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.

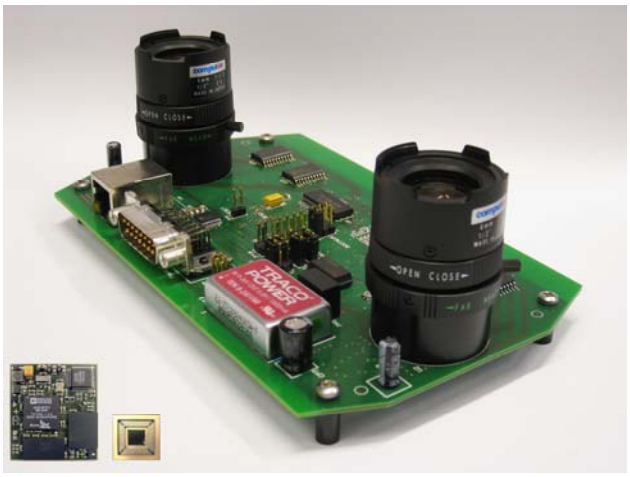

Figure 1: Image of the event-driven stereo vision system. In the lower left corner the DSP Bf537 and the sensor chip are shown. The DSP is mounted on the back of the board.

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.

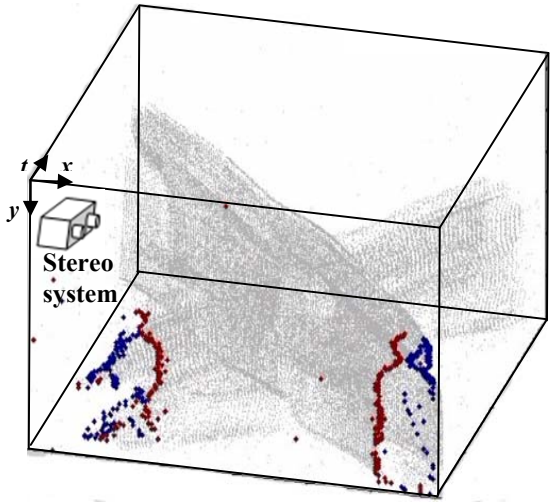

Figure 2: Event representation of scene dynamics (2 persons crossing the field of view) in a space-time domain using 1 DVS. The data are shown in a room-like representation with sensor mounted on the side wall.

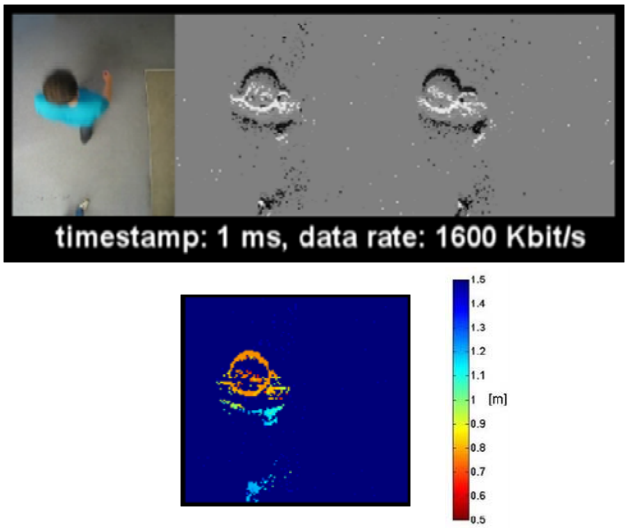

Figure 3: Still image of a person from a conventional video camera (top left); the corresponding events of a pair of dynamic vision sensors (top middle and right) and resulting event "sparse" depth map (bottom) rendered in an image-like representation.

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

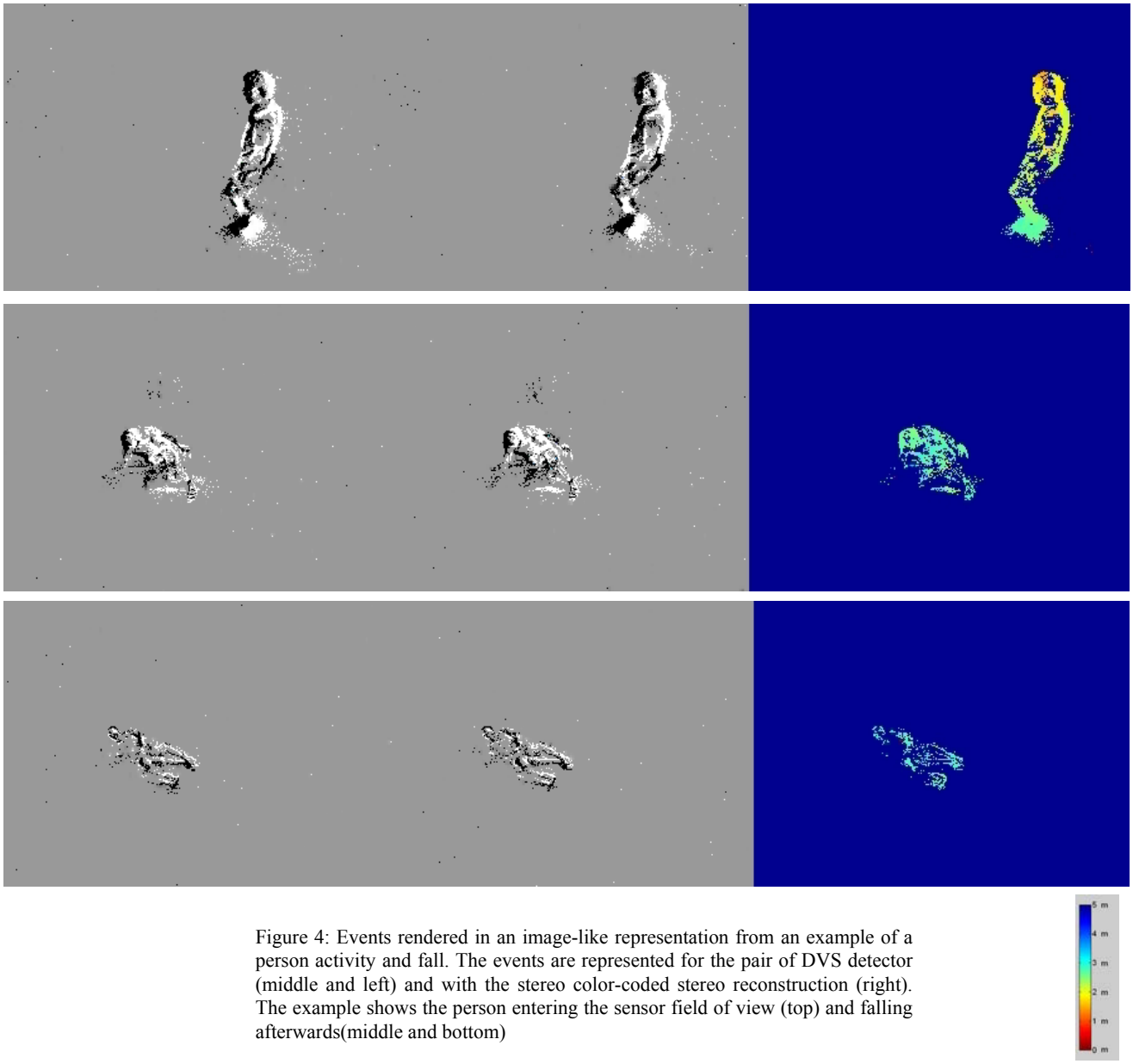

Figure 4: Events rendered in an image-like representation from an example of a person activity and fall. The events are represented for the pair of DVS detector (middle and left) and with the stereo color-coded stereo reconstruction (right). The example shows the person entering the sensor field of view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

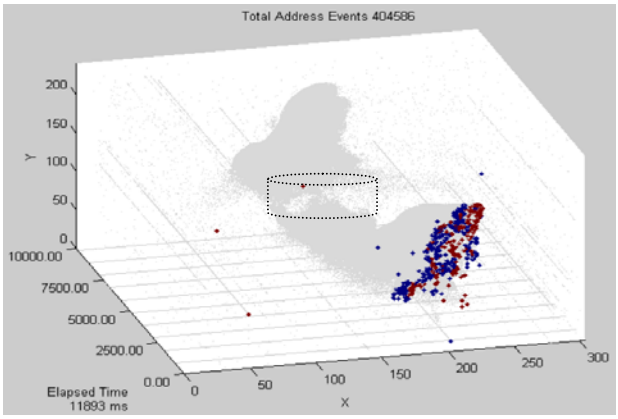

Figure 5: Continuously generated events for 10 sec person's motion including 3 sec inactivity (dashed cube) after a fall

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.

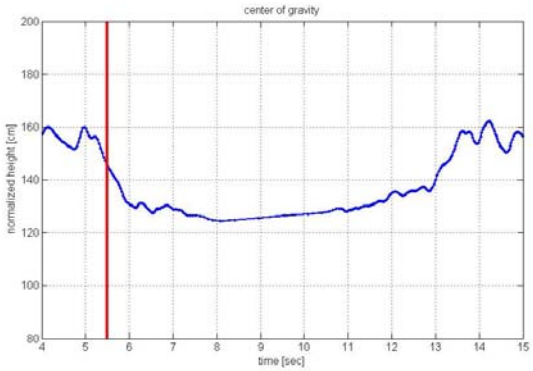

Figure 6: Normalized height of the gravity center of the event during the person motion. The red line shows the start time for the person fall

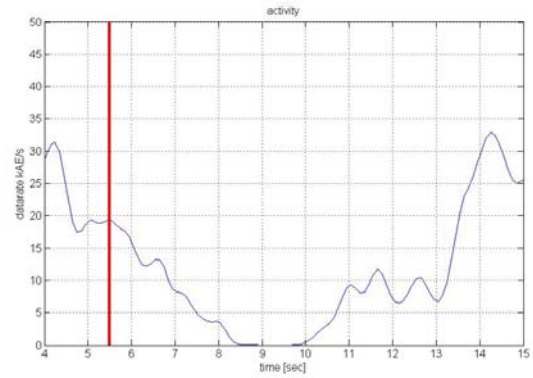

Figure 7: Event rate generated by the sensor during the person motion. The red line shows the start time for the person fall

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

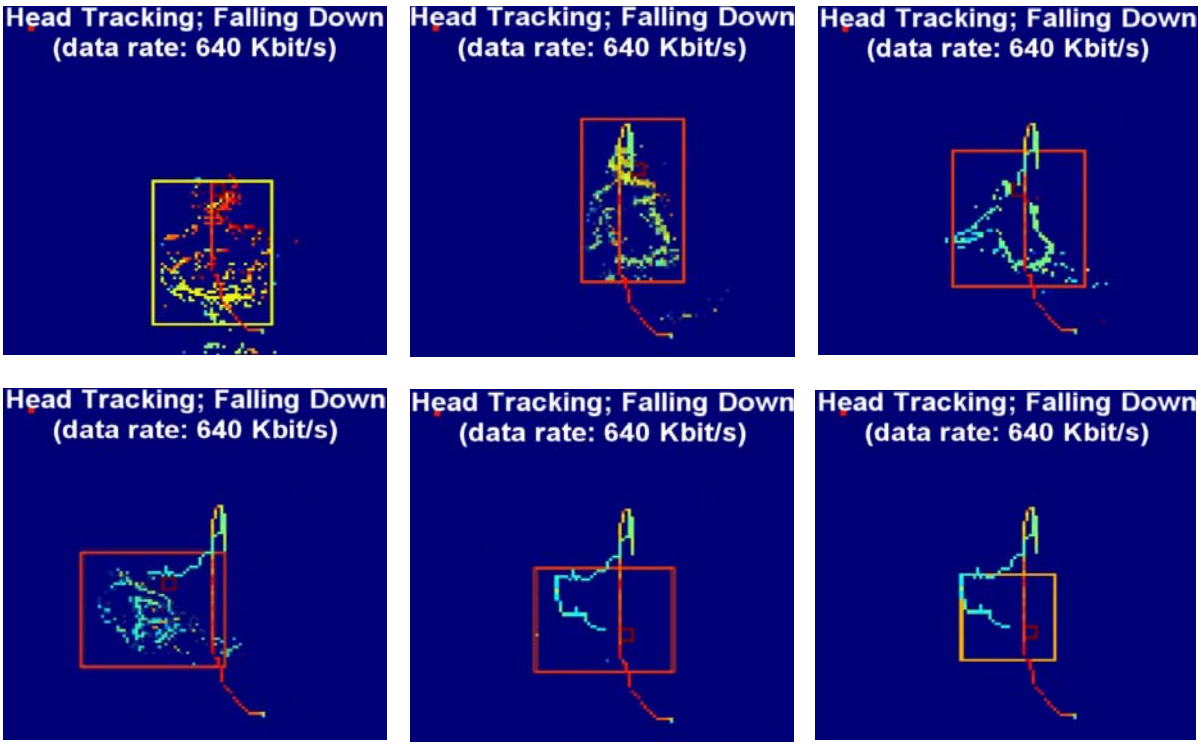

Figure 8: Sequence of images (top left to bottom right) showing the results of tracking and detection of person's fall from standing drastically.

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 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).

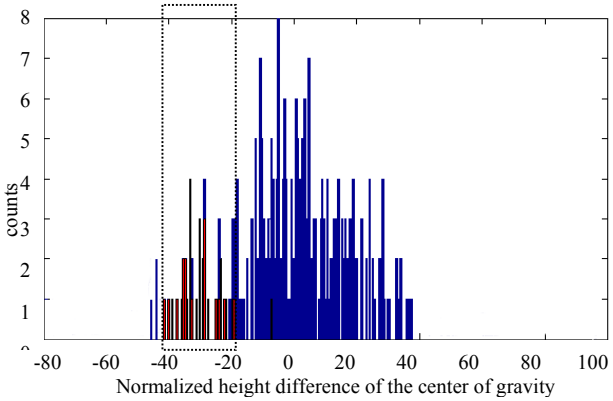

Figure 9: Fall statistics with the analysis of the height feature

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.

Event-driven Feature Analysis in a 4D Spatiotemporal Representation for Ambient Assisted Living

Ahmed Nabil Belbachir, *Member* IEEE, Aneta Nowakowska, Stephan Schraml, Georg Wiesmann
New Sensor Technologies, Safety & Security Dep., AIT Austrian Institute of Technology
Donau-City Str. 1/5, A-1220, Vienna Austria.

{nabil.belbachir; aneta.nowakowska; stephan.schraml; georg.wiesmann}@ait.ac.at

http://www.ait.ac.at

Robert Sablatnig

Computer Vision Lab
Institute of Computer Aided Automation,
Vienna University of Technology
Favoritenstr. 9/183, A-1040, Vienna, Austria

sab@caa.tuwien.ac.at

http://www.caa.tuwien.ac.at/cvl

Abstract

This paper presents a detailed analysis of a 4D representation of events, which are generated by a dynamic stereo vision sensor for the recognition of person's fall. Dynamic vision detectors consist of self-signaling pixels that autonomously react to scene dynamics and asynchronously generate events upon relative light intensity change. Their complete on-chip redundancy reduction, wide dynamic range and high temporal resolution allow efficient and continuous activity monitoring in natural environment. Using a stereo pair of dynamic vision detectors, it is possible to represent the scene dynamics in a 4D space (including time) at a high temporal resolution. In this work, we performed 100 recordings of scenarios including falls in indoor environment using this dynamic stereo vision sensor. Seven features have been extracted and analyzed for three types of falls such that robust parameters will be kept for fall recognition. The result of this analysis is shown in this work with promising outcomes.

1. Introduction

One of the highest risks for elderly persons living completely alone or spending most of the time alone is falling down and being unable to call for help, especially in case of loss of consciousness. While critical situations can occur principally in all home locations and situations, the risk is particularly high in bathrooms, where critical conditions increase the possibility of falls, collapses or cardiac and circulatory troubles. Several technologies with wearing parts are often used for monitoring elderly people in nursing homes including RFID tags and accelerometers. These technologies are either not popular, inconvenient and often disposed in such situations, rendering them of little use for detecting potentially hazardous situations. Furthermore, tags are usually taken off by persons during activities such as taking a bath or fitness and sport where the probability for incidents like falls is high.

Camera-based systems can be stationary mounted to perform visual surveillance for the automated detection of activities and dynamics in the scene. 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 may be computationally expensive, especially if higher temporal resolution is required for continuous tracking of scene dynamics; moreover, it is not always reliable. For instance, pixels in low texture areas are very hard to match.

In [1], a smart ambient approach for 3D representation of activities and scene dynamics for the recognition of person's fall has been proposed. The presented system consists of a new 3D visual sensing technology, stationary mounted in home environment (living room and bathroom) for automatically monitoring and recognition of falls. 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 having an on-chip background subtraction, compared to image-based sensors and efficiently capture scene

dynamics [1][6]. The stereo system described in [1] efficiently calculates the depth information of the dynamics and allows a low-cost and a low-power 3D representation.

This paper presents a detailed analysis of 4D represented data for fall detection, by including the temporal information to the spatially represented data. A set of features have been extracted from several recorded fall scenarios. The stability of every feature over the whole data set has been investigated, to select relevant features for 4D recognition. A neural-net based learning system should have been used for automated and real-time recognition of person's fall, to be implemented in the embedded hardware of the stereo system. The paper is structured as follows: Section 2 provides a brief review of the dynamic stereo vision system presented in [1]. Section 3 gives a short description on how we were handling the continuous event stream. The list and property of all selected features for the analysis of the 4D represented data is given in section 4. The temporal analysis of the features on three fall types is shown in section 5. In section 6, the statistical analysis results of the fall scenario features are discussed. A summary concludes the paper in section 7.

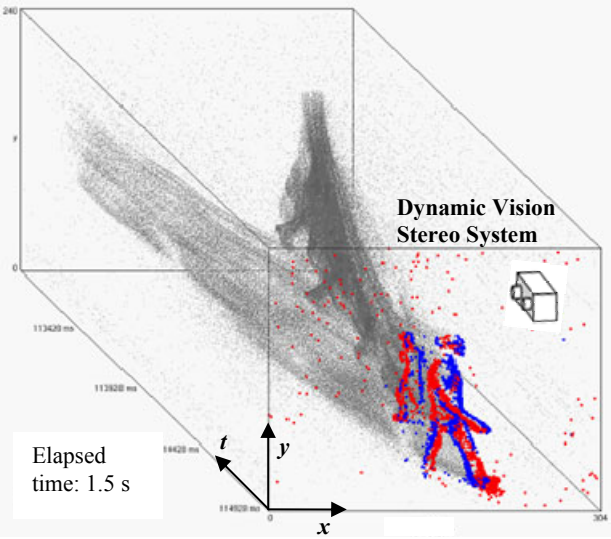

Figure1: Spatiotemporal representation of one dynamic vision detector capturing scene dynamics of 2 persons crossing the field of view. The events data are shown in a room-like representation with sensor mounted on the side wall.

2. Dynamic Stereo Vision Sensor

The dynamic stereo vision sensor consists of an array of 304x240 pixels, built in a standard 0.18µm CMOS-technology. The array elements (pixels) respond to

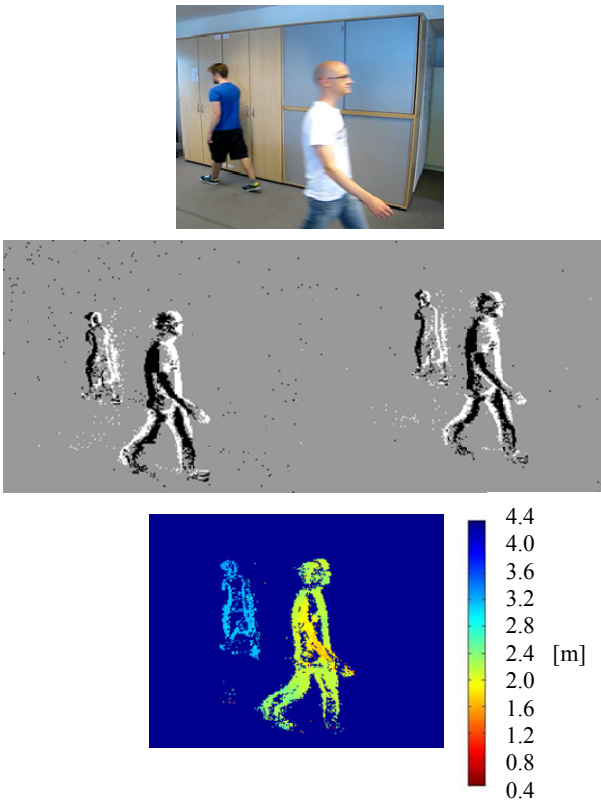

Figure2: Still image of a person from a conventional video camera (top); the corresponding events of a pair of dynamic vision sensors (middle) and resulting event "sparse" depth map (bottom) rendered in an image-like representation.

relative light intensity changes by instantaneously sending their address, i.e. their position in the pixel matrix, asynchronously over a shared 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 an FPGA, which attaches to each AE a timestamp at a resolution of 1µs or less. The combined data (AEs and timestamps) are used as input stream for depth map generation and subsequent processing.

Figure 1 depicts a spatiotemporal data representation of one dynamic vision detector, resulting from a two persons crossing the sensor field of view in a room-like environment. The events are represented in a volume with the coordinates x (0:304), y (0:240) 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

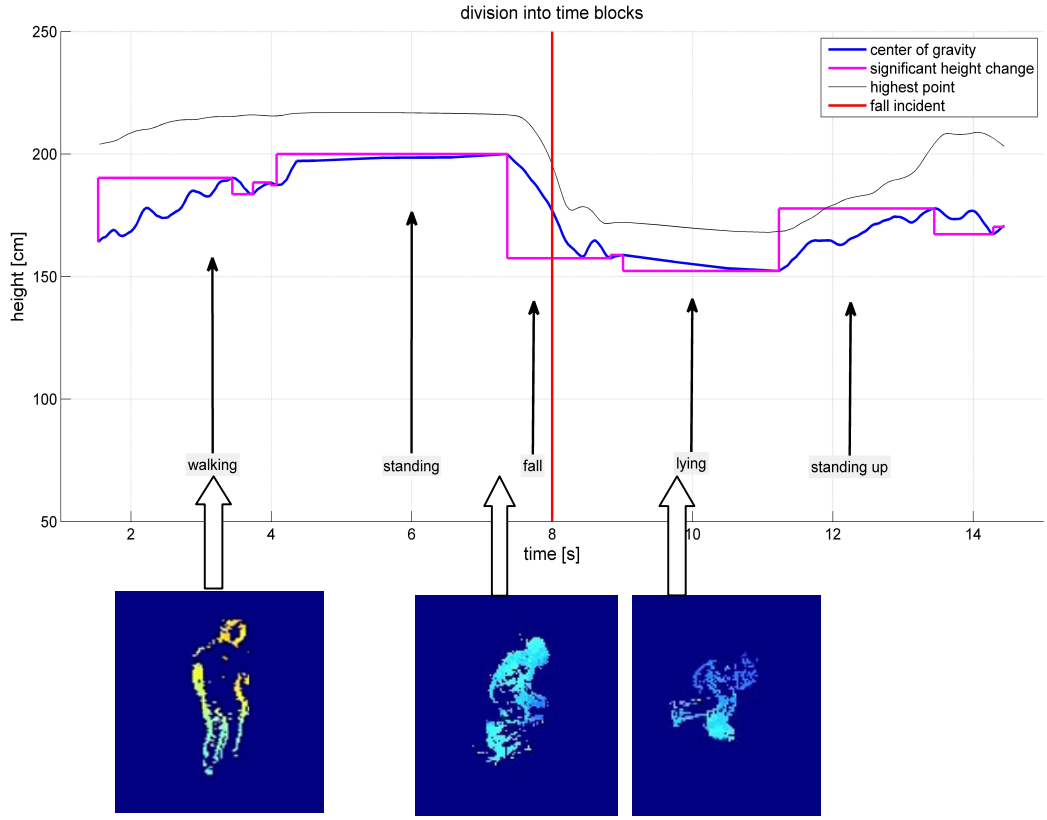

Figure3: Illustration of the temporal segmentation of the person's activity data into blocks using the direction changes of the center of gravity (vertical axis). The three figures in the bottom show the instant poses of monitored person

events generated in the elapsed 1.5 seconds prior to the recent 16ms. This event history highlights the dynamics path in the past 1.5 sec resulting from the moving persons, which is an ideal basis for continuous monitoring by simultaneous detection and tracking in space and time.

Figure 2 shows the the instant picture of the visual scene in figure 1 imaged by a conventional video camera (top) and its corresponding AEs using a pair of dynamic vision detectors (middle) rendered in an image-like representation. The white and black pixels represent spike 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 of the dynamic stereo vision sensor embeds event-based stereo vision algorithms, including the depth generation or the so-called sparse depth map, where the algorithm is described in [8] in detail. The resulting sparse color-coded depth map of the scene dynamics is provided at the bottom in Figure 2.

3. Analysis of the 4D Representation

For the analysis of the spatiotemporal generated events, time blocks have been created at various temporal lengths (figure 3) depending on the persons' movements. The start and end of every time block was triggered by the vertical direction of Center Of Gravity of events for which we use the acronym COGz. If the height of the COGz changes significantly, a new time block is annotated. In this way, time periods with small or similar changes in the height of the person are merged to one time block. The aim of this temporal segmentation of the processing is to regulate the processing of the asynchronous data and their feature extraction with respect to the continuous movements of a person. In this way, it was possible to investigate how the features behave during a several individual time slots and localize the block where a fall may be happened in comparison to time blocks without falls.
Figure 3 depicts procedure of the temporal segmentation of the asynchronous sensor events, and shows screen shots of sparse (event) depth map with person poses (walking, falling, lying).

4. Features Analysis

For the feature extraction, it was necessary to transform the recorded data into world coordinates by means of the depth information, so that every AE is described as data point in the world coordinate system of the room. In this way it was possible to represent the person as a 3D point cloud and every point is characterized by its X, Y and Z coordinate. The Z direction corresponds to the height, while the X and Y- axes span the ground plane of world coordinate system. Using this representation, we were able to extract and analyze seven distinct features.

Feature 1: Mean in Z- direction

The first intuitive feature was the mean (also called center of gravity, COGz) of the events according to the z (person height) coordinate, which were computed on an amount of data points produced by the person within a specific time span (0.04 sec in our case).

Feature 2: Vertical velocity of the COG z

The second feature extracted was the vertical velocity of the mean height (Z coordinate). It was computed in a way that it should be negative when the person movements go towards the ground plane and will be positive when the COGz moves up.

Feature 3: Highest point in Z- direction (Height)

This feature represents the person highest point. To avoid outliers, we computed the height under which 97 % of the persons data points lie.

Feature 4: Ratio of the bounding cone to height

One non-intuitive feature consists of the so-called bounding cone. This latter is computed as the variance of the person's data points projected on the ground (XY) plane. While the height (highest point) is expected to decrease during a fall, the bounding cone is minimal when the person is standing upright and expected to rise during falls. Figure 4 demonstrates the bounding cone which is determined by the projection of the persons' data points onto the ground plane. In the left case, when the person is standing upright, the variance or cone in the ground plane is smaller than in the right case of a lying person.

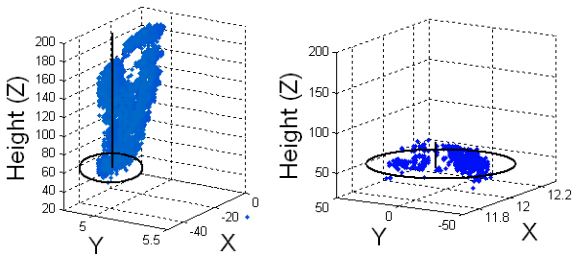

Figure 4: Bounding cone (Ellipsoid in the XY plane) and height of an upright standing person (left) and a lying person (right)

Feature 5: Angle of the main axis

The fifth feature extracted from the data was the person declination to the main vertical axis. The main axis of the body is computed as the axis with the maximal variance with the help of Principal Component Analysis. The angle between the ground plane (XY – plane) and the main axis is in an upright position expected to be about 90 °, while in lying positions it decreases to 0°.

Feature 6: Vertical volume distribution ratio (VVDR)

The VVDR feature was taken from the work [12] and it describes the distribution of the data points with respect to the height. It is computed as the amount of data points, which lie below a specific height (e.g. 30 cm), divided by the whole amount of data points. If the person is lying on the floor, almost all data points generated by the person will lie below the threshold and so the VVDR will rise.

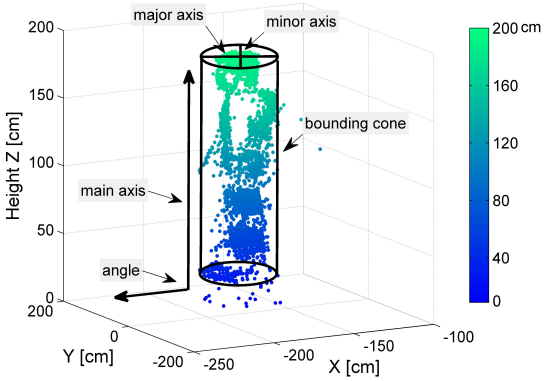

Figure 5: Main axis of an upright standing person

Feature 7: Mean activity / Address Event rate

The last feature used was the mean activity. Due to the asynchronous data generation of the sensor, the amount of produced AEs depends on the amplitude of the scene dynamic and therefore on how much the person is moving. The mean activity measures the amount of address events produced by the person activities in a second. Due to the decrease of activity with respect to the person distance from the sensor, we use the square distance to weight the mean activity.

5. Experimental Results

An amount of 100 scenarios were recorded in a lab environment including 50 simulated falls from walking/ sitting and lying to investigate the significance of possible features for fall detection. These scenarios were analyzed and are discussed in the next subsections.

5.1. 4D Recognition of Falls from Walking

Figure 6 shows the spatiotemporal 4D representation of a person during a fall from walking position. The depth is color coded so that with rising distance between the object and the sensor the object is marked darker (from green to blue). The person walks away from the sensor and falls after 9 sec in the same direction of his walking. Therefore the color of the data points representing the person decreases with time from bright to dark blue.

Figure 7 shows the corresponding features of the fall scenario from walking, where the fall instance is represented by the red line. We can notice the sudden decrease of the mean height (feature 1), of the highest point (feature 3) and the angle (feature 5). The velocity (feature 2) and the activity rate a sudden decrease/increase followed by a swift increase/decrease respectively, which represents the person immobilization after the fall. The bounding cone/ height ratio and the VVDR show a sudden increase, which mean that the person is lying on the floor. From this feature, it can be concluded that the recognition of the fall for this specific scenario is possible, by using a combination of these features.

5.2. 4D Fall Recognition while Sitting

In this subsection, we are providing the scenario of a person falling while trying to sit down. Figure 8 shows data generated by the sensor for scenario where the person falls while trying to sit down on a chair. The depth information of the data is color-coded (like in figure 6) where green means close to the sensor and blue means far from the sensor. The person has fallen after 16 sec of activity. Figure 9 depicts the graphical representation of the features along this scenario. This scenario seems to be very similar to the first scenario (fall from walking) where the mean height (feature 1), the highest point (feature 3) and the angle (feature 5) suddenly decreased at the fall instant. The other features also show the same characteristic as those of the first scenario.

5.3. 4D Fall Recognition while Lying

In this case we deal with spatiotemporal data from a difficult scenario where the fall out of a lying position on the bed has to be recognized. Figure 10 shows the spatiotemporal event representation with the distance information to the sensor (color-coded). The legs of the person are nearer to the sensor and therefore brighter (green) while the body is blue. As figure 11 shows, the COG Z of the height before the fall is lower than that in figure 7 and 9 because the person is lying on a couch. Therefore the height difference of the COG Z during a fall from lying is lower than a fall from sitting or walking.

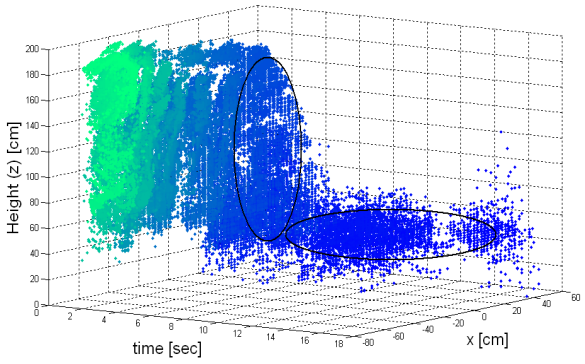

Figure 6: Spatiotemporal 4D representation of scene dynamics of a fall from walking (depth is color coded)

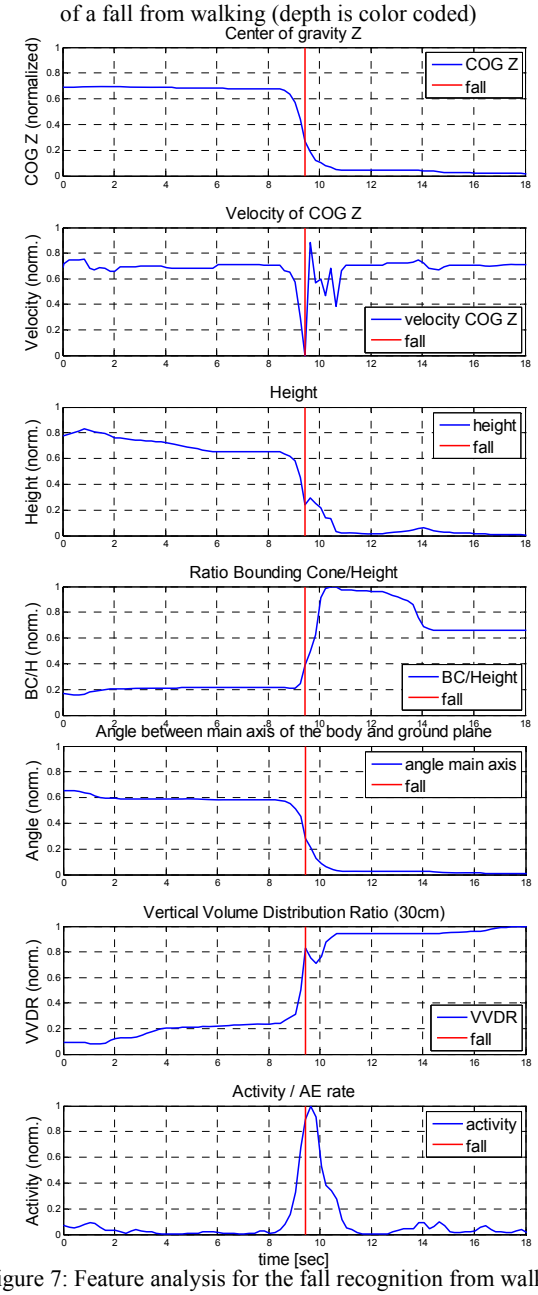

Figure 7: Feature analysis for the fall recognition from walking

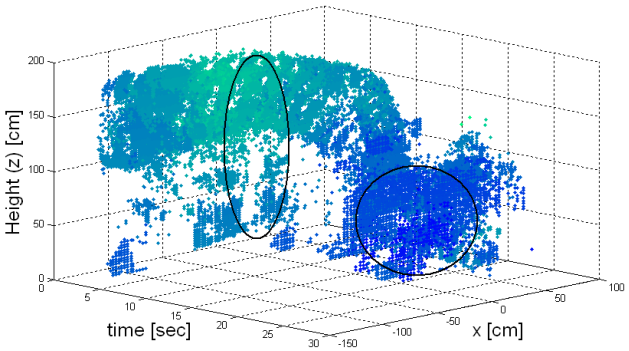

Figure 8: Spatiotemporal 4D representation of scene dynamics of a fall while sitting (depth is color coded)

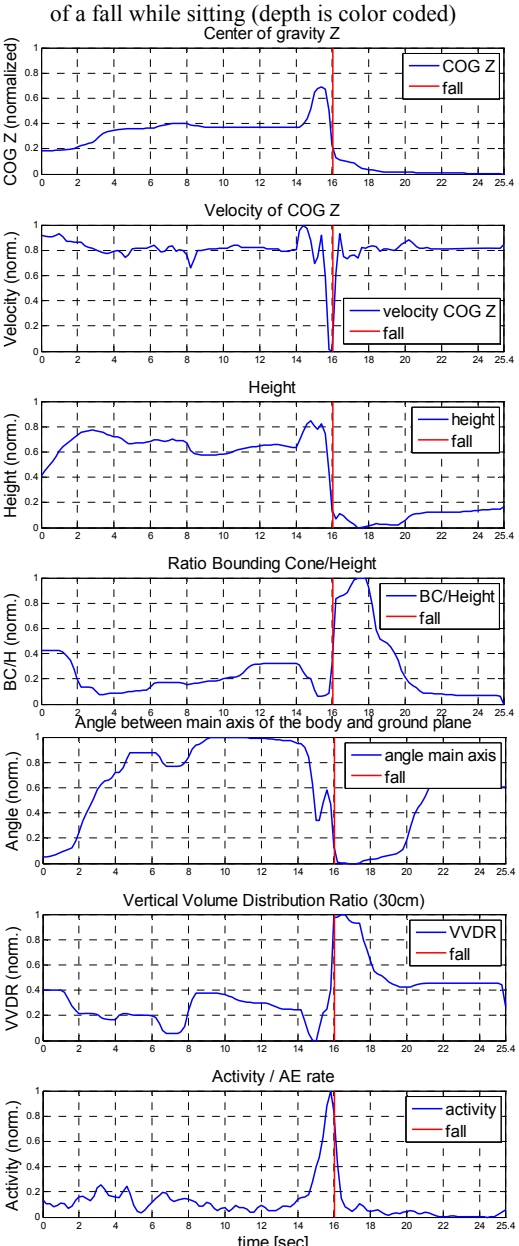

Figure 9: Feature analysis for the fall recognition while sitting

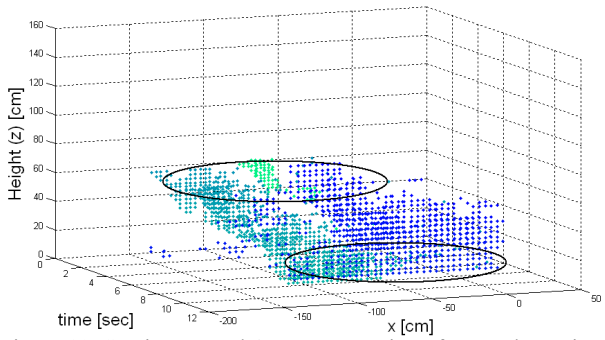

Figure 10: Spatiotemporal 4D representation of scene dynamics of a fall while lying (depth is color coded)

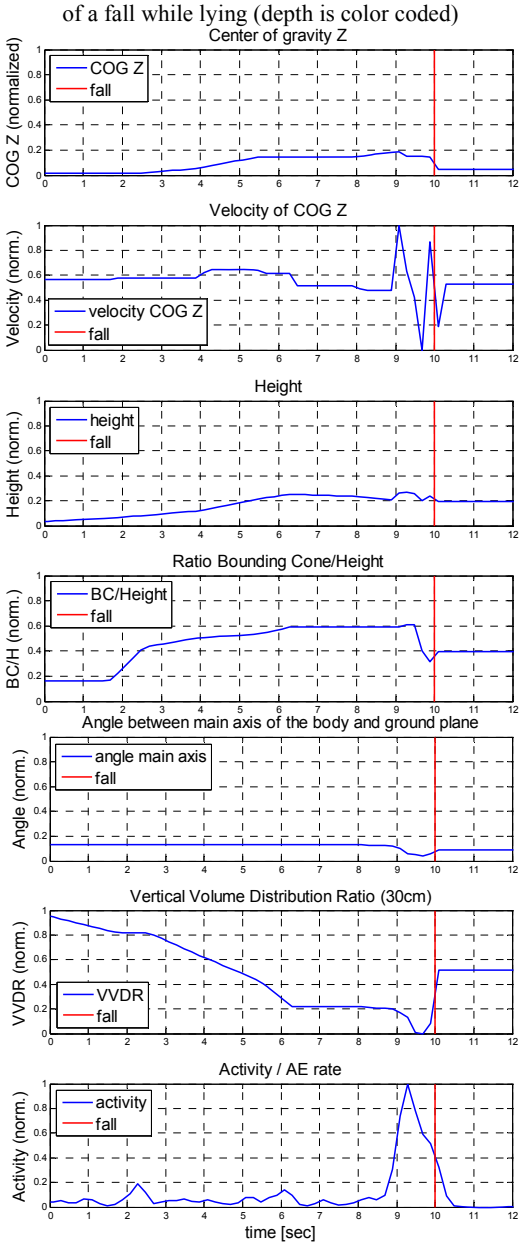

Figure 11: Feature analysis for the fall recognition while lying

The angle of the main axis does not change and is low even before the fall because the pose during a fall from lying remains the same. The activity before the fall, during lying on the couch, is low and therefore the computed features are less significant than in the first two cases (walking and trying to sit down). The activity feature shows a sudden change due to the instantaneous motion of the person during the fall event.

6. Statistical Analysis of the Recordings

Within this statistical analysis, we are analyzing the feature representation according the time blocks of every scenario. As already mentioned, the recorded data of every scenario were divided in time blocks of variable length, in which we were continuously evaluating the features. As a cost function, the decrease of the entropy was computed to specify the significance of the features. It contains information about how well time blocks containing the fall event can be distinguished from other (no fall) blocks using the specific feature. In this way a classification of all time blocks and the computation of the decrease of entropy are done with each of the features. The entropy is computed with following formulas [11]

$$H = \frac{p * Ip}{S} + \frac{n * In}{S} \quad (1)$$

$$Ip = -\log_2(\frac{p}{S}) \quad (2)$$

$$In = -\log_2(\frac{n}{S}) \quad (3)$$

H… Entropy
p,n… number of correct (p), false (n) classified time blocks
S… Number of all time blocks
Ip, In … Information content of the correct (Ip) and false (In) classified time blocks

For a visual analysis of the features, a histogram of the time blocks was plotted. For each time block the difference between the beginning and the end of the time block of each feature is computed. It measures how much a feature alters during a time block. A significant feature alters during fall time blocks more than during no fall time blocks and therefore the difference between the beginning and the end of a time block is higher. For instance, during time blocks with only walking activity, the angle of the main axis should alternate about 80 degrees. In contrast, during a fall time block (fall from standing), the persons' pose changes from standing upright to lying and therefore the angle decreases from about 80° at the beginning of the time block to about 20° at the end of the time block, that is the end of the fall. In the case of the vertical velocity of the COG Z and the mean activity, measuring the difference between the end and beginning of the time block does not make sense because they do not describe the pose but the dynamic of the persons' activity. Therefore for the vertical velocity of the COG Z the minimum and for the mean activity the difference between maximum and minimum of the time blocks are computed.

7. Discussions

Figure 12 shows the histogram of the time blocks for the angle of the main axis. It contains the whole amount of assembled time blocks generated by dividing all 100 records into time blocks. The red bars stand for time blocks containing falls, the blue bars are time blocks without falls. The different types of time blocks can be separated relatively well in comparison with the histogram in figure 13 of the vertical velocity of the center of gravity.

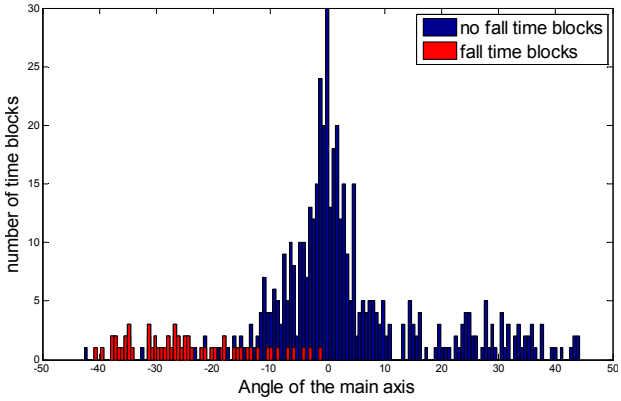

Figure 12: Histogram of the angle of the main axis

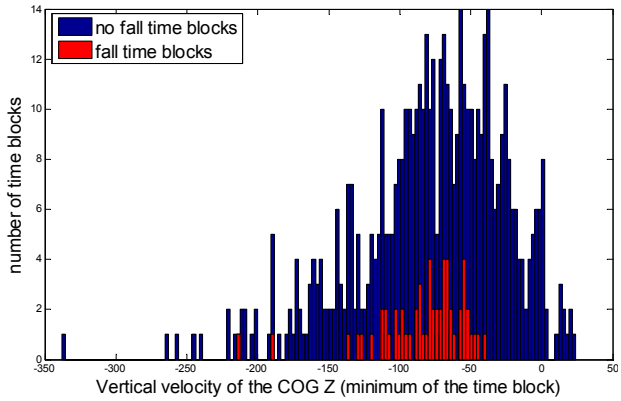

Figure 13: Histogram of the vertical velocity of the COG Z

To compute the decrease of entropy, thresholds for each feature defining the bounds for fall time blocks were extracted empirically from the histograms. As an example in the above histogram (figure 12) of the angle of the main axis, the thresholds could be -20° and -40°. As next step time blocks lying within these thresholds are classified as

fall time blocks, the remaining can be classified as no fall blocks. With the help of the resulting classification error, the decrease of entropy can be computed according to the formulas in section 6.
The following table presents the cost function (entropy decrease) for every feature based on the evaluation performed for the whole data set. The features are sorted according to the decrease of the entropy during classification. According to this evaluation, the most significant feature is the angle of the main axis, showing an entropy decrease of 0.2935, while the least significant features are the mean activity and vertical velocity of the COG Z.

Ranking	Decrease of entropy	Feature description
1	0.2935	Angle of the main axis (difference)
2	0.2219	Ratio bounding cone/ height (difference)
3	0.1942	Center of gravity (difference)
4	0.1717	Highest point (difference)
5	0.1133	Vertical volume distribution ratio (difference)
6	0.0608	Vertical velocity of the COG Z (minimum)
7	0.0304	Mean activity (maximum- minimum)

7. Conclusions

Within this investigation, a feature analysis has been performed using 4D data from event-driven stereo vision sensor. The 4D data represent spatiotemporal scene dynamics collected upon person's activity during in-home monitoring, and is intended to be exploited for developing a real-time method for robust recognition of person's fall. A total of seven features were extracted from data generated by the event-driven dynamic stereo vision system, having a data set including 100 scenarios. These features were individually analyzed for three types of falls (fall from walking, fall while sitting and fall from lying). As a general conclusion of this work, we noticed that the angle of the main axis of the person and the ratio between the bounding cone vs. the height seem to be the most significant features for detecting the fall and were always sensitive (within the 100 scenarios) at the incident instant. Our next step will be to analyze these features for a larger data set such 1500 has to be recorded within the period July - September 2011. A further next step is to teach a decision system, based on the neural network in recognizing falls and to evaluate the whole system for home monitoring and asynchronous and automated detection of person falls.

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. The second author master work was further supported by an additional Austrian grant under the topic "CHANGES" dealing with woman sponsorship and gender studies.

References

[1] A.N. Belbachir, S. Schraml and A. Nowakowska, "Event-driven Stereo Vision for Fall Detection," in Proceedings of the IEEE CVPR Workshops under Embedded Computer Vision Workshop, p.p. 82-87, USA, June 2011

[2] 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.

[3] A.N. Belbachir, "Smart Cameras," Springer, Nov. 2009.

[4] S.Y. Cho, C.G. Park and G.I. Jee, "Measurement System of Walking Distance Using Low-cost Accelerometers," In Proc 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, Austria, Sep. 2007.

[7] 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.

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

[9] 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.

[10] 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.

[11] Shannon C., A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review 5, no. 1, pp. 3–55., 2001

[12] E. Auvinet, F. Multon, A. Saint-Arnaud, J. Rousseau and J. Meunier, "Fall detection with multiple cameras: an occlusion-resistant method based on 3-d silhouette vertical distribution," IEEE T.Inf. Tech. Biomed, pp. 290-300, 2011