Fall Detection using an Address-Event Temporal Contrast Vision Sensor

Zhengming Fu, Eugenio Culurciello
Electrical Engineering Department
Yale University, CT USA

Patrick Lichtsteiner, Tobi Delbruck
Institute for Neuroinformatics (INI)
Zurich, Switzerland

***Abstract*—In this paper we describe an address-event vision system designed to detect accidental falls in elderly home care applications. The system raises an alarm when a fall hazard is detected. We use an asynchronous temporal contrast vision sensor which features sub-millisecond temporal resolution. A lightweight algorithm computes an instantaneous motion vector and reports fall events. We are able to distinguish fall events from normal human behavior, such as walking, crouching down, and sitting down. Our system is robust to the monitored person's spatial position in a room and presence of pets.**

I. Introduction

Human society is experiencing tremendous demographic changes in aging since the turn of the 20th century. According to a report of U.S. Census Bureau, there will be a 210% increase in the population with age of 65 and over within the next 50 years [1]. The substantial increase in the ageing population will cause society to face two challenges: increase of ageing people will require more investment in elderly care services; decrease of working population will cause shortage of skilled care-givers of elders. In the future, this imbalance between the number of elderly people and that of the care-givers will be exacerbated when life expectancies increase. In the past, various solutions were proposed based on emerging technologies. *Video monitoring* is a commonly-used solution in nursing institutions. But considerable human resource is required in order to monitor activities. Patients' privacy is also compromised when they are monitored. Another common solution is to have patients raise alarms when they are in trouble by pushing a button on a wearable or pendant device [2]. This solution depends on the patient's capability and willingness to raise alarm. For example, a fall may result in unconsciousness. A dementia patient may not be able or willing to push the button when necessary [3]. Both scenarios would limit this "push-the-button" solution in applications.

Fall is a major health hazard for the elders when they live independently [4]. Approximately 30% of 65-year or older people fall each year. This number becomes higher in medical service institutions. Although less than one fall in 10 results in an injury, a fifth of fall incidents require medical attention. How to effectively assess, response and assist those elderly patients in trouble becomes an important research topic in medical elderly care services [5].

In this paper we describe a fall detector using an asynchronous temporal contrast vision sensor. Figure 1 illustrates the fall detector setup. The detectors take multiple side-views of the scene in order to detect accidental activities and raise alarms. Our approach is innovative for two reasons: First, an asynchronous temporal contrast (motion-detection) vision sensor features very-high readout speed and reports pixel changes with a latency on the order of milli-second. Second, a lightweight computation algorithm plus a fast readout allow us to compute an instantaneous motion vector and report fall events. This cannot be done with a frame-based temporal-difference image sensor because the frame rate is constant, and redundant information in images saturates the transmission bandwidth.

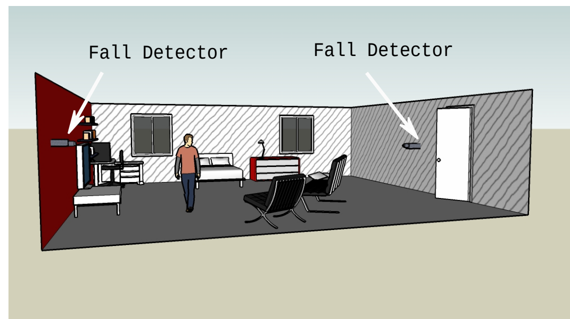


Fig. 1. Address-event fall detectors used for assisted living applications. The detectors are mounted on the wall at the same height as a light switch.

II. Asynchronous temporal contrast vision sensor

A temporal contrast vision sensor extracts changing pixels from the background and reports temporal contrast, which is equivalent to image reflectance change when lighting is constant. A temporal contrast vision sensor can extract motion information because, in normal lighting conditions, the intensity of a significant number of pixels changes as an object moves in the scene. In this paper we are investigating an asynchronous temporal contrast (ATC) vision sensor, which outputs sequences of events [6] [7]. In this ATC vision sensor, each pixel is self-timed and responds to relative changes in light intensity. When the change of light intensity in a pixel passes a threshold, an event is triggered. The corresponding pixel address is transmitted. After the event is acknowledged by an external receiver, the pixel resets itself. A major advantage of this ATC image sensor is that it pushes information to the receiver once a predefined condition is satisfied. This feature is

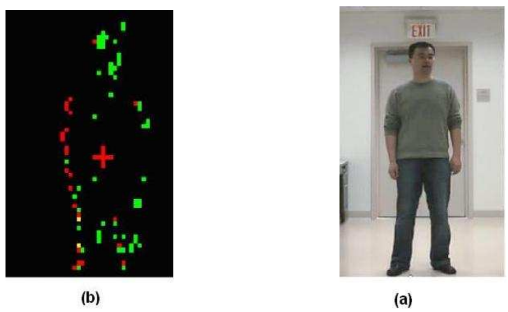


Fig. 2. (a) Temporal contrast image from the ATC image sensor and (b) its targeted scene. The ATC imaging system is placed in front of the subject with a distance of 3 meters and a height of 0.8 meter.

important in high-speed vision systems because a pixel sends information of interest immediately, instead of waiting for its polling sequence. Figure 1 shows the ATC image sensor system used in the experiment. The imaging system streams a series of time-stamped address-events from the vision chip, and sends them to a PC via a USB interface. The silhouette of a moving object can be reconstructed on a PC. The temporal contrast vision sensor used in this work contains a 64×64 array of pixels and responds to relative changes in light intensity. Figure 2 shows an image from the ATC image sensor and its targeted scene. The imaging system is placed in front of the subject with a distance of 3 meters and a height of 0.8 meter.

III. Detecting falls using address-event vision sensor

We use a temporal average of the motion events from the ATC vision sensor, here referred as *centroid event*, to track fall hazards and evaluate its dynamics.

A. Centroid event computation

Centroids are an effective way to estimate object motion in space. Centroids can be computed as temporal averages of a series of events. A single centroid event address, (x_c, y_c), during a fixed period can be calculated using Equation 1.

$$x_c = \lceil \frac{\sum_{i=1}^{N} x_i}{N} \rceil, y_c = \lceil \frac{\sum_{i=1}^{N} y_i}{N} \rceil \tag{1}$$

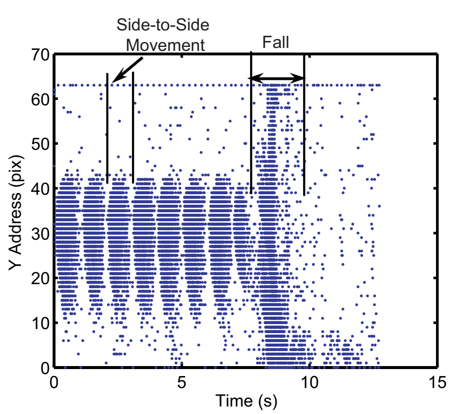


Fig. 3. Vertical address (Y) of events when the camera is monitoring a fall. The fall occurs in 0.9 second, and 5043 events are transmitted.

Centroids are computed as moving averages of a series of events. This can be done with high temporal resolution and low hardware cost. One possible hardware implementation is using a FIFO buffer, which stores the events that occur in a fixed time period. As a new event comes in, a computation cycle starts with removing the expired events and appending the incoming event in the buffer. All events in the buffer are averaged using Equation 1.

B. Evaluation of a fall

Fall detection is of particular importance when a person lives alone. This is one of the most critical scenarios in assisted-living environment and requires immediate attention. When more than one person are in the same room, one of them can call medical assistance if a fall hazard occurs.

The readout speed in the ATC vision sensor, i.e. *event rate*, is correlated with motion speed, size and light contrast in the scene. When the scene (lighting condition) is set, the event rate is useful for characterizing the motion in the scene. For example, a faster motion causes more events generated during a time period. Due to the different event rates, the number of events to be averaged varies depending on the motion in the scene. Figure 3 shows event responses directly measured from the ATC imager when it tracks a person's fall. Before the fall, the subject swings side-to-side in a normal walking manner. This side-to-side motion can be compared to the data collected during the fall. Each side-to-side movement generates approximately 460 event/s, while the fall causes 5600 event/s. The burst of events is due to the fast motion in the scene.

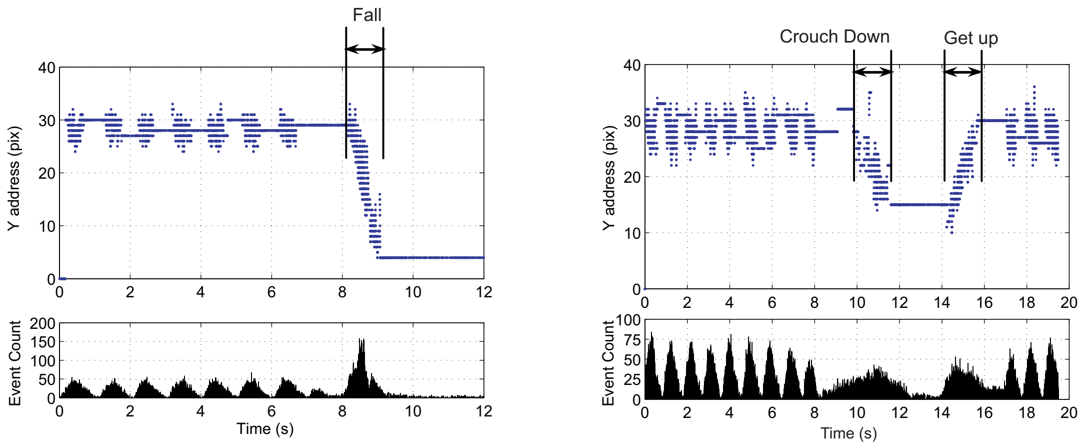
Figure 4 shows centroid event responses when the imager monitors a person's fall and crouch-down. A crouch-down scenario is considered because it is a motion similar to a fall but it happens in a longer time scale. In the experiment, a short walking motion with side-to-side oscillations happens before the fall and the crouch-down. In Figure 4(a), a fall causes the vertical address to decrease from 30 to 5 in 0.9s. An instantaneous event rate of 5600 event/s indicates significant changes in the scene. Figure 4(b) shows changes in the centroid when the imager observes a person crouching down, and then getting up. In this case, the centroid vertical coordinates reports a slower vertical velocity and a slower event rate. The Y address decreases from 30 to 15 in 2 seconds at an event rate of 310 event/s.

In order to numerically evaluate the dynamics in the ATC imager's event outputs, a centroid vertical velocity is computed in Equation 2.

$$V_y = \frac{\Delta y_c}{\Delta t} = \frac{(y_{c,i} - y_{c,j})}{t_i - t_j} \tag{2}$$

where $\Delta t = t_i - t_j < T$ and T is a fixed time period.

It is important to provide a measurement of the centroid that is invariant to the distance between camera and objects. In order to be invariant to distance, the vertical velocity (V_y) is divided by the height of the object, y_d, as shown in Equation 3. The height is the difference between the maximum and



(a) fall down

(b) crouches down

Fig. 4. Vertical address of the centroid events (Y_c) (a) when a person falls, and (b) when a person crouches down. In Figure (a), when the person falls, the centroid vertical address decreases by 25 pixels in 0.9 second. The event rate is approximately 5600 event/s. In Figure (b), when the person crouches down, the centroid vertical address decreases by 20 pixels in 2 seconds. The event rate is approximately 310 event/s. The fall causes faster decrease in the vertical address than the crouch-down.

minimum vertical address during a time period. The unit of a normalized vertical velocity is second^{-1}.

$$V_{y,norm} = \frac{V_y}{y_d} = \frac{\frac{\Delta y}{\Delta t}}{y_d} = \frac{(y_i - y_j)/y_d}{t_i - t_j} \qquad (3)$$

Figure 5 shows centroid vertical velocities in four scenarios: 1) a person crouches down; 2) a box free-falls; 3) a person falls forward; 4) a person falls backward. The velocities are normalized for distance using Equation 3. The centroids in Figure 5 show both positive and negative velocities because of the difference between the moving-average and the physical centroids when tracking a fast moving object. The ATC vision sensor stochastically fires events. The good estimation of physical centroid requires time to collect events. We set the averaging windows small in order to keep a high temporal resolution in tracking. This causes the events' average oscillates around the physical centroid when the vision sensor monitors a fast motion. For example, in a human fall case, the first average of events, which mostly describe the lower part of the person, is followed by the second average of events, which mainly describe the upper body. Even though the vertical address of the physical centroid decreases, the estimated centroid, i.e. average, could still show a positive velocity.

Notice that, in Figure 5, a fall shows a peak velocity of $-3s^{-1}$ in the vertical address decrease. The peak velocity is close to that of the free-falling box. When the vision sensor monitors a person crouching down, the centroid vertical velocity reaches a peak of $-1s^{-1}$, a value that is three times smaller than the fall's velocity. By estimating the peak vertical velocity, a fall is differentiated from other human behavior. More importantly, the fall shows more than three times peak-to-peak vertical velocity of the side-to-side swing before the fall, which is comparable to a normal walking.

We evaluated various scenarios using laboratory trials. The ATC vision sensor monitored a group of 3 people individually at a distance of 3 meters. The view was from the side and the vision sensor was mounted at a height of 0.8m, which is approximately the same height of a light switch on the wall. Each actor performed a predefined task list, including fall and other normal human behavior. The fall scenarios we tested in this work included a variety of fall types, such as fall forward, fall backward, and fall side-way. We also monitored other scenarios besides the falls, which frequently happen in home-assisted living applications. These tasks include a person walking, crouching down, sitting down and a cat walking. Pets are considered in this work because they are common company to the elders in home-assisted living.

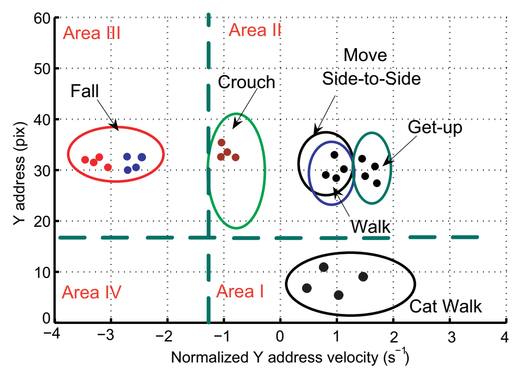


Fig. 6. Centroid address-event evaluation matrix of five common home-assisted living scenarios. The centroid vertical velocities are normalized for distance.

Experimental results are illustrated in Figure 6. They show the comparison of the initial centroid vertical address and peak centroid vertical velocity. The space in Figure 6 is divided into four areas. The centroids in Area II are those of human behavior, including crouching-down, walking and getting-up. They have higher coordinates than the pet centroids in Area I, which are closer to ground. The centroids in Area III are reported as fall hazards. They have higher peak vertical velocities than other human behavior. The centroids of people are around the middle of the camera view when people move. The centroids fluctuate between 30 and 40. This coincides with

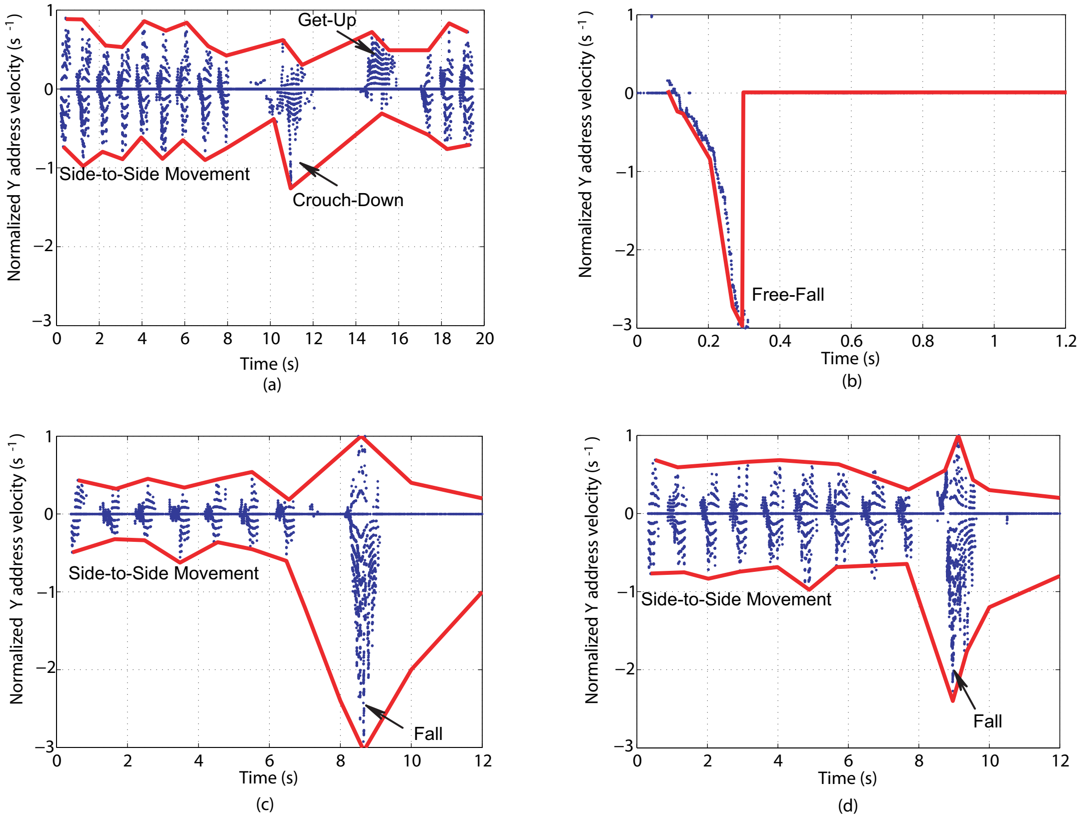


Fig. 5. Centroid vertical velocities ($V_{y,norm}$) in address-event streams demonstrate the difference among four scenarios: (a) a person crouches down; (b) a box free-falls; (c) a person falls backward; (d) a person falls forward. The velocities are normalized for distance between the object and the image sensor. The red lines in the figures outline the boundary of the peak vertical velocities.

TABLE I
STATISTICS OF THE EVENT RATE IN DIFFERENT EXPERIMENTAL SCENARIOS

Behavior	**Average Event Rate (event/s)**	**Variance (event/s)**
Walk	2100	11.3
Crouch Down	3500	15.2
Fall Down	5120	10.2
Sit Down	3150	15.2
No Motion	300	30.5

the expected height of human body. The pet's centroid moves at a lower coordinate, generally closer to ground. Both the fall and crouch-down demonstrate negative vertical velocities, which are due to the decrease in vertical address. The fall's centroid decreases at a much higher speed of $-3s^{-1}$. The fast decrease in the centroid vertical address distinguishes a fall from other normal human behavior.

Table I shows the statistics of the event rate when the vision system observes human behavior. When the system monitors a fall, the burst event rate is twice that of a person walking. This is because that the light intensity in more pixels changes in a unit time period when the system monitors a faster motion. When there is no motion in the scene, the event rate is as slow as 300 event/s with a larger variance of 30.5 event/s. This is due to the noise events associated with source/drain junction leakage in pixels' reset transistors.

IV. Conclusion

In this paper we evaluate an address-event temporal contrast fall detector. We propose using this detector to detect falls, a major health hazard in elderly home-care applications. A fall detection Matlab code is available at http://www.eng.yale.edu/elab/FallDetect.html.

The fall detector presented in this paper possesses the following features: (1) The motion detector is non-intrusive. Because of its low power budget and self-contained system property, a detector node is small in size. Implementing these nodes in elderly people's apartment will cause minimal change in their living patterns. (2) The motion detector protects patient's privacy. An address-event imager takes no image snapshot, and filters out detailed visual appearance of the patients. All images are processed locally. No data is released outside the node until an emergency is detected.

References

[1] U.S. Census Bureau Population Data Public Web Site, "URL http://www.census.gov/population/www/."

[2] T. Hori, Y. Nishida, and H. Aizawa, "Sensor network for supporting elderly care home," *IEEE Proceeding of Sensors*, vol. 2, pp. 575 – 578, April-June 2004.

[3] S. Allin, , Bharucha, J. Zimmerman, D. Wilson, and M. Roberson, "Toward the automatic assessment of behavioral disturbances dementia," in *International UbiHealth 2003)*, October 2003, pp. 25 – 29.

[4] E. Duthie, "Falls," in *Medical Clinics of North America*, April-June 1989, pp. 1321 – 1335.

[5] A. Sixsmith and N. Johnson, “A smart sensor to detect the falls of the elderly,” in *IEEE Pervasive Computng*, vol. 2, April-June 2004, pp. 42 – 47.
[6] P. Lichtsteiner and T. Delbruck, “A 64×64 AER logarithmic temporal derivative silicon retina,” in *IEEE Ph.D. Research In Micro-Electronics & Electronics (PRIME)*, 2005, pp. 202–205.
[7] Lichtsteiner and T. Delbruck, “64×64 event-driven logarithmic temporal derivative silicon retina,” in *2005 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, 2005, pp. 157–160.