Fall Detection Application on an ARM and FPGA Heterogeneous Computing Platform

Hong Thi Khanh Nguyen¹,³, Cecile Belleudy¹ and Pham Van Tuan²

LEAT, University of Nice Sophia Antipolis, Nice, France¹
Dept. of Electronic & Telecommunication Engineering, University of Science and Technology.
The University of Danang, Vietnam²
Dept. of Electrical Engineering, College of Technology, the University of Danang, Vietnam³

ABSTRACT: Heterogeneous computing platform, Zynq 7000 all programmable system-on-chip, not only accomplishes high efficiency solution in emerging the power consumption, execution time for implementing the Fall Detection application but also takes the advantage of Open source Computer Vision (OpenCV) libraries. The main goal of this work is to design and implement the Fall Detection Application on ARM Cortex A9 processor of Zynq Platform. Besides, real power consumption, estimated execution time and calculated energy are extracted from the implementation. Based on the observed measurements, pre-processing module based on morphology filter which occupies most execution time will be replaced by Sobel Filter. Then Sobel Filter will be implemented on hardware (FPGAs) part of the platform. The result analysis leads to a potential low power exploration of HW/SW co-design flow for performance improvement of the fall detection application.

KEYWORDS: Fall Detection; power consumption; execution time; Hardware solution; HW/SW co-design.

I.INTRODUCTION AND RELATED WORKS

It is necessary to have systems being able to automatically monitor human activities in order to reduce the pressure on training and expanding force for health solutions. As a result, it is important to develop an automated Fall Detection application to prevent fall risk of elderly and rehabilitants and provide immediate help to them.

1. FALL DETECTION APPROACHES

In general, automatic fall detection can be performed by many different techniques:

- Indoor sensors: Arrays of infrared sensors [1], laser scanners [2] and even floor vibration and sound [3].
- Wearable sensors: Accelerometers, gyroscopes or even barometric pressure sensors [4].
- Video systems: mono camera [5], multi cameras [6], Kinect sensors [7].

Among them, the wearable sensors enable to capture the high velocities, which occur during the critical phase, and the horizontal orientation during the post fall phase. However, in these methods the users have to wear the device all time, and therefore, if it is inconvenient, it could bother them. Additionally, such systems require recharging the battery frequently, which could be a serious limitation for practical application. On the other hand, video systems enable an operator to rapidly check if an alarm is linked to an actual fall or not. A block diagram of Fall Detection based on video processing is described in Fig. 1.

An interesting object (i.e. the foreground) will be extracted from the background of each video frame in a video clip. This detection technique is called background subtraction. This approach detects the foreground by thresholding the difference between the current frame and the modelled background in pixel by pixel process. After blobbing and
smoothing the object in Filter module, this result will be tracked by 2D modeling such as point tracking, kernel tracking (rectangle, ellipse, skeleton...), or silhouette tracking. Then, the feature extractions to understand object gestures are calculated based on one of these modeling. The problems are that these features must encapsulate unique characteristics for the same activity but acting by different people. In order to avoid misdetection and false alarms for this system, it not only depends on the techniques but also confronts some challenges such as

- Dynamic background: in some cases there are many moving objects in an indicated scene and how to extract the exact object.
- Brightness: The changes of light intensity at different times of the day or the suddenly turning of lights also affect the background processing.
- Occlusions: are often considered as artifacts, undesirable in many image motion computations. We have argued that occlusions show the related important information about the camera motion (position and angle of camera) and scene structure (static or dynamic environment; brightness, colour costume of objects).
- Static object: when the object is sleeping, sitting or walking calmly, it is easy to mistake the extraction of object.

After the object features are tracked and extracted, the problem of the system has to understand the meaning of the object actions through their features in the recognition block.

2. IMPLEMENTATION OF FALL DETECTION APPLICATION

Recently, a video processing framework called OpenCV is regularly developed on an open source library for computer vision of Intel. However, how to implement video applications with OpenCV on a hardware platform are really big challenges. In the work of Floris Driessen [8], he proposed the combination of embedded processors and customized accelerators on heterogeneous computation platform - the Zynq-7000 all programmable system-on-chip -, which offers a high-end embedded processor combined with field programmable gate array (FPGA) based reconfigurable logic. Moreover, another fall detection system which was implemented on Terasic's DE2-115 development board including Altera Cyclone IV (EP4CE115) FPGA device, a 5 megapixels CMOS camera sensor and a LCD touch panel. This system was also designed with highly exploitation of the parallel and pipeline architecture of the FPGA [9].

Shimmer technology and applied the orthogonal matching pursuit (OMP) algorithm for advanced data compression was shown in [10]. This system has been simulated and implemented on the Virtex-5 and Zynq 7 (FPGA) using Vivado High Level Synthesis tool. It is used to estimate the area, power and computation time for the detection of falls with different scenarios. In the other side, the combination of Kinect and Wireless Accelerometer is used to extract object of images in a dark room [11]. This system was implemented on the PandaBoard ES with real-time indication.

In the next section, the research objective is mentioned. Fall Detection application will be described in Section III with four steps: object segmentation, filter, feature extractions and recognition. In Section IV an insightful experiment of implementation and evaluation is described. Finally, Section V contains the conclusions of this paper.

II. RESEARCH OBJECTIVE

To deploy applications of embedded systems on heterogeneous platforms which contain various computation units such as Field Programmable Gate Array (FPGA), microprocessor and graphics processing unit GPU, we need to determine a specific activity of design phase. Then, the designers are going to profile the partition for software (SW) or hardware (HW) implementation.

In common practice, deployment decisions are taken at an early stage of the design phase, and it branches into two separated flows: HW and SW design flows. Then, during the implementation phase they evolve separately until the final integration. In this scenario, the design phase is affected by issues such as HW or SW flow interruptions, redesigns and unplanned iterations which negatively impact the overall development process in terms of efficiency, quality and costs, and the system lifecycle. Especially, this method is time-consuming to separate implementation on HW and SW for whole application.

In our study, we extract the execution time, power consumption of whole fall detection application which is deployed on ARM processor of Zynq-7000 AP SoC platforms. After that, the modules in Fall Detection which take the most execution time will be implemented on HW (with FPGAs).

III. FALL DETECTION APPLICATION
Object segmentation block is responsible for detecting and distinguishing between moving objects and the rest of the frame which is called background. Background subtraction method is applied in this study. A pixel is marked as foreground if
\[ |I_i - B_i| > \tau, \quad (1) \]
where \( I_i \) is current video frame; \( B_i \) is modelled background frame and \( \tau \) is a “predefined” threshold.

The updated background is estimated as:
\[ B_{i+1} = \alpha I_i + (1 - \alpha)B_i \quad (2) \]
where \( \alpha \) is kept small to prevent artificial “tails” forming.

2. FILTER

There are some supporting methods to improve the quality of image from the object binary image such as Morphology Mathematic (MM), Edge Detection Filter (Sobel, Canny, Prewit Filter). In this module the object will be smoothed, blobbed and the noise will be removed.

Morphology Mathematic (MM) is a technique for the analysis and processing of geometrical structures. It is also used in morphological image processing by transforming images according to the characterizations of object. Some of MM are dilation, erosion, opening, closing or the combination of these.

Sobel edge detection algorithm is the most commonly used techniques in image processing for edge detection [12]. In this paper two types of Sobel operators, horizontal, vertical, were used. The operator calculates the gradient of the image intensity at each point, giving the direction of the largest possible increase from light to dark and the rate of change in that direction. The Sobel kernels are given by
\[ G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (3) \]

Here the kernel \( G_x \) is sensitive to changes in the \( x \) direction, i.e. edges that run vertically, or have a vertical component. Similarly, the kernel \( G_y \) is sensitive to changes in \( y \) direction, i.e. edges that run horizontally, or have a horizontal component. The two gradients computed at each pixel \((G_x, G_y)\) by convolving with above two kernels can be regarded as the \( x \) and \( y \) components of gradient vector. This vector is oriented along the direction of change, normal to the direction in which the edge runs. Gradient magnitude and direction are given by:
\[ G = \sqrt{G_x^2 + G_y^2} \quad (4) \]
An approximate magnitude is computed using:
\[ |G| = |G_x| + |G_y| \quad (5) \]
The angle of orientation of the edge (relative to the pixel grid) giving rise to the spatial gradient is given by:
\[ \theta = \arctan \left( \frac{G_y}{G_x} \right) \quad (6) \]

3. FEATURE EXTRACTION

ELLIPSE MODEL is usually used for tracking object because it is easy to fit an ellipse around the object. In this study, there are three following parameters needed to set an ellipse: Centroid of ellipse, Vertical angle of the object, Major and Minor Axis of the Object (Fig. 2)

![Fig. 2. Ellipse model](image)

**Centroid of ellipse:** for each binary frame, centroid coordinate of an ellipse \( O_x, O_y \) is determined: Abscissa \( (O_x) \) and ordinate \( (O_y) \) are average of the entire \( x \) coordinates and the entire \( y \) coordinates of the white pixels.
Vertical angle of the object: after determining the centroid coordinate, the system calculates the angle between major axis ellipse and horizontal line (θ or current angle is calculated by Equation 7).

\[ \theta = \frac{1}{2} \arctan \left( \frac{\sum \sum x \cdot y \cdot P(i,j)}{\sum \sum x^2 \cdot P(i,j) - \sum \sum y^2 \cdot P(i,j)} \right) \]  

(7)

Where, i, j is position of pixel (i=1...width of frame, j=1...height of frame), x = i – O_x and y = j – O_y (O_x, O_y: position of centroid), and P(i,j): value of pixel (i,j).

**Major and Minor Axis of the Object**

a and b are respectively semi-major axis and semi-minor axis of the ellipse. \( d_1 \) and \( d_2 \) are distances from O to \( O_1(x_1, y_1) \) and \( O_2(x_2, y_2) \), respectively. \( O_1 \) and \( O_2 \) are calculated: \( (x_1, y_1) \) and \( (x_2, y_2) \) coordinates are the average of the entire x coordinate and y coordinate of the white pixels \( W(W_x, W_y) \) satisfying 2 following conditions: y-coordinate of these white pixels \( W_y \) are smaller than the y-coordinate of centroid \( O_y \).

These \( W \) are in the limited angle so that 
\[ |W_0\theta - \theta| < 60^\circ \] for \( O_1 \) and 
\[ |W_0\theta - (\theta + \pi/2)| < 60^\circ \] for \( O_2 \)

Finally, major and minor axes are calculated: \( a = 2d_1 \), \( b = 2d_2 \).

**FEATURE EXTRACTION:** defines 5 features

**Current angle:** cf. Equation 7.

**Coefficient of motion (Cmotion)**

At the same position versus time gray pixels are white pixels of the previous frame. Brightness of the gray pixel increases versus the time frame by frame containing white pixel at the same position. The gray frame uses to determine the object’s motion rate also known as Motion History Imaging (MHI). The motion coefficient of object Cmotion is determined by:

\[ C_{motion} = \frac{\text{Gray pixel}}{\text{Gray pixel} + \text{White pixel}} \]  

(8)

Where “White pixel” is the number of white pixels; “Gray pixel” is the number of gray pixels. Cmotion’s value must be in range of [0, 1].

**Deviation of the angle (Ctheta)**

Ctheta is standard deviation of 15 angles θ from 15 successive frames. Ctheta is usually higher when a fall is occurring.

**Eccentricity** at current frame is computed by:

\[ e = \sqrt{1 - \frac{b^2}{a^2}} \]  

(9)

Where e: eccentricity; a, b: semi-major and semi-minor axis of the ellipse; e is smaller when direct fall happens.

**Deviation of the centroid (Ccentroid)** is defined as standard deviation of 15 ordinates from 15 successive frames. Ccentroid rapidly decreases when the fall occurs.

4. **RECOGNITION BASED ON TEMPLATE MATCHING**

Based on the direction of falls and the type of falls, four models have been built to detect falling accidents. The first model is face fall, in this case Cmotion, Ctheta, Ccentroid is high but Theta is low. The second model is cross fall, with high Cmotion and Ctheta, whilst Theta, Ccentroid and Eccentricity have medium values. In the next model, the victim fall in the perpendicular direction to the camera, consequently Theta is almost constant, Cmotion is in average, the Eccentricity is low while Ccentroid is quite high. The other cases are included in the last model. The features are combined with each other depending on the fall models, and the thresholds are selected from the survey of training videos.

**IV.IMPLEMENTATION & EVALUATION**

We design this system in High Level Languages specified in C/C++ integrated OpenCV and cross-compiled along with libraries which implement the communication Application Programming Interfaces (APIs) and runtime layer using gcc/g++ toolchains. These toolchains generate an .elf file which is downloaded to the processor ARM Cortex A9 on Zynq platform supported by SDK tools. Moreover, this system is emerged the power consumption; estimated the execution time of each frame from input data on this processor. The frequency of this core is set up at 666 MHz.
The execution time is extracted from available time.h library. The measurements of power and thermal are taken by the Fusion Digital Power Designer GUI. The TI USB Adapter is connected to controller Power Management Bus (PMBus) on Zynq platform and to PC for displaying measurement results, as shown in Fig. 3. In this study, the input video is recorded by the Camera Web Cam-Philips SPC 900NC PC that is mounted on the wall at the distance of 3m from the floor. The data are captured with both resolution 320x240 pixels and 680x360 pixels.

1. SOFTWARE IMPLEMENTATION

CLASSIFICATION PERFORMANCE:

The DUT-HBU database [5] is used in this system. All video data are compressed in .avi format and captured by a single camera in a small room with the changeable conditions such as brightness, objects, direction of camera, etc. DATABASE: the fall direction is subdivided into three basic directions

- **Direct fall**: object falls face to the camera.
- **Cross fall**: occurs when the object falls cross the camera.
- **Side fall**: the object perpendicularly falls to both sides of the camera.

In terms of non-fall videos, usual activities which can be misrecognized with fall action such as lying, sitting, creeping, bending are also classified into three directions above.

In this study, we create two dataset (as shown in Table 1):

<table>
<thead>
<tr>
<th>Classification of videos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database</td>
</tr>
<tr>
<td>Train</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Fall</td>
</tr>
<tr>
<td>Side-Fall (F1)</td>
</tr>
<tr>
<td>Direct-Fall (F2)</td>
</tr>
<tr>
<td>Cross-Fall (F3)</td>
</tr>
<tr>
<td>Non Fall</td>
</tr>
<tr>
<td>Bending (N1)</td>
</tr>
<tr>
<td>Lying (N2)</td>
</tr>
<tr>
<td>Creeping (N3)</td>
</tr>
<tr>
<td>Sitting (N4)</td>
</tr>
<tr>
<td>Sum</td>
</tr>
</tbody>
</table>

*Train set*: Clear data consists of videos which have stable background. These videos are captured in a small room under good brightness condition. The object is not obscured by furniture in the room. Train set contains 21 videos of fall and 26 videos of daily activities.

*Test set*: Contents and activities in the video clips for testing are basically performed similar to the ones for training, just a small difference of environment condition. In each clip, there is only one object with stable background and include 21 fall videos and the rest is 33 videos.
CLASSIFYING EVALUATION:

ROC (Receiver Operating Characteristics) is one of the methods to evaluate the efficient and accuracy of a system by calculating the Precision (PR), Recall (RC) and Accuracy (Acc), see in the Equation 6.

\[
PR = \frac{TP}{TP + FP} \quad ; \quad RC = \frac{TP}{TP + FN} \quad ; \quad Acc = \frac{TP + TN}{TP + TN + FP + FN}
\]  

Where TP, TN, FN, and FP are defined as follows:

- **True positives** (TP): amount of fall actions which are correctly classified as fall.
- **False positives** (FP): amount of non-fall actions which are wrongly considered to be fall.
- **False negatives** (FN): amount of fall actions which are wrongly rejected and classified as non-fall actions.
- **True negative** (TN): amount of non-fall actions which are correctly classified as non-fall.

CONFUSION MATRIX

A confusion matrix gives information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. The following Table 2 shows the confusion matrix for two classes which are classified FALL or NON FALL for both dataset of Train and Test which are implemented on ARM Cortex A9 of Zynq-7000 AP SoC platform.

### Table 2 Confusion matrix

<table>
<thead>
<tr>
<th>Database</th>
<th></th>
<th>Video</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td></td>
<td>F1</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F2</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Non Fall</td>
<td></td>
<td>N2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N4</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

From the confusion matrix, the Recall, Precision and Accuracy are calculated and shown in Fig. 4. The result of clear data in Train set is higher than Test set in all Recall, Precision, and Accuracy. The reason is that Template Matching uses “hard threshold” and the combination of features is quite simple to detect a fall event. Four models of the fall are not enough to describe all falls may occur.
2. THE MEASUREMENT

As shown in the Fig 5, the mean of total execution time of Fall Detection application is approximately 0.107s/frame. The Frame Filter module based on Morphology Filter takes around 2/3 times of total execution time. The similar observation has been obtained when using higher resolution input video of 680x360 pixels. In which the execution time is 0.221ms/frame for Frame Filter and 0.3s/frame for total execution time.

In this case, the measured power consumption of whole Fall Detection application is closed to 0.403W. Therefore, the energy per frame is multiplied by the power consumption (P) and total execution time (T) as following:

\[ E = P \times T = 0.403 \times 0.107 = 0.043 \text{ (J/frame)} \]

As the result of this experiment, the frame rate of this system is calculated by:

\[ \text{Frame rate} = \frac{1}{0.107} = 9.3 \text{ fps} \]

It is found that the overall of this system does not keep on operation at 24 frames per second. Thus, this parameter will have an effect on the recognition ability of this system. That is also a big challenge in video design to get the reasonable precision, accuracy, performance when we change the offline procession by online one.

![Fig. 5. Comparison execution times at two image resolutions](image)

3. HW IMPLEMENTATION

The Zynq®-7000 family is based on the Xilinx All Programmable SoC architecture. These products integrate a feature-rich dual-core ARM® Cortex™-A9 based processing system (PS) and 28 nm Xilinx programmable logic (PL) in a single device. The ARM Cortex-A9 CPUs are the heart of the PS and also include on-chip memory, external memory interfaces, and a rich set of peripheral connectivity interfaces [13]. As a result, the Zynq-7000 AP SoCs are able to serve a wide range of applications including video application.

As discuss in the section II and from the observed results in Fig. 5, the Frame Filter is chosen for implementation on HW. In order to reduce the execution time, the Sobel Filter is replaced by the Morphology Filter. In Sobel Filter, two types of Sobel operators, horizontal and vertical, were used. The operator calculated the gradient of the image intensity at each point, giving the direction of the largest possible increase from light to dark and the rate of change in that direction [14]. By this way, the power and execution time of this module are estimated by using the Vivado_HLS tool.

From the power consumption and execution time shown in Table 3, the energy consumption per frame is calculated as

\[ (*) \quad E = P \times T = 1.65 \times 10^{-1} \text{ (J/frame)} \]

\[ (***) \quad E = P \times T = 1.158 \times 10^{-3} \text{ (J/frame)} \]

With the same function for Frame Filter, the Sobel Filter implemented on HW spends forty times less energy comparison with Morphology Filter implemented on ARM processor.

After that the previous experiment is performed to select which modules or algorithms are candidates for HW or SW implementations. Therefore, the HW/SW architecture is proposed for this Fall Detection application, in which Frame Filter based on Sobel Filter is selected to obtain the better performance and low power consumption.
Table 3 Power and execution time of frame filter on HW

<table>
<thead>
<tr>
<th>Frame Filter</th>
<th>Power Consumption (mW)</th>
<th>Execution Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>680x360(*)</td>
<td>198</td>
<td>8.34</td>
</tr>
<tr>
<td>320x240(**)</td>
<td>193</td>
<td>6.15</td>
</tr>
</tbody>
</table>

V. CONCLUSION

In this paper, a Fall Detection Application is implemented on ARM Cortex A9 processor of Zynq-7000 AP SoC platform with two video input resolutions. Its recognition performance has been evaluated in terms of recall, precision and accuracy. The SW implementation of the application shows an average accuracy of almost 80% in different testing conditions. From the extracted measurements on real power consumption and estimated execution time, the Sobel filter has been selected for HW implementation on FPGAs of the same platform. The related parameters such as power consumption, execution time and energy have been estimated by using Vivado_HLS. The overall observation leads to a suggestion of SW/HW co-design for taking advantages of technical features of FPGAs which accelerate digital signal processing algorithms. The investigation of an optimized HW/SW architecture will be studied on this platform.

REFERENCES


Hong Thi Khanh Nguyen was born in Da Nang city, Vietnam in September 1979. She received the Engineering degree in Electronic and Telecommunication Engineering at the Polytechnic University of Da Nang in 2002 and the Master degree in Electronic Engineering at The University of Da Nang in 2007. Currently, she is a PhD student in the Electronics, Antennas, Telecommunications Laboratory (LEAT), the University of Nice Sophia Antipolis, France.

Her main focus is on designing and developing fall detection algorithms. She proposed the optimization method for the hardware resources and power consumption of FPGAs. Her ongoing research interests involve in image and video processing, and multi-core processor architectures.
She has several publications in relation to these areas in scientific conferences such as 2013 RUNSUD (UNS-Scientific meetings UD), 2014 International Conference on Advances in Electronics Engineering (ICAEE 2014), and 2014 IEEE Faible Tension Faible Consommation (FTFC).

Cécile Belleudy joined the Electronics, Antennas, Telecommunications Laboratory (LEAT) in the University of Nice Sophia Antipolis as an assistant professor. Her current work focuses on studying and developing low power strategies (off-line, on-line) for embedded system. More especially, she addresses the problem of low power RTOS for monoprocessor and multiprocessor platforms and energy aware memory allocation. She has also developed new approach for managing energy in wireless sensor network at different level (PHY, MAC, and Routing) and for autonomous node (Energy harvesting). In the field of low power SoC design methodologies, she participates in national and European collaborative projects.

Tuan.V. Pham was born in Hoi An ancient town, Vietnam. He received Engineering and MSc. degrees in Electronics & Telecommunications and Dr. Techn. degree in Electrical Engineering from Danang University of Technology (1999), Hanoi University of Science and Technology (2001) and Graz University of Technology (2007). Tuan was granted with a post-doctoral scholarship at the SPSC Lab of TUGraz from 2007 to 2008 and promoted to Associate Professor at ETE Department, DUT since 2012. He has been a lecturer at DUT since 1999, Vice-Chair of Electronic & Telecommunication Engineering Dept., DUT since 2010 and Vice-Director, Center of Excellence, DUT since 2011.