Infrared Physics & Technology 72 (2015) 127–134

Contents lists available at ScienceDirect

Infrared Physics & Technology

journal homepage: www.elsevier.com/locate/infrared

A robust system for counting people using an infrared sensor and a camera

Fatih Erden [a,*], Ali Ziya Alkar [b], Ahmet Enis Cetin [c]

[a] Department of Electrical and Electronics Engineering, Atılım University, Ankara 06836, Turkey
[b] Department of Electrical and Electronics Engineering, Hacettepe University, Ankara 06800, Turkey
[c] Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey

HIGHLIGHTS

- The multi-modal system consists of a PIR sensor and a regular camera.
- Entry/exit motions and ordinary body movements are distinguished by the PIR sensor.
- Motion types are classified by a Markovian decision algorithm in wavelet domain.
- The camera is turned off, unless the PIR sensor detects an entry/exit type motion.
- Accuracy of the camera-only system is improved and the processing cost is lowered.

ARTICLE INFO

Article history:
Received 20 June 2015
Available online 31 July 2015

Keywords:
Infrared sensors
Markov models
Multi-modal systems
People counting

ABSTRACT

In this paper, a multi-modal solution to the people counting problem in a given area is described. The multi-modal system consists of a differential pyro-electric infrared (PIR) sensor and a camera. Faces in the surveillance area are detected by the camera with the aim of counting people using cascaded AdaBoost classifiers. Due to the imprecise results produced by the camera-only system, an additional differential PIR sensor is integrated to the camera. Two types of human motion: (i) entry to and exit from the surveillance area and (ii) ordinary activities in that area are distinguished by the PIR sensor using a Markovian decision algorithm. The wavelet transform of the continuous-time real-valued signal received from the PIR sensor circuit is used for feature extraction from the sensor signal. Wavelet parameters are then fed to a set of Markov models representing the two motion classes. The affiliation of a test signal is decided as the class of the model yielding higher probability. People counting results produced by the camera are then corrected by utilizing the additional information obtained from the PIR sensor signal analysis. With the proof of concept built, it is shown that the multi-modal system can reduce false alarms of the camera-only system and determines the number of people watching a TV set in a more robust manner.

1. Introduction

Determining the number of people in a given area is a critical problem for many surveillance applications. Presence or absence of an unexpected number of people in an observed area may indicate an unusual situation [1]. A real-time and accurate estimation of people in a shop or a shopping mall can provide substantial information for managers. Control systems can manage power and energy consumption efficiently by correctly estimating the people count in buildings, e.g. they can adjust climate and lighting conditions according to the number of people present in the building [2]. The schedule of a public transportation system may be arranged according to the number of passengers waiting [3]. TV ratings are important information for the media industry. Conventional techniques [4–6] assume a fixed number of population (in the location where the measurement is taken) to find out how many people are watching a certain TV program. The information about which programs are being watched, as well as the tuning behaviors during programs and commercial breaks is delivered to the clients. However, audience measurement may be provided more accurately if the number of people sitting in front of the screen is exactly known. Hence, several works have addressed the problem of estimating the number of people in a definite area, such as [7–12].

* Corresponding author. Tel.: +90 312 290 1477; fax: +90 312 266 4126.
E-mail address: erden@ee.bilkent.edu.tr (F. Erden).

http://dx.doi.org/10.1016/j.infrared.2015.07.019
1350-4495/

In this paper, it is aimed to determine the number of people in front of a TV set using a pyro-electric infrared (PIR) sensor and a camera. PIR sensors are low-cost infrared sensors and provide both differential motion and infrared signature information about an observed area. This paper focuses on processing continuous-time PIR sensor signals to improve the counting results of the camera-only system.

Current PIR sensor based systems have many potential applications in automation of electrical appliances [13,14], flame detection systems [15], falling person detection [16], design and implementation of a home embedded surveillance systems [17], hand gesture recognition [18], battery-operated presence detection, etc. and are all based on the on/off decisions of the analog circuitry of the PIR sensors. There are a number of recent studies using both the analog circuitry of the PIR sensor and the continuous-time real-valued signals that the PIR sensor produces during a motion, but on different tasks other than counting people [15,19,20].

Yun and Lee [19] have recently developed a PIR sensor based system to detect the movement direction, speed and identity of a person. They collect the raw data coming from 3 modules, each of which consists of 4 PIR sensors, and form a reduced feature set, i.e. voltage peak value, time of the peak, and passage duration. Then they feed these features to a list of classifiers.

Wahl et al. [2] use a distributed PIR-based approach for estimating the people count in office environments. In this approach movement direction of a person passing through a gateway is aimed to be discriminated based on the timing of motion events reported by pairs of PIR sensors. The proposed work here differs from this study in the sense that it has a multi-modal structure and uses the continuous-time signal of the PIR sensor rather than the binary PIR sensors. In addition, the algorithm proposed in [2] estimates the number of people entering into an area but not the number of people present at any time in that area.

Dan et al. [8] present a people counting system using a video-plus-depth-camera mounted on the ceiling. This system is based on fusing the depth and vision data provided by a camera, rather than fusing different type of sensors.

Video processing based people counting methods can be categorized into two groups [21]: (i) detection-based and (ii) map-based methods. Detection-based methods use some form of segmentation and object detection to first detect people individually and then count them [9,22,23]. Map-based methods, instead, use the measurement of some feature to count people which does not require to detect each person in the scene separately [10,11]. Map-based methods are more suitable for precise measurement of people counting. Since the goal here is to count the number of people watching a TV set, a map-based method proposed by Viola and Jones [23] to detect human faces is used in this paper because it is computationally efficient enough to run in real-time. It also works well even in low-resolution video. Other video-based human detectors which may be more suitable for a given application can also be incorporated to the multi-modal system.

In this novel multi-modal system a differential PIR sensor is used in addition to a regular camera to overcome the problems faced by the camera-only system in counting people. Two types of human motion; (i) entry to and exit from the observed area and (ii) ordinary activities in the observed area are distinguished by the PIR sensor using a Markovian decision algorithm. It is not possible to differentiate these two motions using an ordinary PIR sensor providing only binary information. The wavelet transform of the continuous-time real-valued sensor signal received from the PIR sensor circuit is used for feature extraction. Wavelet parameters are then fed to a set of Markov models representing the two motion classes. The class affiliation of a test signal is determined according to the model yielding the highest probability. People counting results produced by the camera are corrected by the classification results of the PIR sensor signal analysis. It is experimentally shown that the multi-modal system can reduce false alarms and determine the number of people in a surveillance area more accurately. Since the camera is activated only when the analog decision circuitry of the PIR sensor detects an entry/exit type motion in the viewing range of the sensor, the resultant system is an energy efficient system. As far as is known, this is the first study on people counting based on the fusion of a PIR sensor and camera.

The organization of the paper is as follows. Operating principles of a differential PIR sensor and signal data acquisition are described in Section 2. The wavelet based sensor signal processing and the training of the Markov models representing the motion classes are presented in Section 3. The decision mechanisms of the PIR sensor and the multi-modal system are described in Section 4. Experimental results are presented in Section 5.

2. Infrared sensor and data acquisition

A differential PIR sensor basically measures the difference of infrared radiation density between the two pyro-electric elements inside. Fig. 1 shows the block diagram of a typical differential PIR sensor, (s1) and (s2) are the outputs of the pyro-electric elements and (g) is ground. Normal temperature alterations and changes caused by airflow are canceled by the two elements connected in parallel. If these elements are exposed to the same amount of infrared radiation, they cancel each other and the sensor produces a zero-output at (d). Thus the analog circuitry of the PIR sensor can reject false detections very effectively.

Commercially available PIR motion detector circuits produce binary outputs. However, it is possible to capture a continuous-time analog signal representing the amplitude of the voltage signal which is the transient behavior of the circuit. The corresponding circuit for capturing an analog output signal from the PIR sensor is shown in Fig. 2. The circuit consists of four operational amplifiers (op amps), U1A, U1B, U1C and U1D. U1A and U1B constitute a two stage amplifier circuit whereas U1C and U1D couple behaves as a comparator. The very-low amplitude raw output at the 2nd pin of the PIR sensor is amplified through the two stage amplifier circuit. The amplified signal at the output of U1B is fed into the comparator structure which outputs a binary signal, either 0 V or 5 V. Instead of using binary output in the original version of the PIR sensor read-out circuit, the analog output signal at the output of the 2nd op amp U1B is captured directly. The analog output signal is digitized using a microcontroller with a sampling rate of 100 Hz and transferred to a general-purpose computer for further processing. A typical sampled differential

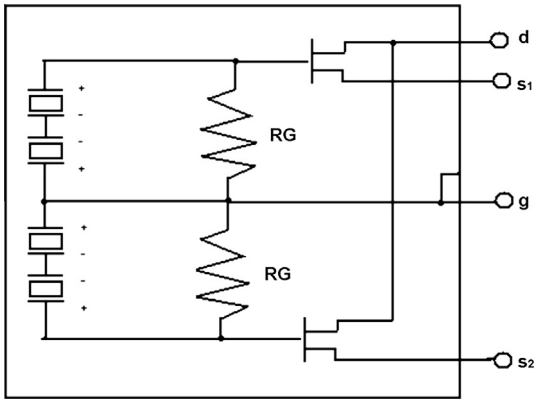

Fig. 1. Model of the inner structure of a differential PIR sensor.

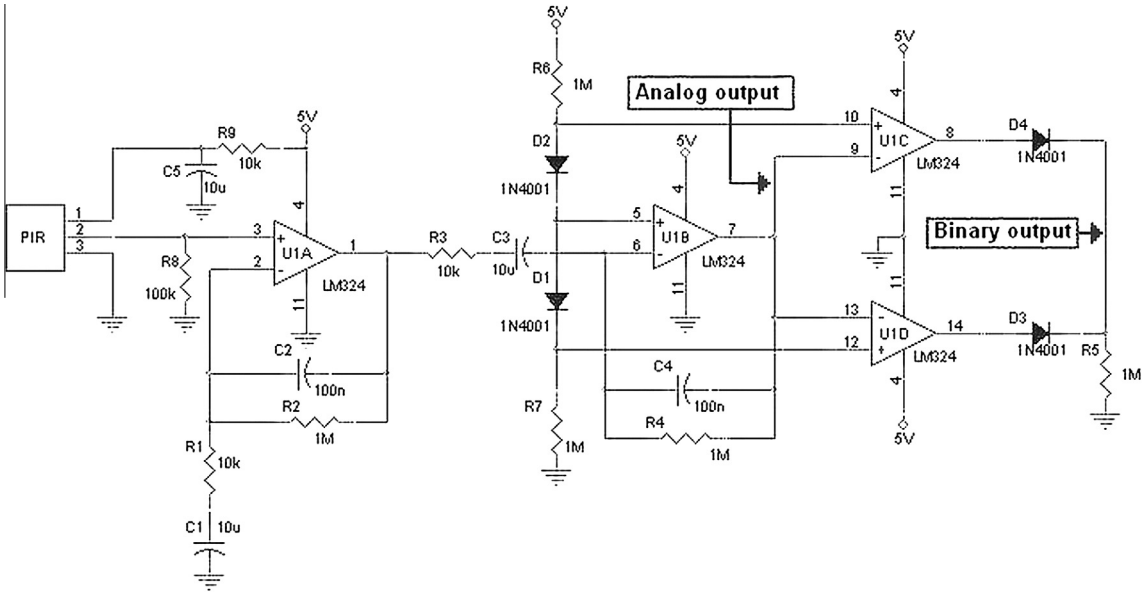

Fig. 2. The circuit diagram for capturing an analog output signal from a PIR sensor.

PIR sensor output signal for no activity case using 8 bit quantization is shown in Fig. 3.

3. Sensor signal processing and Markov models

Wavelet transform is used to extract features from the PIR sensor signal. Wavelet domain analysis provides robustness to variations in the sensor signal caused by temperature changes in the environment.

Let $x[n]$ be a sampled version of the signal received from the PIR sensor with a sampling frequency of 100 Hz. Wavelet coefficients $w[k]$ corresponding to [25 Hz, 50 Hz] frequency band information of $x[n]$ are obtained after a two-stage sub-band decomposition. In the decomposition process, the input signal is filtered with integer arithmetic filters corresponding to Lagrange wavelets followed by resolution halving. The transfer functions of the low-pass and the high-pass filters are given by:

$$H_l(z) = \frac{1}{2} + \frac{1}{4}(z^{-1} + z) \tag{1}$$

and

$$H_h(z) = \frac{1}{2} - \frac{1}{4}(z^{-1} + z), \tag{2}$$

respectively.

The wavelet transforms of the two sample signals of four seconds duration in the training set are shown in Fig. 4. Fig. 4(a) is for a person entering to the observed area and Fig. 4(b) is for simple hand/arm movements of a person in the observed area. The two wavelet signals both have peaks at around index 30. The wavelet signal obtained due to the entry motion of a person to the viewing range of the PIR sensor has a greater peak height at the time of the main motion compared to the arm movement and it also has follow-up oscillations. The amplitude of the peaks and the duration of the motions make the difference in Markov models representing the ordinary activities and entry/exit motions.

Two three-state Markov models are trained in the wavelet domain to represent the two types of motion: (i) entry to and exit from the surveillance area and (ii) ordinary activities such as hand, arm and leg motions in the surveillance area. First, states are defined. Let A and B be the training signal sequences formed by concatenating many sample signals in the "entry/exit motions" and "ordinary activities" classes, respectively. Each wavelet coefficient in A and B is mapped to a state by investigating the relation of the absolute value of the current wavelet coefficient, $|w[k]|$, to two non-negative thresholds, T_1 and T_2. The state of $w[k]$ is labeled as S_0, if $|w[k]|<T_1$. If $T_1<|w[k]|<T_2$, state S_1 and if $|w[k]|>T_2$, state S_2 is attained. The procedure to determine the thresholds will be introduced in the next subsection.

Next, the state sequences C_A and C_B are formed and the number of every possible transition in each state sequence is counted. Let a_{ij} and b_{ij} denote the number of transitions from state S_i to S_j in C_A and C_B. Since the peak height of a wavelet signal in the "entry/exit motions" class is greater than the one in "ordinary activities" class, it is expected that a_{22} will be greater than b_{22}. Moreover, more transitions between different states are supposed to occur in the entry/exit type signal because of the follow-up oscillations. The two three-state Markov models are shown in Fig. 5.

The training of the Markov models ends with the computation of the state transition probabilities for each class. If L_A and L_B are the lengths of C_A and C_B, then the state transition probabilities are computed as follows:

$$p_{a_{,b}}(i,j) = 1/L_{A,B}(a,b)_{ij}, \tag{3}$$

where $p_{a,b}(i,j)$ is the probability of a transition from state S_i to state S_j in $C_{A,B}$.

3.1. Threshold estimation

When there is no activity in the viewing range of the PIR sensor, the corresponding sensor output signal is a noise signal. In order to characterize the no activity case, the wavelet coefficients of the noise signal are mapped to the state S_0. In other words, T_1 is set to a value such that almost all absolute valued wavelet coefficients of the noise signal are below it. This value is chosen to be greater than $\mu + 2\sigma$ due to the well-known 68-95-99.7 rule, where μ is the mean and σ is the standard deviation of the training no activity signal in the wavelet domain, respectively.

In addition, the outputs of the training process $\bar{p}_a$and,$\bar{p}_b$ each of which is a function of T_1 and T_2, are supposed to reflect the

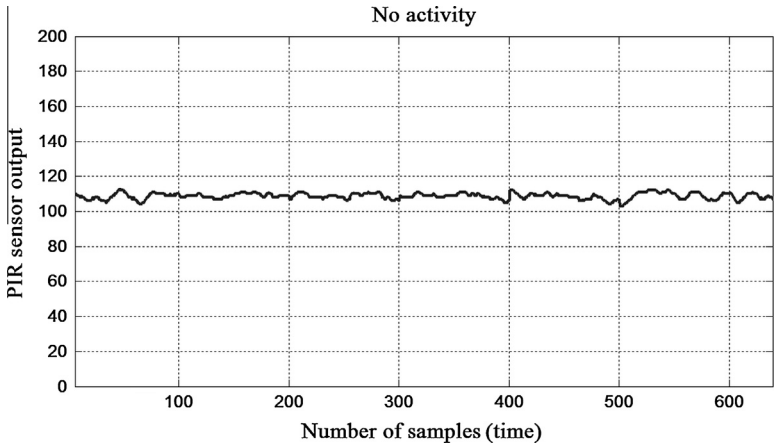

Fig. 3. A typical differential PIR sensor output signal when there is no activity within its viewing range. Sampling frequency is 100 Hz.

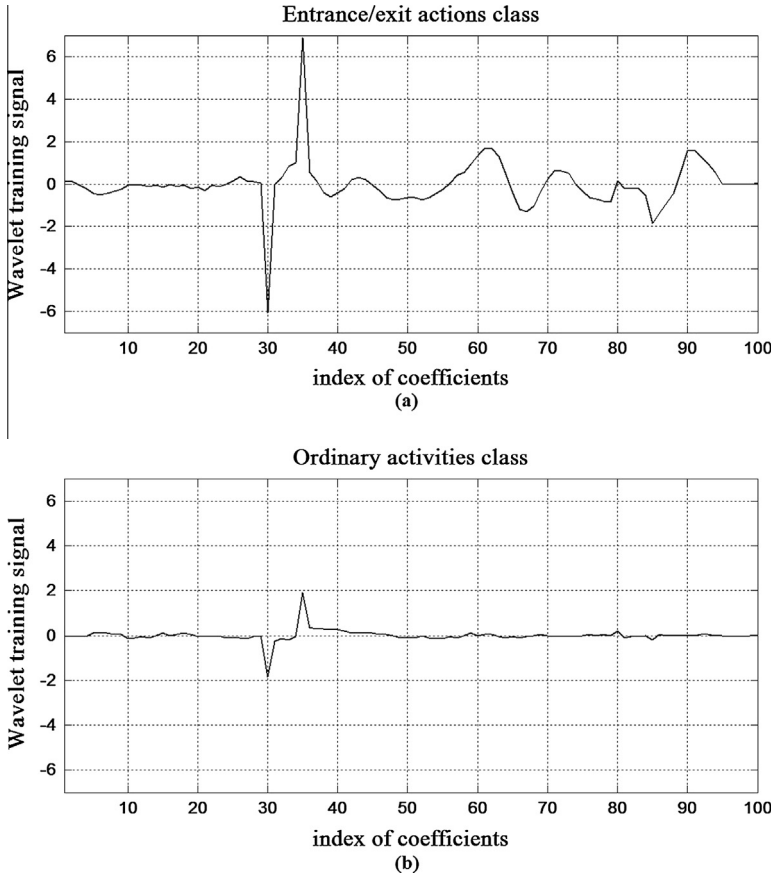

Fig. 4. Wavelet transformed PIR sensor training signal obtained due to (a) a person entering to the surveillance area and (b) the hand/arm movements of a person in the surveillance area.

distinction between the two classes. Thus, (T_1, T_2) is chosen such that they maximize the dissimilarity

$$D(T_1, T_2) = ||\bar{p}_a - \bar{p}_b||^2, \tag{4}$$

where $||\bar{x} - \bar{y}||$ is the L_2 distance between the points $\bar{x}$ and $\bar{y}$. A typical plot of the dissimilarity function in Eq. (4) is shown in Fig. 6. It is obvious from the figure that the dissimilarity function is non-differentiable and highly nonlinear. Therefore, it is maximized by using a genetic algorithm with the objective function $D(T_1, T_2)$.

4. Decision mechanism

The PIR sensor by itself cannot count the number of people in a surveillance area, but it can differentiate if the motion is an entry/exit type motion or just a hand/arm gesture. The class affiliation of a test signal is decided using a probabilistic approach. The test signal is first divided into windows of 300 samples covering a 3 s time interval and then wavelet transformation is carried out on each window. Since the resolution is halved in each stage of the wavelet decomposition tree, the resulting wavelet signal window is of length 75. Then the corresponding state sequence is

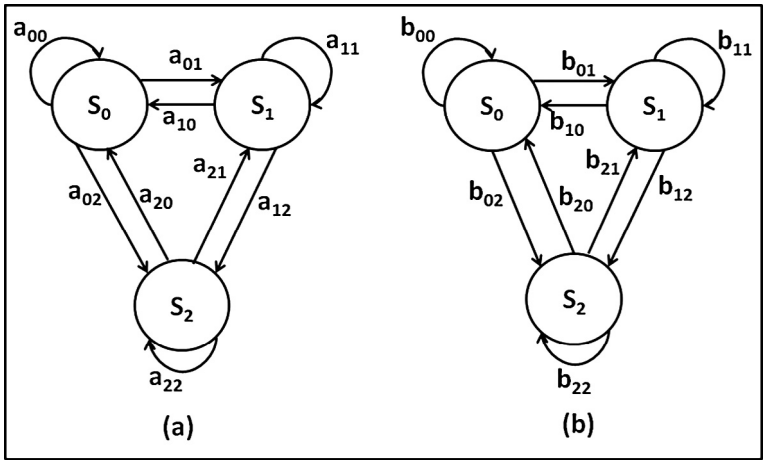

Fig. 5. The two three-state Markov models corresponding to the (a) "entry/exit motions" and (b) "ordinary activities" classes.

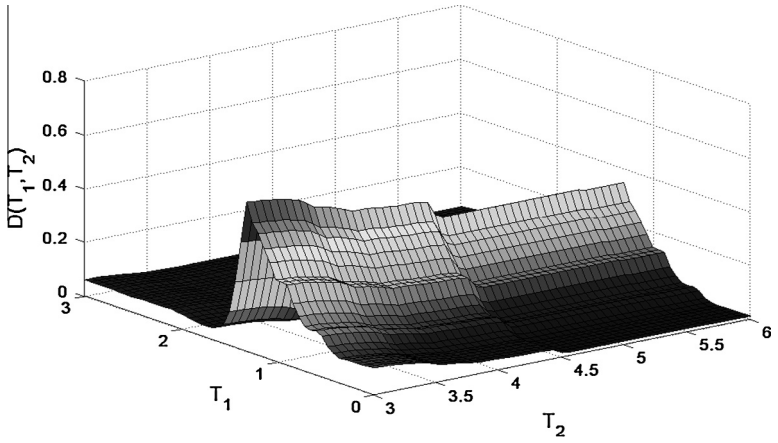

Fig. 6. A typical plot of the dissimilarity function $D(T_1,T_2)$.

generated. Let C be the state sequence of a test window. The probabilities of belonging to the "entry/exit motions" and "ordinary activities" classes for that window are calculated as follows:

$$P_{a,b}(C) = \prod_{i=0}^{L-1} p_{a,b}(C_i, C_{i+1}), \tag{5}$$

where L is the length of C and $p_{a,b}(C_i, C_{i+1})$ is the probability of a transition from the ith element to the $(i+1)$th element in C calculated in the training phase of each model.

If t_{ij} denotes the number of transitions from S_i to S_j in C, then Eq. (5) can be rearranged as follows:

$$P_{a,b}(C) = \prod_{i=0}^{2}\prod_{j=0}^{2} p_{a,b}(i,j)^{t_{ij}}. \tag{6}$$

The model yielding the higher probability for the current test signal window is reported as the class affiliation of that window. Since the class affiliation decision is based on the magnitude of the probabilities, taking the logarithm of both sides in Eq. (6) does not affect the result. This leads to a reduction in the computational cost of the decision mechanism, because multiplication is replaced by summation in the probability equations after taking the logarithm. The new probability equations become:

$$P'_{a,b}(C) = \sum_{i=0}^{2}\sum_{j=0}^{2} t_{ij}\log_{10}(p_{a,b}(i,j)). \tag{7}$$

In the classification process, just two models representing the "entry/exit motions" and "ordinary activities" classes are used. It is not necessary to form a model for the "no activity" case. The "no activity" case is easily detected when 90% or more of the elements of C are S_0.

Classification algorithm of a test signal window producing a state sequence C of length L can be summarized as in Algorithm I. In the next subsection video based face detection is described.

Algorithm 1. Markov models based classification algorithm.

```
if, test window ∈ "no activity" class
else
if, P'a(C) > P'b(C) test window ∈ "entry/exit motions" class
else, test window ∈ "ordinary activities" class
end
```

4.1. Video processing

Faces in front of a TV set are detected by the camera with the aim of counting people. As pointed out in Section 1, the method proposed by Viola and Jones [23] is used for this purpose because of its good performance in real-time. The errors of the camera-only system are then debugged by the PIR sensor signal analysis. Any other face detection algorithm satisfying the real-time constraints

may also be used to implement the proposed idea. But, the contribution generated by introducing the PIR sensor is independent of which vision-based method is used since all have to overcome similar challenges such as occlusion and lighting variances.

The method [23] for face detection is briefly reviewed here. It uses Haar-like features for feature extraction. A small number of these visual features are selected from a larger set by an AdaBoost based learning algorithm to create efficient classifiers. An image representation called "Integral Image" is used to scan the whole image and detect the presence of the features in sub-regions of that image very quickly. Each classifier looks for a set of features and if the result is positive, a rectangular region is selected for each potential face region. Then, largely overlapping rectangular regions are labeled and their intersections are reported as the locations of the faces in the image. The classifiers are cascaded according to their weights determined by the AdaBoost algorithm.

A camera-only system is not very reliable for counting people because lighting conditions, illumination and face-camera angle variations may give rise to false negative detections. This situation is illustrated in Fig. 7(a). Although there are two people in the surveillance area, only one is detected. Furthermore, there may be also specific problems due to the video analysis algorithm chosen. For example, rectangular regions turned by individual classifiers of the Viola–Jones face detector may largely overlap in a location other than the face region and this situation may lead to a false positive detection as shown in Fig. 7(b). In this case there is only one person in the surveillance area, but there are two detections.

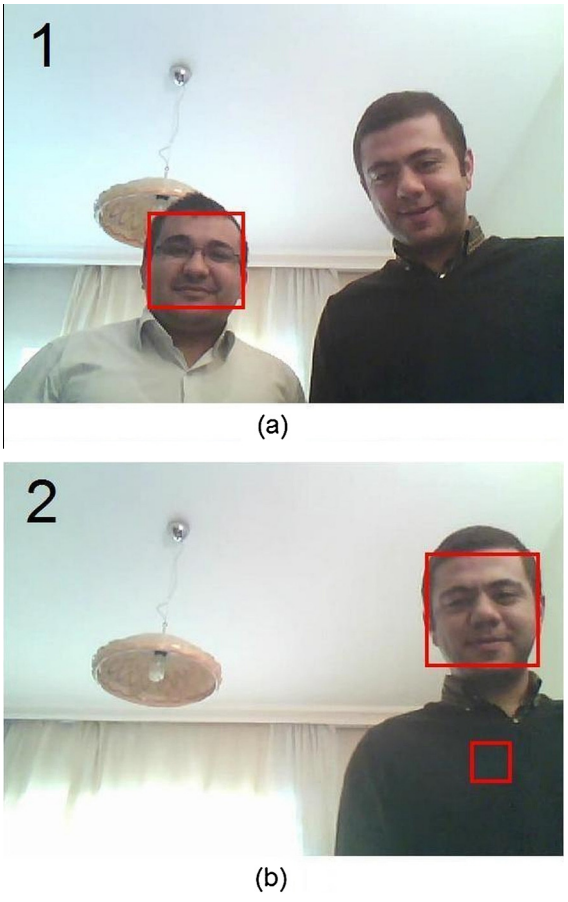

Fig. 7. An example of (a) a false negative, and (b) a false positive detection by the camera-only system using the Viola–Jones face detector [14]. Each rectangle indicates a separate detection.

To reduce the number of false detections produced the camera-only system, classification results of the PIR sensor signal analysis are used. The multi-modal system operates as follows. It starts counting faces in the surveillance area by using only the camera. If the result of counting remains fixed for a number of frames, it is assumed that the number of people present in that area is counted accurately and the camera is turned off. Then the PIR sensor is activated to analyze the motions in the surveillance area. As long as no activity or just an ordinary activity is detected, the camera stays in standby mode and does not count people again. The multi-modal system assumes that there is the same number of people in the surveillance area, which is the case indeed. Since the camera does not count people continuously unless there is an entry or exit motion, false negative and positive detections of the camera-only system are significantly reduced. In addition, the unnecessary image processing is avoided. Whenever the motion in the surveillance area is interpreted as an entry/exit type motion by the PIR sensor, the camera is activated again to count people and the same process is repeated.

5. Experiments and results

People counting experiments with the proposed multi-modal system are carried out in a 7 m × 7 m room. The PIR sensor and the camera are placed on top of a TV set. The distance between the door and the TV set is about 2.5 m. There are 2–6 people in the room at any time and they sit at a distance of 2–5 m to the TV set. The subjects present in the room continue with their ordinary activities such as hand/arm or head movements while watching the TV. Others are asked to enter to the room, have a seat immediately and watch the TV or leave the room randomly. 12 test video clips each of length 7 min on the average are recorded at 640 × 280 pixels and with a rate of 12 frames per second. Each video clip includes 24–28 entry/exit motions. The number of people in the room is counted by the camera-only and the multi-modal system.

Success rates of the camera-only system in counting people are presented in Table 1. A false positive detection indicates that there are less people, and a false negative detection indicates there are more people in the room than detected. The success rate is the ratio of the number of frames in which the number of people is estimated correctly to the number of total frames. The average success rate for the 12 test video clips using the camera-only system is 83.1%.

Performance of the camera-only system is improved by integrating a PIR sensor to the system. The PIR sensor signal is recorded in synchronization with the video in each test. During the training of the Markov models, 120 sample signals, each of which covers a

Table 1
People counting results of the camera-only system for 12 test video clips.

Test video	Number of frames	False positives	False negatives	Success rate (%)
#1	5040	74	756	83.5
#2	5053	36	1095	77.6
#3	5012	78	654	85.3
#4	5082	92	483	88.6
#5	5040	99	735	83.4
#6	5020	53	657	85.8
#7	5022	111	489	88.0
#8	5072	88	1003	78.4
#9	5064	44	939	80.5
#10	5110	67	774	83.5
#11	5089	40	1018	79.2
#12	5004	79	814	82.1
Avg	5050.6	71.7	777.2	83.1

Table 2
Results for the Markov models based classification of the entry/exit type motions using the PIR sensor.

Test sequence	Number of test motions	Detections	Success rate (%)
#1	27	27	100
#2	26	25	96.1
#3	25	25	100
#4	25	25	100
#5	27	27	100
#6	24	24	100
#7	26	25	96.1
#8	26	26	100
#9	27	27	100
#10	28	27	96.4
#11	25	25	100
#12	28	28	100
Avg	26.1	25.9	99.2

Table 3
People counting results of the multi-modal system for 12 test video clips.

Test video	Number of frames	False positives	False negatives	Success rate (%)
#1	5040	0	287	94.3
#2	5053	0	673	86.6
#3	5012	0	306	93.8
#4	5082	0	191	96.2
#5	5040	0	212	95.6
#6	5020	0	303	95.7
#7	5022	0	385	92.3
#8	5072	0	201	96.0
#9	5064	0	294	94.1
#10	5110	0	542	89.3
#11	5089	0	334	93.8
#12	5004	0	266	93.4
Avg	5050.6	0	332.8	93.4

3 s time interval, are recorded first for each class. The sample signals of the same class are then concatenated to estimate the parameters of each Markov model. The results for the Markov models based classification of the entry/exit type motions by using the PIR sensor are presented in Table 2. The test set consists of about 72 minutes-long records in total, including 314 entry and exit motions. Table 2 shows that the entry/exit motions can be distinguished from the ordinary activities with an overall success rate of 99.2% on the average. Only 3 of a total of 314 entry/exit motions are missed. Besides, during 72 minutes-long testing only 7 false alarms due to the unusual body movements (such as waving hands or arms) are produced. A false alarm does not lead to deterioration in the counting results, it just triggers the camera unnecessarily and causes power consumption.

People counting results of the multi-modal system for the same test set are presented in Table 3. In the multi-modal system, the camera does not count people unless the PIR sensor detects an entry motion. Thus, the multi-modal system does not produce any false positive detection. Similarly, since the camera stays idle unless an exit motion is detected, the false negatives, which are mainly caused by the changes in the face-camera angle, are significantly reduced. The multi-modal system achieves an average improvement of about 10% in comparison to the camera-only system. The improvements are lower in cases #2, #7 and #10; because an entry/exit motion is missed by the PIR sensor in these cases and consequently the number of people in the surveillance area is not updated by the camera. Nevertheless the overall performance of the multi-modal system is better.

Dan et al. [8] report a 98% accuracy, which is better than it is reported in this paper, for people counting by using both the depth and vision data of a 3D camera. It is obvious that it is possible to achieve higher success rates using different vision-based methods. But this paper aims to show that the accuracy of a camera-only system for people counting can be increased by adding a PIR sensor to the system. The validity of the proposed idea is independent of which vision-based method is being employed, because all of them suffer from similar problems such as occlusion, and illumination.

A test setup with a camera and a PIR sensor is used to estimate the computational gain by using the multi modal approach presented in this paper with respect to a camera only approach. Following a detection of an entry/exit type motion by the PIR sensor, it takes 5 seconds on the average to satisfy the condition to ensure that the camera detects the number of people correctly. By considering the 12 test sequences which include 314 entry/exit type motions, this duration approximately corresponds to 26 minutes in total. This means that the camera is turned off for 46 minutes in the 72 minutes-long testing. On the other hand, the PIR sensor is on for this period. But the cost of processing the 1-D PIR sensor signals is much lower than that of the images captured by the camera. If a camera was used by itself, the camera would be on for the entire test duration. As a result, the multi-modal system is more efficient than the camera-only system in terms of computation and power consumption.

6. Conclusion

A novel multi-modal system consisting of a low-cost PIR sensor and a regular camera to count people in a given area is successfully demonstrated. As far as is known, this is the first study on people counting based on the fusion of PIR sensors and cameras. It is shown that the entry/exit type motions can be discriminated from the ordinary body motions of a person by processing the continuous-time real-valued signals of a PIR sensor using a Markovian decision algorithm. The camera of the multi-modal system does not count people unless the PIR sensor detects an entry/exit type motion and it is assumed that the number of people in the surveillance area remains the same. Thus, the multi-modal system estimates the number of people in a more robust manner than the camera-only system. In addition, since the camera is triggered by the PIR sensor, the resulting multi-modal system consumes less power than a camera-only system.

Conflict of interest

There is no conflict of interest.

References

[1] M.J.V. Leach, E.P. Sparks, N.M. Robertson, Contextual anomaly detection in crowded surveillance scenes, Pattern Recognit. Lett. 44 (2014) 71–79.
[2] F. Wahl, M. Milenkovic, O. Amft, A distributed PIR-based approach for estimating people count in office environments, in: Proc. IEEE 15th International Conference on Computational Science and Engineering, Paphos, Cyprus, 2012, pp. 640–647.
[3] D. Conte, P. Foggia, G. Percannella, F. Tufano, M. Vento, Counting moving people in videos by salient points detection, in: Proc. 20th International Conference on Pattern Recognition, Istanbul, Turkey, 2010, pp. 1743–1746.
[4] W.L. Thomas, L. Daozheng, Audience measurement system utilizing ancillay codes and passive signatures, US Patent 5,481,291, 1996.
[5] H.B. Wheeler, L. Daozheng, Source detection apparatus and method for audience measurement, US Patent 6,675,383, 2004.
[6] E. W. Aust, L. Daozheng, Audience measurement system incorporating a mobile handset, U. S. Patent 6,467,089, 2002.
[7] S.W. Kim, J.Y. Jung, S.J. Lee, A.W. Morales, S.J. Ko, Sensor fusion-based people counting system using the active appearance models, in: Proc. IEEE International Conference on Consumer Electronics, Las Vegas, NV, 2013, pp. 65–66.
[8] B.K. Dan, Y.S. Kim, Suryanto, J.Y. Jung, S.J. Ko, Robust people counting system based on sensor fusion, IEEE Trans. Consum. Electron. 58 (3) (2012) 1013–1021.

[9] T. Zhao, R. Nevatia, B. Wu, Segmentation and tracking of multiple humans in crowded environments, IEEE Trans. Pattern Anal. Mach. Intell. 30 (7) (2008) 1198–1211.
[10] H. Celik, A. Hanjalić, E.A. Hendriks, Towards a robust solution to people counting, in: Proc. IEEE International Conference on Image Processing, Atlanta, GA, 2006, pp. 2401–2404.
[11] V. Rabaud, S. Belongie, Counting crowded moving objects, in: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, New York, NY, 2006, pp. 705–711.
[12] P. Kilambi, E. Ribnick, A.J. Joshi, O. Masoud, N. Papanikolopoulos, Estimating pedestrian counts in groups, Comput. Vis. Image Underst. 110 (1) (2008) 43–59.
[13] C.H. Tsai, Y.W. Bai, C.A. Chu, C.Y. Chung, M.B. Lin, PIR-sensor-based lighting device with ultra-low standby power consumption, IEEE Trans. Consum. Electron. 57 (3) (2011) 1157–1164.
[14] S. Lee, K.N. Ha, K.C. Lee, A pyroelectric infrared sensor-based indoor location-aware system for the smart home, IEEE Trans. Consum. Electron. 52 (4) (2006) 1311–1317.
[15] F. Erden, B.U. Toreyin, E.B. Soyer, I. Inac, O. Gunay, K. Kose, A.E. Cetin, Wavelet based flickering flame detector using differential PIR sensors, Fire Saf. J. 53 (2012) 13–18.
[16] B.U. Toreyin, E.B. Soyer, I. Onaran, A.E. Cetin, Falling person detection using multisensor signal processing, EURASIP J. Adv. Signal Process. 2008 (2008) 1–7.
[17] Y.W. Bai, L.S. Shen, Z.H. Li, Design and implementation of an embedded home surveillance system by use of multiple ultrasonic sensors, IEEE Trans. Consum. Electron. 56 (1) (2010) 119–124.
[18] P. Wojtczuk, A. Armitage, T.D. Binnie, T. Chamberlain, Recognition of simple gestures using a PIR sensor array, Sens. Transducers J. 14 (1) (2012) 83–94.
[19] J. Yun, S.-S. Lee, Human movement detection and identification using pyroelectric infrared sensors, Sensors 14 (5) (2014) 8057–8081.
[20] B. Yang, J. Luo, Q. Liu, A novel low-cost and small-size human tracking system with pyroelectric infrared sensor mesh network, Infrared Phys. Technol. 63 (2014) 147–156.
[21] Y. Hou, G. Pang, People counting and human detection in a challenging situation, IEEE Trans. Syst. Man Cybern. A Syst. Humans 41 (1) (2011) 24–33.
[22] J. Rittscher, P. H. Tu, N. Krahnstoever, Simultaneous estimation of segmentation and shape, in: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, San Diego, CA, 2005, pp. 486–493.
[23] P. Viola, M.J. Jones, Robust real-time face detection, Int. J. Comput. Vis. 57 (2) (2004) 137–154.