available on the storage device, the option to stop the deletion of the data in case of a fall has to be provided. This way we have the time to retrieve the important video of this fall.

An entire house or most of it will be monitored, so a network to transport the video streams to the central unit is required. Installing a wired network in all rooms of the monitored house would require a lot of cables. Even holes in walls have to be drilled. Most people don't want these structural changes, so another option has to be chosen. The open options are wireless transmission or transmission over the power grid. A lot of people think of wifi as an insecure network, but with the present advanced encryption methods this is not the case. To have a higher mental feeling of security, Ethernet over power (Homeplug[13]) is a good alternative. Both networks use Advanced Encryption Standard 128 bit (AES128) [14], which is presently almost impossible to hack. The network only has a limited capacity, so compression of the video streams is needed.

Nowadays, the power consumption of all these devices gets more important. Certainly for our detection prototype, the power consumption has to be taken into account. But also already in the acquisition phase, this should be considered. Running a server computer with for example 8 cameras needs over 800 Watt continuously, 600 Watt for the server plus 8 times 25 Watt for the cameras. This would represent a significant extra cost.

H. Our Chosen Acquisition System

Like described above, we need a camera system both for recording real-life data and for the simulated data. The approach that we want to follow is using a hybrid system. We are going to use IP-based network cameras, connected with Ethernet over Power. For data-storage and the detection algorithm, a standard desktop PC will be used.

We have chosen 2 different IP cameras. The first type is rather cheap, the Allnet ALL2292 IP-Cam [15]. It does not have a changeable lens, but the one attached, supports a 80 degrees horizontal viewing angle. The second one, the EyeView CMI-H260 IP-CAM [16], is a basic midrange camera that does have a changeable lens. In this way we can compare the different quality of recordings and more objectively chose the one that can be used for the real-life recordings. We will also test the built-in motion detection of these cameras. If the detection is not sufficient for our goals, then we will move to a complete centralized system in the data-acquisition phase. Each room is normally monitored with at least two cameras.

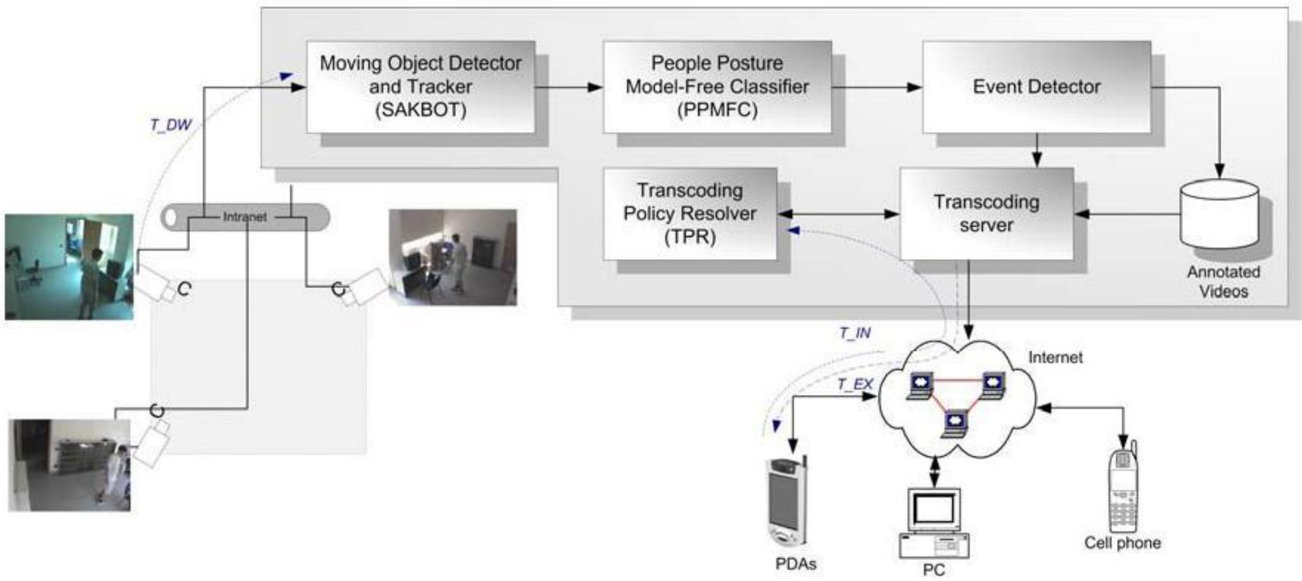

Fig. 3 Overall scheme of the fall detection system proposed by Cucchiara *et al.*

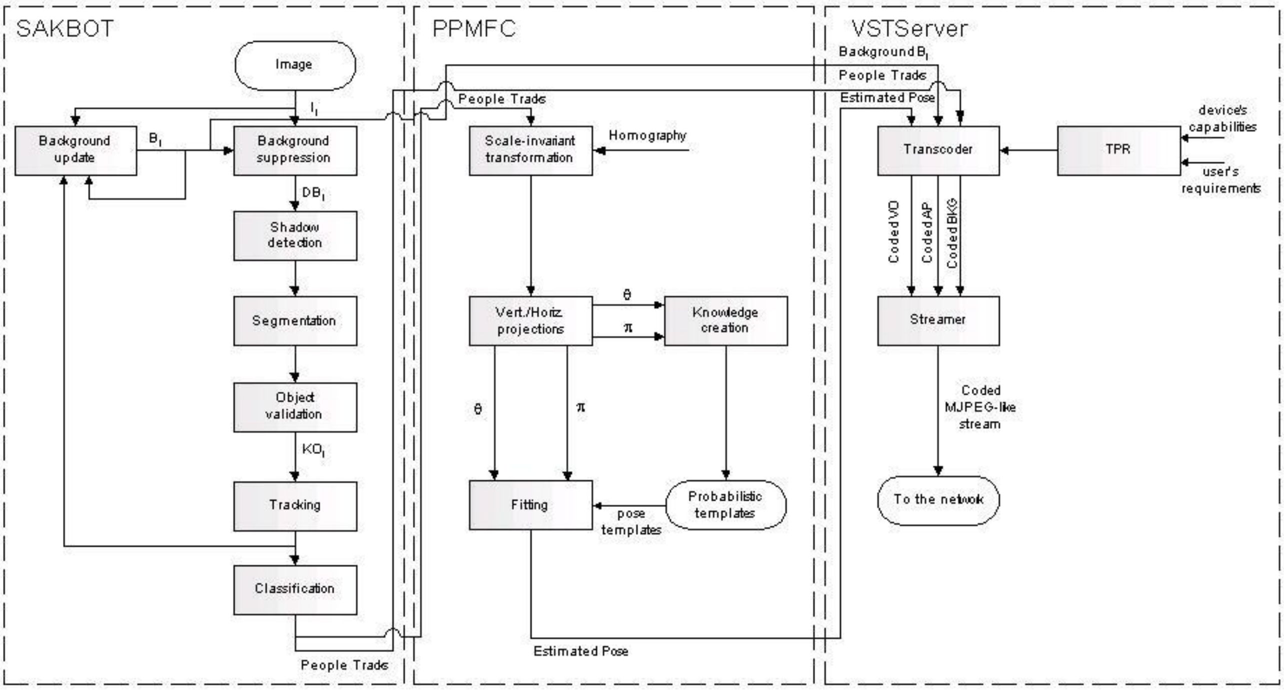

Fig. 4 Details on the three main modules of the fall detection system proposed by Cucchiara *et al*.

The data storage and processing will be done on a regular desktop PC. To get a higher reliability, a mirrored Redundant Array of Independent Disks (RAID) system will be used to store the data. The reason for not using a server PC are mainly cost-efficiency. The regular pc is cheaper both in purchase and in running costs. The server pc is made for speed and reliability, energy efficiency is lower on the requirements list. This will limit us in processing power, since only one processor will be available, but the system resembles a final fall detection system closer.

The network will be based on the HomePlug AV standard. It has a theoretical bandwidth of 150 Mbps. We will test the real throughput to check how many cameras can stream together. If needed, different standards will have to be combined.

The next section will give a small overview of a fall detection system.

IV. Fall Detection System

The main goal of our research is off course the building of a complete fall detection system. Cucchiara *et al.(2007)* give a good overview of such a system in “A Multi-Camera Vision System for Fall Detection and Alarm Generation” [17]. In Fig. 3, a general overview of their system is shown. Fig. 4 shows the main modules more in detail. Their approach exploits computer vision techniques to detect and track people inside a single room with a single camera. A method based on Hidden Markov Models (HMM) is used for classifying the posture and detecting falls. Additionally, multiple cameras are used to cover different rooms and the camera handoff is treated by warping the person's appearance in the new view by means of homography [18]. After detecting a fall, an alarm is generated and a SMS message is sent to a Personal Digital Assistant (PDA) with the link for live video streaming. We will explain a bit more in detail how this system works.

The single camera module, named Statistical And Knowledge-Based ObjecT detection (SAKBOT) and described in detail in [19], extracts moving objects from each camera by exploiting background suppression with selective and adaptive update in order to quickly react to the changes and to also take "ghosts" (i.e., aura left behind by an object that begins to move) into account. After the object extraction, a sophisticated tracking algorithm is used to cope with occlusions and split/merge of objects. A probabilistic and appearance-based tracking is used to handle objects with non rigid motion, variable shape (like people), and frequent occlusions. This tracking algorithm maintains, in addition to the current blob, the appearance image (or temporal template) and the probability mask of the track. The appearance image is obtained with a temporal integration of the colour images of the blobs, while the probability mask associates to each point of the map a probability value that indicates its reliability. Comparing the current blob with the appearance image of the tracks allows the detection of occlusions. Finally, tracks that satisfy some geometrical and colour constraints are classified as people and submitted to the posture classifier. Four main postures are considered: standing, crawling, sitting, and lying. To this aim, they exploit a classifier based on the projection histograms computed over the blobs of the segmented people.

The probabilistic tracking is able to handle occlusions and segmentation errors in the single camera module. However, the strong hypothesis to be robust to occlusions is, that the track has been seen for some frames without occlusions in order for the appearance model to be correctly initialized. This hypothesis is erroneous in the case the track is occluded since its creation. Cucchiara *et al.(2007)* propose to exploit the appearance information from another camera (where the track is not occluded) to solve this problem. If a person passes between two monitored rooms, it is possible to keep the temporal information stored into its track extracted from the first room and use them to initialize the corresponding track in the second room.

The last module of their system is a transcoding video server. To only need a low bandwidth to transmit the video to check the alarm, the image is divided in different sections. The interesting sections have a higher quality setting. In this case the region where a fall is detected, provides of a good quality to allow the verification of the fall.

We will base our fall detection system on this approach. We use multiple cameras per room, to assure that there are no blind spots.

V. CONCLUSION

Fall detection is becoming more and more important to ease the fears of an older person or someone with a higher fall risk. In this way these persons are able to live independently longer and also more comfortable. In this paper, we have described how we will create a camera-system that works in real-life. First, the structure of our project was explained. In the third section, we summarized the different possible cameras and different systems to process the data. Both the advantages and disadvantages were described and then used to distil the approach that we are taking in our acquisition system, a centralized PC platform in combination with two regular IP cameras in each room of the monitored house. The last part that was discussed is an overview of an existing fall detection system. We are now at the point that we are going to collect simulated data as well as real-life data. In conjunction with this, we will start to implement the different needed algorithms. These will be verified with the acquired data, both the simulated as the real-life data. At the end of the project we will verify the real prototype at the homes of real elderly.

ACKNOWLEDGMENT

The authors greatly acknowledge the financial support of this work by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen), Belgium. The aim of the TETRA fund is to stimulate Technology Transfer to a large group of small and medium enterprises.

REFERENCES

[1] K. Milisen, E. Detroch, K. Bellens, D. Dierckx, W. Smeulders, S. Teughels, E. Dejaeger, S. Boonen, W. Pelemans, "Valincidenten bij thuiswonende ouderen: een pilotstudie naar prevalentie, omstandigheden en gevolgen in Vlaanderen," *Tijdschr Gerontol en Geriatr*, vol.35, no.1, pp.15-20, 2004.

[2] M.E. Tinetti, "Preventing falls in elderly persons," *N Engl J Med*, vol.348, pp.42-49, 2003.

[3] T. Masud and R.O. Morris, "Epidemiology of Falls," *Age Ageing*, vol.30, no.4, pp.3-7, 2001.

[4] P.A. Stalenhoef, J.P.M. Diederiks, J.A. Knottnerus, A.D.M. Kester, H.F.J.M. Crebolder, "A Risk Model for the Prediction of Recurrent Falls in Community-Dwelling Elderly: a Prospective Cohort Study," *J Clin Epidemiol*, vol.55, pp.1088-1094, 2002.

[5] SeniorWatch. Fall detector. Technical report, Case Study of European IST Seniorwatch Project, IST-1999-29086, 2001.

[6] K. Doughty, "Fall prevention and management strategies based on intelligent detection, monitoring and assessment,", in *New Technologies in Medicine for the Elderly*, Charing Cross Hospital, London, November 2000.

[7] IWT homepage [Online]. Available: http://www.iwt.be/iwt_engels/

[8] FallCam homepage [Online]. Available: http://www.fallcam.be

[9] S.K. Nayar, "Catadioptric Omnidirectional Camera", Proc. Of IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, June 1997.

[10] K. Yamazawa, Y. Yagi, M. Yachida, "Obstacle detection with omnidirectional image sensor HyperOmniVision", Robotics and Automation, 1995. Proceedings., May 1995

[11] XenICs Goby-384 [Online]. Available: http://www.xenics.com/content/view/207/490/

[12] K. Delac, M. Grgic and S. Grgic, "Image Compression Effects in Face Recognition Systems" , Face Recognition, Book edited by: K. Delac and M. Grgic, ISBN 978-3-902613-03-5, pp.558, I-Tech, Vienna, Austria, June 2007

[13] HomePlug Power Alliance website. [Online]. Available: http://www.homeplug.org
[14] Advanced Encryption Standard (AES) Questions and Answers website. [Online]. Available: http://www.nist.gov/public_affairs/releases/aesq&a.htm
[15] Allnet ALL2292 IP-Cam datasheet. [Online]. Available: http://www.allnet-italia.it/download/ALL2292.pdf
[16] EyeView CMI-H260 IP-CAM datasheet. [Online]. Available: www.eyeview.com.tw/updf/catalog/MPEG4%20IP%20BOX%20CAMERA.pdf
[17] R. Cucchiara, A. Prati, R. Vezzani,, "A multi-camera vision system for fall detection and alarm generation," *Expert Systems*, vol. 24, no. 5, pp. 334-345, November 2007.
[18] E. Vincent, R. Laganiere, “Detecting Planar Homographies in an Image Pair”, Image and Signal Processing and Analysis, 2001. Proceedings., 2001
[19] R. Cucchiara, C. Granao, M. Picardi, A. Prati, "Detecting Moving Objects, Ghosts, and Shadows in Video Streams," *IEEE Trans Pattern Anal Mach Intell*,, vol. 25, no. 10, October 2003.
[20] J. Willems, “How to detect human fall in video? An overview," unpublished.