Abstract

This paper presents a low-cost, low-power automated home-based surveillance system, capable of monitoring activity level of elders living alone independently. The proposed system runs on an embedded platform with a specialised ceiling-mounted video sensor for intelligent activity monitoring. The system has the ability to learn resting locations, to measure overall activity levels and to detect specific events such as potential falls. We build a probabilistic spatial map of resting locations using the head position of the subject, represented as cluster centres discovered by K-means in the camera view space. A novel edge-based object detection algorithm capable of running at a reasonable speed on the embedded platform has been developed. The head location of the subject is also estimated by a novel approach capable of running on any resource limited platform with power constraints.

1. Introduction

The United Nations Department of Economic and Social Affairs predicts a worldwide increase in life expectancy coupled with a decline in the number of children and total fertility [18]; thus a drastic reduction in support provided by young individuals working in home and health care for elders. This calls for an automated home-based system, capable of unobtrusive monitoring of the status of elders living alone independently. Ambient intelligence is an emerging discipline that brings intelligence to our everyday environments and makes those environments sensitive to us [4]. The main aim of Ambient Assisted care is to provide quality of life (mainly to the elderly) in all stages of their life.

Many elders living at home on their own may have more than one chronic disease requiring ongoing care. Earlier research into offering some form of holistic health management has focused on medical monitoring like blood glucose levels, heart rate, weight, blood pressure and other similar symptoms [10]. Over the past decade, new research trends have focused on unobtrusive monitoring of behaviours, including activity levels, falls and adherence to health behaviours. Smart home technology developed specifically for Ambient Assistive care or monitoring health behaviours is faced with three basic challenges [10]; the hardware or physical device, the software that runs on the hardware, and the social issues related to the use of technology.

In this paper we present a smart home sub-system for monitoring activity levels for elders, which seeks to address the three basic Ambient Assistive technological challenges. The system is part of a wider development; an end-to-end assistive care solution including multiple sensors. It uses special particle ceiling-mounted intelligent activity monitoring sensor to infer the level of activity and behaviour. The system has the ability to characterise behaviour as learned resting locations, to measure overall activity levels and to detect specific events such as potential falls. The design has been accomplished on a low-cost embedded platform, which minimizes on-board power consumption.

2. Related Work

Zhou et al. [20] presented an automated activity analysis and summarization for the monitoring of the elderly. They used an adaptive learning method to estimate the physical location and moving speed of a person from a single camera view. Chan et al. [3] presented a multi-sensor home monitoring system to help elderly people by observing mobility changes indicative of abnormal events. The design assesses changes in occurrence, time and duration from a statistical perspective. People with dementia often have low physical activity and some sleep problems [17]. The daily life activities and sleeping conditions has been used by Suzuki et al. [17] for the early detection of dementia. Adami et al. [1] described a system for unobtrusive detection of movement in bed that uses load cells installed at the corners of a bed. The system focused on identifying when a movement occurs based on the forces sensed by the load cells.

Cuddihy et al. [5] presented an Automatic Inactivity Detection system capable of constructing an individual models
of normal activity within a home using motion sensor data. Alerts are generated when a period of inactivity exceeds a normal length for a particular residence. Gil et al. [7] used data mining and visualisation to illustrate activity from sensor data that can reveal individual life patterns and changing circumstances. They measure activity to determine whether it would be possible to build a busyness model of a person’s life at various levels of granularity.

Hayes et al. [8] presented a system for continuous and unobtrusive monitoring capable of observing changes in physical behaviour over time. Similarly, Shieh et al. [16] presented an automatic monitoring system where activities are monitored by infrared positioning sensors. Williams et al. [19] presented a design, implementation and evaluation of a distributed network of smart cameras whose function is to detect and localise falls. Their design demonstrates that a distributed low-power system can perform adequately for the detection of fall when monitoring an elderly.

Dickinson and Hunter [6] presented a novel method for detecting unusual modes of behaviour in video surveillance data, suitable for supporting home-based care of elderly patients. Their approach is based on learning a spatial map of normal inactivity for an observed scene to detect unusual patterns of inactivity. Magno et al. [13] presented a multi-modal video sensor node designed for low-power video surveillance able to detect changes in the environment. The system has a CMOS video camera and a Pyroelectric InfraRed (PIR) sensor exploited to reduce significantly the power consumption of the system in absence of events. The same combined sensor system has successfully been used to detect abandoned and removed objects [14]. The indoor monitoring sub-system proposed in this paper is based on a similar concept as in [13, 14], to reduce the overall power consumption, but without the use of a PIR. Our system is such that the camera will be operational only when there is movement in the observed environment.

3. System Architecture

The proposed energy efficient assistive care sub-system has been developed on an ultra low cost small form factor, high performance leopardboard [9]. Figure 1 is a block-level diagram of the proposed system. The main processor is TMS320DM365 (ARM926EJ-S), an ARM9-based RISC processor offering speeds up to 300MHz. The ARM926EJ-S is a 32-bit processor core that performs 32-bit and 16-bit instructions and processes 32-bit, 16-bit, and 8-bit data. The board is also packaged with the Micron DDR2 SDRAM, it is a +3.3V powered, 8M×16×8 banks, 1Gb DDR2 SDRAM totalling 128MB. The camera board used in this development is the Aptina 1/2.5” CMOS sensor MT9P031 with 2592×1944 pixels. The kernel that runs on the leopardboard has been developed with RidgeRun Embedded Linux Software Development Kit (RRSDK) [15].

For unobtrusiveness, a high-mounted fish-eye lens with viewing angle of 105 degrees has been used. This effectively reduces the number of cameras needed in any single living environment. The lens of the CMOS sensor has been replaced with a 1.9mm (fish-eye) lens. Figure 2 is the fish-eye lens used in this development with a sample output image when mounted high up in the ceiling. A powerful 5-Channel Power Management IC with 2 step down converters and 3 low input voltage LDOs chip TPS65053 is provided on the leopardboard 365. The board is powered by +5VDC power supply and consumes less than 2W, which includes the MT9P031 camera board running at 30fps.

The main goal of our system is to be able to determine the presence of a single subject (an elder) and estimate the location in the room with minimal power consumption. Apart from the constraints (in terms of memory and processing power) of the embedded development platform chosen for this implementation, the design is also constrained by a low-power budget. This is to enable the system run for longer period when battery-powered. To satisfy these requirements, a novel edge-based differencing algorithm has been developed and implemented on the embedded ARM9 processor, rather than the standard background differencing algorithm [2], which requires the camera and the leopardboard to run continuously to update the background model. The system described in this paper uses edge-based differencing, as edges are less sensitive to illumination changes. A low power motion detector is used to trigger the camera only when activity occurs.

4. Video-based Analysis

This section describes the video-based analysis performed on the leopardboard for monitoring the behaviour of an elderly person. The system has two basic sub-systems: the detection and behavioural sub-systems as shown in figure 3. Both sub-systems are activated every time the camera switches to RUN mode when there is enough activity in the scene. The detection sub-system has three major phases: edge detection, training (re-training) and object detection. The edge detection phase uses the Sobel edge [12], which
calculates the gradient of the image intensity at each point.

The Sobel edge detector implemented on the leopard-board (ARM9) uses a pair of $3 \times 3$ convolution masks as shown in figure 4, $G_x$ is for estimating the gradient in the x-direction (columns) and $G_y$ for estimating the gradient in the y-direction (rows). The mask is slid over the image, manipulating $3 \times 3$ pixels at a time. The magnitude of the gradient is generally computed as the square-root of the sum of squares of the horizontal and vertical gradient as shown in equation 1. To enable ease of implementation of the gradient magnitude $G$ on the ARM9 processor, an approximation has been used as shown in equation 2. For each pixel in the image (excluding the boundary pixels), the values of $G_x$ and $G_y$ are estimated using the convolution masks shown in figure 4. The sum of the two values is compared with a threshold value and used to determine if that pixel is an edge pixel or not.

$$|G| = \sqrt{G_x^2 + G_y^2}$$  \hspace{1cm} (1)
$$|G| = |G_x| + |G_y|$$  \hspace{1cm} (2)

To establish the background model $M$ using the edge-map $E$, extracted from an image $I$, a burn-in period is used in the training phase. The burn-in period defines the minimum number of frames that can be used to model the background. During this period, the accumulated value $A_i$ of the number of edges that appear at every pixel location for all the frames is estimated. The burning period is also used to estimate the average number of edges $E_{avg}$ that appear in any particular frame.

$$A_{i+1} = \begin{cases} 
A_i + 1 & \text{if } I_i \text{ is an edge pixel} \\
A_i & \text{otherwise}
\end{cases}$$

where $i$ is the $i^{th}$ pixel in the image $I$. The training phase is activated by the following three conditions:

1. when the system is powered on for the first time,
2. when the number of edges in the current edge-map $E$ is greater than 25% of the average number of edges $E_{avg}$, acquired over the previous training phase, and
3. when the number of edges in $E$ exceeds 6.25% of $E_{avg}$ continuously for the same number of frames as in the burn-in period.

The retraining of the systems becomes necessary when the camera drifts slightly or when too many objects appear in the scene. The retrain conditions (2 and 3) allow the system to handle major changes; e.g. resetting the camera or rearranging furniture and significant lighting changes. The percentages used in conditions 2 and 3 are empirical values determined experimentally. After the burn-in period, the background model $M$ is extracted using the accumulated values at each pixel location. The use of edges rather than intensity values to model the background minimizes the need to continuously update the model to handle minimal illumination changes.

$$M_i = \begin{cases} 
1 & \text{if } A_i > \text{threshold} \\
0 & \text{otherwise}
\end{cases}$$

where $M_i$ is the $i^{th}$ element of the background model $M$ taken at pixel location $i$.

To extract the edges of any moving object, a simple edge-based differencing algorithm is used. The edge-map extracted from the current frame $E$ is compared with the background model $M$ to extract any new edges that appears in
the scene. Figure 5 is a sample output of the object detection phase. The edges of the input image is compared with the background model to extract the new edges, which forms the outline of the moving object. The images in figure 5 are actual processed images on the leopardboard.

The central location of the moving object is estimated in the object location phase of the behavioural sub-system. To estimate the area where most edges are concentrated, a tiling operation is adopted. Similar to a compression operation, the entire image is divided into blocks each of size 16×16 pixels. Thus for a VGA sized image there are 40×30 blocks. The number of edge pixels in each block is compared with a threshold value to determine if the resulting block is set or cleared. This is demonstrated in figure 6 with two 8×8 blocks, each with 46 pixels set as edge pixels. If the threshold is less than 46 (as shown to the left), the block is set. Similarly, if the threshold is greater than 46, the resulting block is cleared, as shown to the right in figure 6.

A block-based horizontal and vertical histogram is generated to estimate the area with highest edge blocks. The intersection block (blue block in figure 7) is the block (or area) with the maximum number of edges. This is then used to establish the central location of the moving object. The algorithm is very simple and well-suited for the low-powered embedded platform. The vertical block-based histogram is estimated by accumulating the number of blocks set for each row in the grid. Similar, the horizontal block-based histogram is estimated by accumulating the number of blocks set for each column in the grid. The position of the subject is estimated using the maximum vertical and horizontal block-histogram peaks. To compute the level of activity $T$ between two frames, a distance measure (Manhattan distance) given as $T = |x_t - x_{t-1}| + |y_t - y_{t-1}|$ is used. This is accumulated over a period of time to measure the overall activity level.

The central location $(x_t, y_t)$ of an object is collected over a period of time and used in the K-means clustering phase to determine resting positions. In K-means clustering, given a set of $n$ data points in d-dimensional space and an integer $K$, the problem is to determine a set of $K$ points, also called the centres, so as to minimise the mean squared distance from each data point to its nearest centre [11]. The value of $K$ is pre-defined depending on the assumed number of resting positions in the observed environment. The collected central positions are used to initialize the values of $K$ and updated until the changes in cluster centres falls below a specified threshold. After obtaining the K cluster centres with acceptable radius, an alarm is raised if an object stays outside the defined region for more than a specified period (ie. beyond a threshold distance from a cluster centre).

5. Retraining and Energy Analysis

To test the performance of the monitoring system implemented on the DM365 leopardboard, four different rooms have been used. The leopardboard has been used to collect data continuously over period of two weeks. The robustness and retraining capabilities of the systems alongside its energy efficiency has also been evaluated. To test the ability to retrain, the system was deployed in a room with wide glass window exposed to external day light over a whole weekend. Over the period of three days, Friday to Monday, the system automatically retrained six times. It should be noted that because the room was not used over the weekend, the only cause of retraining was changes in external light and reflection in the room. Also the reflection in the room changes as the sun changes position during the day. However, the other three rooms used for testing have at least a single occupant for a minimum of nine hours everyday.

The object detection and location capabilities of the system has also been tested and compared with a PC based implementation. As shown in figure 8, the moving object
in the input image to the top is detected and its centre extracted as shown in the bottom image. The block-based vertical and horizontal histogram is used to estimate the centre of the moving object, shown in red on figure 8. Object locations extracted from all four test rooms have been examined to determine if the resting locations are correctly identified. Segments of the processed video streams on the leopardboard are visually compared with the PC based implementation to verify the object centre is correctly labelled.

Figure 9 is a 3D graph of object locations in one of the experimental rooms with data collected over the period of 8am to 11am. The plot shows the frequency of occurrence of movement at any particular location in the 2D image plane as shown in figure 10. In 10, we show three resting locations. The location with maximum frequency is the area marked with a large green blob. This area has a kettle, which is used by other office occupants. Hence the level of movement around the kettle is high early in the morning. The area marked with a smaller green blob is the second highest in the scene followed by the area in blue.

The power efficiency of the entire system is demonstrated by the number of times the system needs to process an entire image frame or retrain in order to correctly identify moving objects. For a period of 3 hours 20 minutes in a normal day settings, the system retrained twice to be able to detect objects correctly. The power consumption of the leopardboard with the 5 mega-pixel camera is 2 watt (2 joules per second). Thus if the camera should run continuously for the entire 200 minutes, it will consume 24 kilowatt (kW). Within this period of time the camera recorded 5,224 activities with a significant level, thus consuming 10.4 kW of power, a 50% savings in power. This is the worse case estimate when there is fairly continuous activity in the scene. In the best case, when the camera will only run to retrain the system, the system will consume approximately 4 watt, more than 99% saving in power consumption. Table 1 gives an estimate of the worse case power consumption when the camera was deployed in various test rooms at different time intervals.
6. Conclusion

This paper presents a vision based system for monitoring the activity level of an individual (elder) living alone. The paper presents a novel edge-base differencing algorithm capable of detecting a new or moving object. The system is also capable of training and retraining to build a map of the background using edges. A simple and easily implemented object location algorithm using block-base horizontal and vertical histograms has also been presented. The system is designed to be energy efficient so it can run longer if battery powered, unobtrusive using a fish-eye lens and require no intervention from the user during setup. The system as it is designed requires the camera to be mounted high up in the ceiling of a room. Also, to successfully locate the moving object, the assumption that only one object will be moving in the scene is made. Future work will involve developing the system to locate multiple subjects and locate track them in the scene. Work is ongoing to compare the proposed edge-based with background differencing algorithm in terms of speed and power consumption.

Acknowledgements

This work was supported by TSB under the TOTAL-CARE Project. Special thanks goes to the project team; Brian Fuchs, Janko Mrsic-Flogel, Davide Guidi, Vesso Novov and John Darlington of Imperial College, Nick Hunn of Wifore, Paul Nelson of Phrisk and Mike Beizsley of Tactical Systems Designers.

References