Event-driven Stereo Vision for Fall Detection

Ahmed Nabil Belbachir, *Member IEEE*, Stephan Schraml, Aneta Nowakowska
New Sensor Technologies, Safety & Security Department, AIT Austrian Institute of Technology
Donau-City Strasse 1/5, A-1220, Vienna Austria.
{ nabil.belbachir ; stephan.schraml ; aneta.nowakowska}@ait.ac.at

Abstract

This paper presents a system concept for efficient fall detection in real-time for elderly security in ambient assisted living applications. Event-driven sensors are biologically-inspired and autonomously reacting to scene dynamics and generating events upon relative light intensity change. Their wide dynamic range and high temporal resolution properties enable efficient activity monitoring in natural environment. Using a stereo pair of event-driven sensor chip, it is possible to represent the scene dynamics in a 3D volume at high temporal resolution. Therefore, the person's activity in a home environment can be efficiently recorded, with a low data volume and high temporal resolution that allows efficient incident detection, like person's falls. In this paper, a dataset with scenarios including 68 person's falls has been analyzed for real-time detection with event-driven stereo vision systems and the results are promising.

1. Introduction

Assisted living systems can be subdivided in non-vision and vision systems. RFID tags [13] and accelerometers [8] are examples of non-vision systems. They allow the detection of activities (walking, running, walking stairs), to count steps, to estimate the distance walked [4] and to detect falls [14][15]. Their main disadvantage is their contact with the body of elderly persons as they have to wear sensors and tags, which can be forgotten or lost easily. Furthermore, tags are usually taken off by the supported person during activities such as taking a bath or a shower, where on the other hand the probability for incidents like falls is high.

Vision systems mainly consist of cameras performing visual surveillance and the detection of activities. Besides the privacy issue, regular camera systems provide insufficient spatial information in locating an object in a room out of a sequence of frames. Stereo vision systems have the advantage to provide information on the distance between object and camera and thus, 3D positions of objects can be calculated. However, correct spatial information relies on the automatic matching between corresponding pixels in each image. This process is computationally expensive; moreover, it is not always reliable. For instance, pixels in low texture areas are very hard to match.

The biologically-inspired (neuromorphic) dynamic vision sensors [5] feature massively parallel pre-processing of the visual information in on-chip analogue circuits and stand out for their excellent temporal resolution, wide dynamic range and low power consumption. These vision sensors are event-driven and have the property to be less sensitive to illumination conditions than traditional frame-based sensors as well as they protect privacy to a certain extent. Furthermore, these sensors involve a drastic reduction of the data volume, compared to frame-based sensors and efficiently capture scene dynamics [1][3][6]. Therefore, stereo vision can be performed very efficiently and at low-cost with the neuromorphic dynamic vision sensor.

This paper presents a compact and low-cost stereo vision system for easy deployment and intelligent monitoring and fall detection in ambient assisted living applications. This system integrates the neuromorphic vision technology with an embedded processing unit and communication technology. An analysis of the asynchronously generated events for person's activity including incidents (like fall) has been performed for real time detection. The events of the stereo vision sensor are represented in a spatiotemporal domain. The fall detection method and its implementation on the Blackfin BF 537 from Analog Device are analyzed for real-time indoor monitoring scenarios towards a compact remote stand-alone 3D vision system. The paper is structured as follows: Section 2 provides a brief review of the architecture of the event-based 3D vision system including core algorithms. Examples of real recording of falls in a lab environment using the Dynamic Vision Sensor (DVS) system are shown in section 3. The analysis of falls and detection method are discussed in section 4 including evaluation results. A summary is provided in section 5 to conclude the paper.

2. Dynamic Stereo Vision Sensor

This section briefly describes the existing dynamic stereo vision sensor reported in [2][10][12] including data examples generated by the system. The system, including the sensor board, DVS chip and DSP board, is depicted in Figure 1. It includes two DVSs as sensing elements [5], a buffer unit consisting of a multiplexer (MUX) and First-In First-Out (FIFO) memory, and a digital signal processor (DSP) as processing unit.

This DVS consists of an array of 128x128 pixels, built in a standard 0.35µm CMOS-technology. The array elements (pixels) respond to relative light intensity changes by instantaneously sending their address, i.e. their position in the pixel matrix, asynchronously over a shared 15 bit bus to a receiver using a "request-acknowledge" 2-phase handshake.

Such address-events (AEs) generated by the sensors arrive first at the multiplexer unit. Subsequently, they are forwarded to the DSP over a FIFO. The DSP attaches to each AE a timestamp at a resolution of 1ms. The combined data (AEs and timestamps) are used as input stream for 3D map generation and subsequent processing.

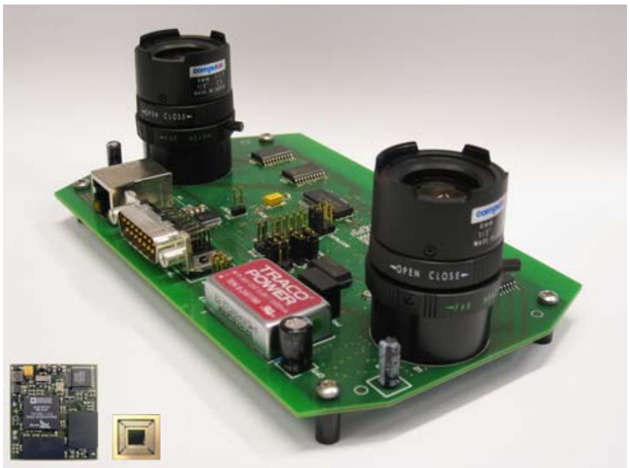


Figure 1: Image of the event-driven stereo vision system. In the lower left corner the DSP Bf537 and the sensor chip are shown. The DSP is mounted on the back of the board.

Figure 2 depicts a space-time representation of one DVS' data, resulting from a two persons crossing the sensor field of view in a room-like environment. The events are represented in a 3 D volume with the coordinates x (0:127), y (0:127) and t (last elapsed ms), the so-called space-time representation. The bold colored dots represents the events generated in the recent 16 ms. The blue and red dots represent spike activity generated by a sensed light-intensity increase (ON-event) and decrease (OFF-event) resulting from the person motions, respectively. The small gray dots are the events generated in the elapsed 2 seconds prior to the recent 16ms. These highlight the event path in the past 2 sec of the moving persons, which is an ideal basis for continuous monitoring by simultaneous detection and tracking in space and time.

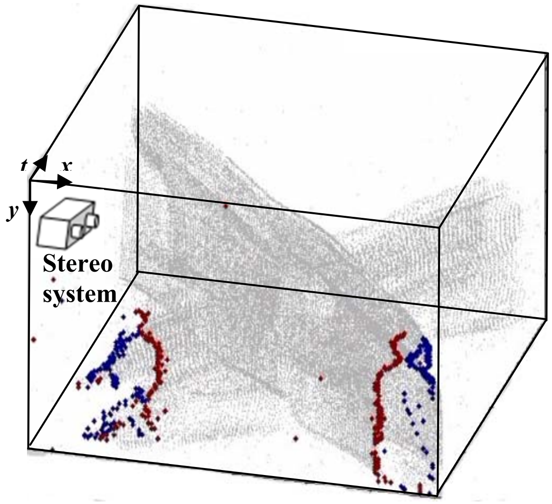


Figure 2: Event representation of scene dynamics (2 persons crossing the field of view) in a space-time domain using 1 DVS. The data are shown in a room-like representation with sensor mounted on the side wall.

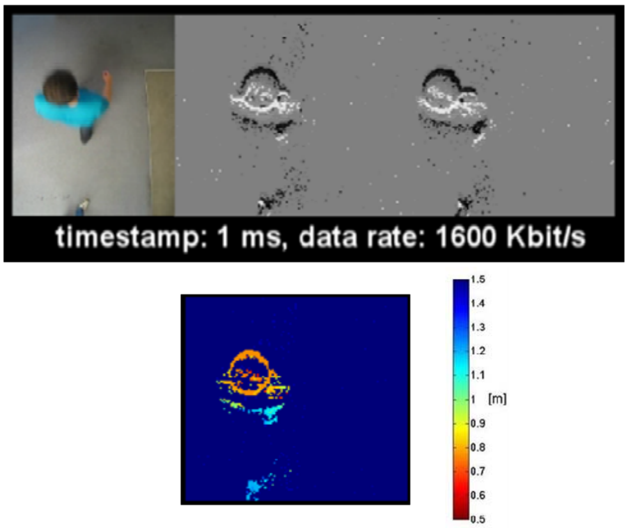


Figure 3: Still image of a person from a conventional video camera (top left); the corresponding events of a pair of dynamic vision sensors (top middle and right) and resulting event "sparse" depth map (bottom) rendered in an image-like representation.

A description of the algorithm for real-time depth estimation is given in [10]. Figure 3 shows an example of a visual scene imaged by a conventional video camera (top left) and its corresponding AEs using a pair of DVSs (top middle and top right) rendered in an image-like representation. The white and black pixels represent spike

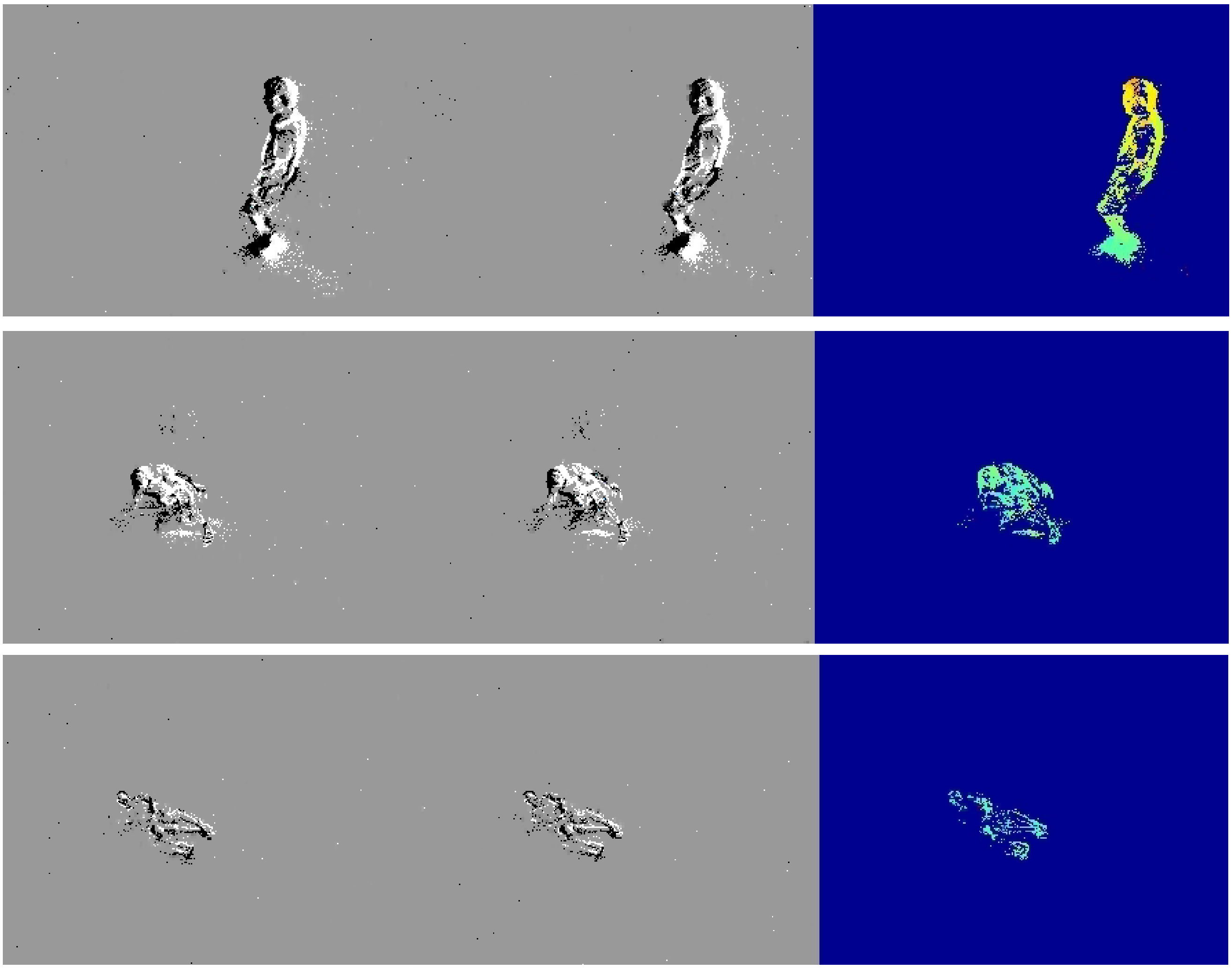
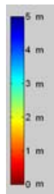


Figure 4: Events rendered in an image-like representation from an example of a person activity and fall. The events are represented for the pair of DVS detector (middle and left) and with the stereo color-coded stereo reconstruction (right). The example shows the person entering the sensor field of view (top) and falling afterwards(middle and bottom)



activity generated by a sensed light-intensity increase (ON-event) and decrease (OFF-event) resulting from one persons motions, respectively. The gray background represents regions with no activity in the scene. The non-moving parts in the scene do not generate any data. The processing unit (DSP) embeds event-based stereo vision algorithms, including the depth generation or the so-called sparse depth map. The resulting sparse color-coded depth map of the scene depicted in Figure 3(left) is provided at the bottom in Figure 3.

3. Example of Fall Recordings

The targeted stereo vision system for fall detection will use the dynamic vision sensor chips with resolution 304 x 256 [9]. The final stereo vision system is still under development and will use as well the DSP BF537 for real-time fall detection. Using the preliminary version of system, it was possible to record more of 110 activity scenarios with about 68 falls. Figure 4 shows one scenario with a person entering a room and falling down afterwards after stumbling across something. The person was lying for about 3 sec on the floor. The events are rendered in an image-like representation and shown for both sensor chips as well as including the color coded depth information. The depth is provided with respect to the distance between the sensor (0: dark red; 5m: dark blue). The algorithm for the stereo reconstruction is described in [10]. These events with the depth information are using for tracking the person position at home and detection fall incidents.

4. Fall Analysis and Evaluation Results

For the analysis of person motion and detection of possible incidents like falls, three steps are performed. First the spatiotemporal analysis of the person activity including the fall is performed. Further, the adequate parameters featuring the fall aspect are extracted. Finally, an intuitive method for real-time fall detection is developed, implemented in the Blackfin DSP and evaluated for the fall detection

4.1. Fall Analysis

In order to provide an adequate analysis for the person activity before, during and after the fall, we plotted in Figure 5 a segment of 10 sec scenario. It shows spatiotemporally generated events upon person motion in the scene. The bold colored events were generated in the last 16 ms while events history within 10 s is shown in small grey dots prior to the last 16ms. The dashed cube embodies the inactivity period after the fall incident. This period shows a drastically reduced number of events compared to the previous time.

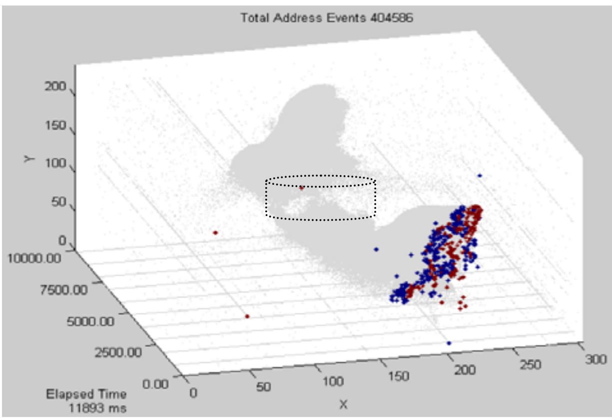


Figure 5: Continuously generated events for 10 sec person's motion including 3 sec inactivity (dashed cube) after a fall

4.2. Feature Analysis

For fall detection two intuitive features seem to be relevant: the person position and activity magnitude. The person position can be identified using the depth map, to find out if the person is lying on the floor. The activity magnitude can be determined by calculating the event rate per sec using the number of events generated by the sensor upon person motion.

Figure 6 depicts the normalized height of the gravity center for the events generated upon person activity in the room. These data was generated by the sensor along 11 sec during the person fall. The red line at 5.5 sec illustrates the start of person fall and show the drastic decrease of the height. Between 8 and 10.5 sec, the height is at the bottom, while the person was on the floor. Starting from 13 sec the person is standing and thus the height was drastically increasing.

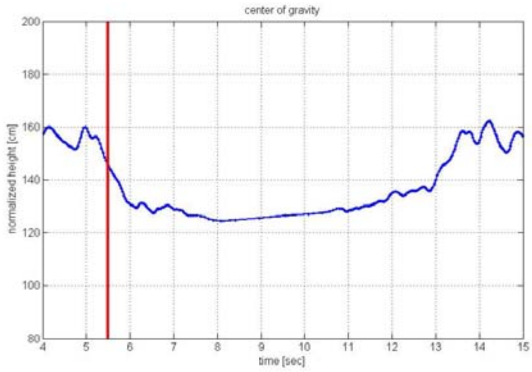


Figure 6: Normalized height of the gravity center of the event during the person motion. The red line shows the start time for the person fall

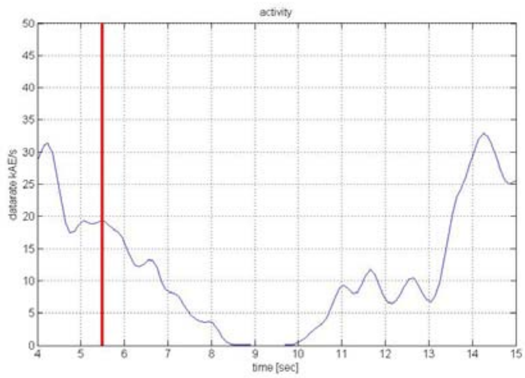


Figure 7: Event rate generated by the sensor during the person motion. The red line shows the start time for the person fall

The second parameter consisting of data rate (activity) is shown in Figure 7. In this 11 sec data, we can clearly notice the person fall after the red line (between 5.5 and 10 sec) illustrated in the decrease of event rate. Between 8.5 and 10 sec, the person was completely immobilized and therefore not generating any events. Afterwards the person is standing up and thus the event rate increases

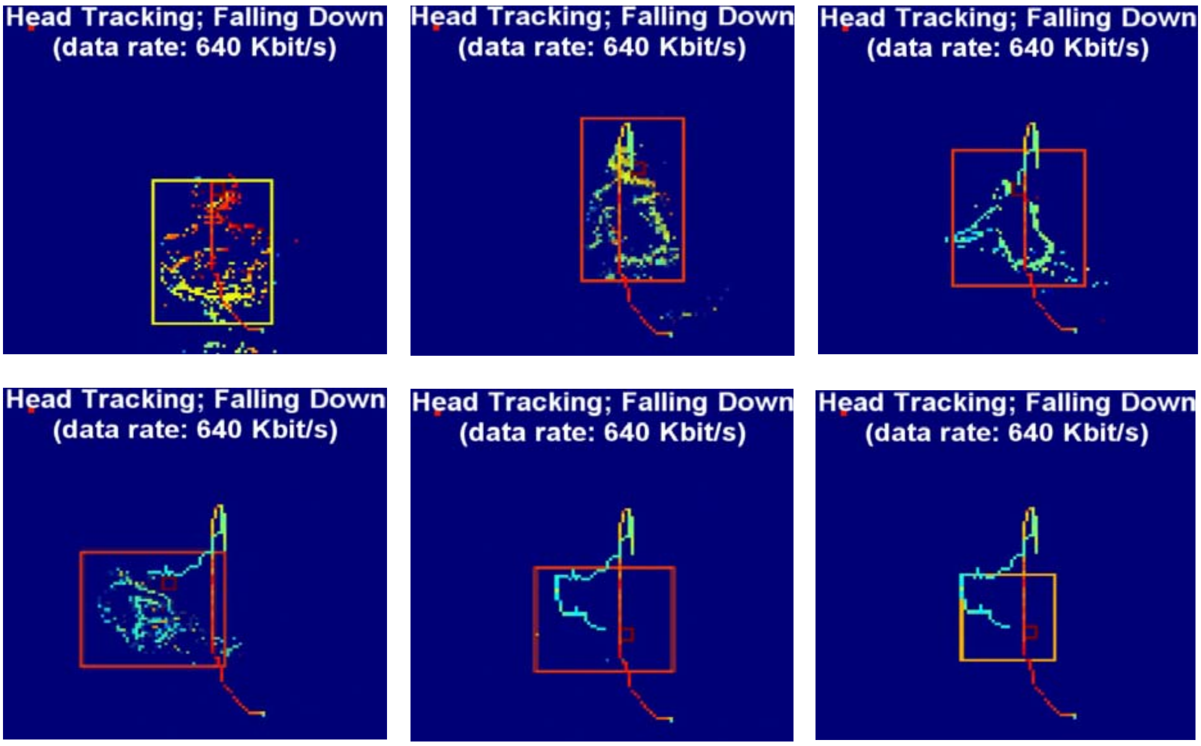


Figure 8: Sequence of images (top left to bottom right) showing the results of tracking and detection of person's fall from standing

drastically.

4.3. Real-time Fall Detection

The analysis performed in last subsection clearly showed that the case of a person fall from standing down to the floor can be detected using two features/parameters. Using the event-driven stereo vision system, the data rate and person height can be useful for the above-describe fall. Both features show a drastic decrease till the lowest point and therefore the detection might be possible.

In order to verify this assumption, a evaluation of real-time fall detection have been performed in real-time and implemented in the embedded system of Figure 1. In this case we performed height detection and tracking together with data rate evaluation. For the height detection, we assessed the highest point (head) as seen in Figure 8. The images in Figure 8 show the events generated from the stereo sensor, rendered in an image-like representation, including the depth information (color-coded), detection box and head-tracking results. The top left image of Figure 8 shows the events generated when the person entered the scene. The yellow box encodes that this person is newly detected. The box color is switched to red after few seconds (images middle/right top and left/middle bottom). The colored line shows the results of tracking the head and its distance from the sensor. The person fall is illustrated in from the middle top image till the left bottom

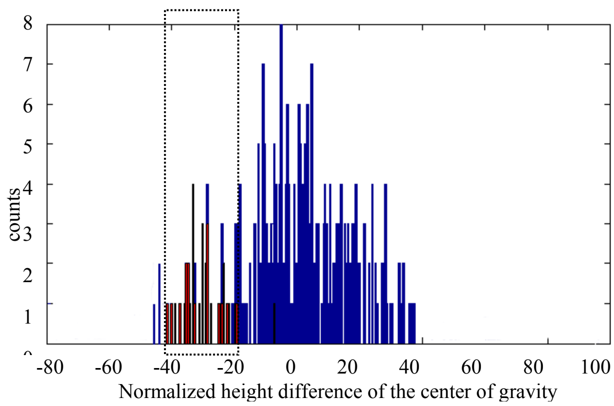


Figure 9: Fall statistics with the analysis of the height feature

image. The color of the tracked line changed from red (standing) to blue (falling) as the head get farther than the sensor. The middle and right bottom images do not show any events due to the person inactivity. For this reason a fall alarm is raised, which is illustrated in the right bottom image as orange box over the fall position. This case shows that the usage of height tracking (head tracking in this case) and the event rate can be useful for efficient detection of this kind of falls (from standing down to floor).

Figure 9 shows a statistical analysis of the temporal difference of the height from 110 acquisitions of sequences with person's activity at home. This data set includes 68 sequences with falls. The dashed square encloses the red bars consisting of fall scenarios. We can clearly notice that the height information of these falls is distributed in a distinguishable form (left corner). Therefore we believe that the height information together with the event rate is useful for the detection of fall incidents from standing position down to floor.

For other types of falls other features have to be investigated in order to consolidate the event rate and the height information for robust detection.

5. Conclusions

This paper presents a real-time event-driven stereo vision system for home monitoring and asynchronous detection of person incident like falls towards safety in ambient assisted living. Inspired from the biology, this stereo vision system allows efficient activity monitoring and with depth information such that standard falls (from standing) can be detected using data rate and depth information. The first analysis on 110 recordings including 68 fall cases shows promising results. Future investigations include assessment of false alarm rate using new recordings of person's activity at home. Furthermore automated detection of falls using e.g learning-based methods has to be investigated as well as the extension of this system evaluation to other types of falls. Additional features have to be investigated to enhance robustness of the detection and the reliability of the system

Acknowledgement

This work is supported by the AAL-EU JP project Grant CARE "aal-2008-1-078". The authors would like to thank all CARE participants who contributed to these results.

References

[1] A.N. Belbachir, M. Litzenberger, C. Posch and Peter Schoen, "Real-Time Vision Using a Smart Sensor System," in the International Symposium on Industrial Electronics, ISIE2007, Vigo, Spain, June 2007.

[2] A.N. Belbachir, "Smart Cameras," in Springer, New York, November 2009.

[3] A.N. Belbachir, M. Hofstaetter, N. Milosevic and P. Schoen, "Embedded Contours Extraction for High-Speed Scene Dynamics Based on a Neuromorphic Temporal Contrast Vision Sensor," Embedded Computer Vision'08, Workshop of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR'08, pp. 1-8, USA, June 2008.

[4] S.Y. Cho, C.G. Park and G.I. Jee, "Measurement System of Walking Distance Using Low-cost Accelerometers," In Proceedings of the 4th Asian Control Conference, Singapore, 2002.

[5] P. Lichtsteiner, C. Posch and T. Delbruck, "A 128×128 120dB 15us Latency Asynchronous Temporal Contrast Vision Sensor," IEEE Journal of Solid State Circuits, Vol. 43, Issue 2, pp. 566 – 576, Feb. 2008.

[6] M. Litzenberger, A.N. Belbachir, P. Schoen and C. Posch, "Embedded Smart Camera for High Speed Vision," in the IEEE International Conference on Distributed Smart Cameras, ICDSC'2007, pp. 81-86, Vienna, Austria, Sep. 2007.

[7] N. Milosevic, S. Schraml and P. Schön,"Smartcam for Real-Time Stereo Vision - Address-Event Based Stereo Vision," in Proceedings Image Understanding / Motion, Tracking and Stereo Vision, INSTICC Press, pp. 466 – 471, Barcelona, Spain, 2007.

[8] G. Pang and H. Liu, "Evaluation of a Low-cost Mems Accelerometer for Distance Measurement," Journal of Intelligent and Robotics Systems vol.30 pp. 249–265, 2001.

[9] C. Posch, R. Wohlgenannt, D. Matolin, "A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS", IEEE Journal of Solid-State Circuits, issue 46, vol. 1 pp. 259 – 275, 2011.

[10] S. Schraml, A.N. Belbachir, N. Milosevic and P. Schoen, "Dynamic Stereo Vision for Real-time Tracking," in Proc. of IEEE ISCAS, June 2010.

[11] S. Schraml, A.N. Belbachir, "A Spatio-temporal Clustering Method Using Real-time Motion Analysis on Event-based 3D Vision," in Proc. of the CVPR2010 Workshop on Three Dimensional Information Extraction for Video Analysis and Mining, San Francisco, 2010.

[12] S. Schraml, N. Milosevic and P. Schön, "Smartcam for Real-Time Stereo Vision - Address-Event Based Stereo Vision," in P. of Computer Vision Theory and Applications, INSTICC Press, pp. 466 – 471, 2007.

[13] V. Stanford, "Using Pervasive Computing to Deliver Elder Care," IEEE Pervasive Computing vol. 1, issue 1. pp.10-13, 2002.

[14] D.Willis, "Ambulation Monitoring and Fall Detection System using Dynamic Belief Networks," Bachelor Thesis, Monash University, Australia, 2000.

[15] K.H. Wolf, A. Lohse, M. Marschollek and R. Haux, "Development of a Fall Detector and Classifier Based on a Triaxial Accelerometer Demo Board," In Proceeding of the UbiComp 2007 Workshop, pp. 210-213, Innsbruck, Austria, 2007.