People Counting System by Using Kinect Sensor

Aylin Coşkun[1], Anil Kara[1], Mustafa Parlaktuna[2], Metin Ozkan[1],
[1]Department of Computer Engineering
[2]Department of Mechatronics
Eskisehir Osmangazi University, Eskisehir, Turkey
{aylincsknn, anilkraa, mparlaktuna}@gmail.com, meozkan@ogu.edu.tr,

Osman Parlaktuna
Department of Electrical-Electronics Engineering
Eskisehir Osmangazi University
Eskisehir, Turkey
oparlak@ogu.edu.tr

***Abstract*—This paper presents a people counting system as an application of a people counting method. The method uses depth and vision data captured by a low-cost Kinect sensor. The sensor is placed perpendicular to the ceiling. The perception of the people is achieved by using water filling method. A new approach called as people tracking increases the performance of the people counting system. Some real applications are given to show the efficiency of the proposed approach.**

***Keywords*—people counting; people tracking; depth camera**

I. INTRODUCTION

People counting system is required to ensure security and to obtain statistical information. The system determines the in/out people number in a specific counting area. The service providers may monitor passengers in a transportation system or consumers in shopping malls, museums, hospitals, and so on [1-2].

There are various sensors used for the people counting systems. Vision sensor, infrared sensor, ultrasonic sensor, and depth camera are some examples to these sensors. Some approaches use the camera installed at various locations with diverse camera views. A camera mounted with inclined view has advantages. The camera may be used for both surveillance and people counting purposes. However, this approach needs to solve occlusion problem [1-3]. To solve this problem, some approaches prefer to mount the camera on the ceiling downward as seen in Fig. 1 [4-5].

In the literature, there are various techniques to construct fast and reliable people detectors for surveillance applications. In [6], a wide-angle camera is used and an algorithm is developed based on motion and size criteria. In this algorithm, they detect the people from the motion histogram. In [7], a new model-based approach is used for tracking people. But, they do not consider counting people. a system is proposed to analyze the captured images and estimate imported feature [8]. In [9], template matching method is used with the images captured from a stereo camera.

In the proposed system, the vertical depth information generated by low-cost Kinect sensor is used. The heads of the people are determined by using water filling algorithm [4] after performing image enhancement. To capture depth information from the Kinect sensor, the Robot Operating System (ROS) is used. The image processing is performed by using open source library OpenCV. The proposed approach has two contributions for people counting. The people are tracked while they are in the region of view angle of the camera. When a person enters into the field of vision, an ID is assigned to him. Until leaving the field of vision, the person is counted just once. The other contribution is that a person may be identified when he/she gets lost suddenly and appears shortly after. For example, the head of a short person may not be appeared if he/she walks close to long people for a short time while passing through the field of view of the camera. The proposed system is presented by real experiments.

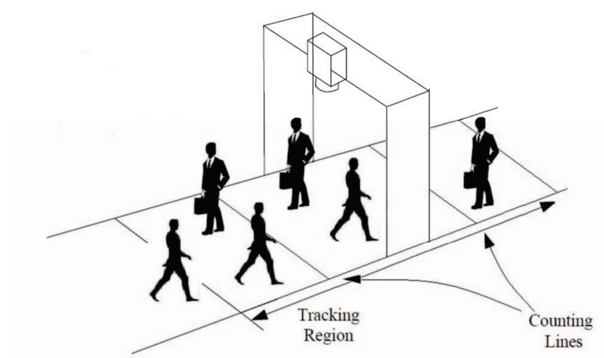


Fig. 1. People counting system.

The paper is organized as follows. In Chapter 2, the proposed system is presented in detail. Chapter 3 consists of the presentation of the experiment. The conclusion is given in Chapter 4.

II. PROPOSED METHOD

The people counting system is set up as seen in Fig. 1. A sensor is placed on the ceiling and pointing to floor. The people are tracked in the region of vision. However, a person is counted if he passes the counting lines. As a

978-1-4673-7751-5/15/$31.00 ©2015 IEEE

sensor, Kinect which provides depth data is used. Kinect sensor has monochrome depth sensing video stream which is in VGA resolution (640 × 480 pixels) with 11-bit depth, and provides 2,048 levels of sensitivity.

The flow chart of the people counting system is given in Fig. 2. First, the depth data is cleaned from noise, then background subtraction method is applied to eliminate irrelevant data. After that, local maximum points are found by using water filling method and filtering is done on the result of this image. Finally detection, tracking and counting processing are applied on filtered image, respectively.

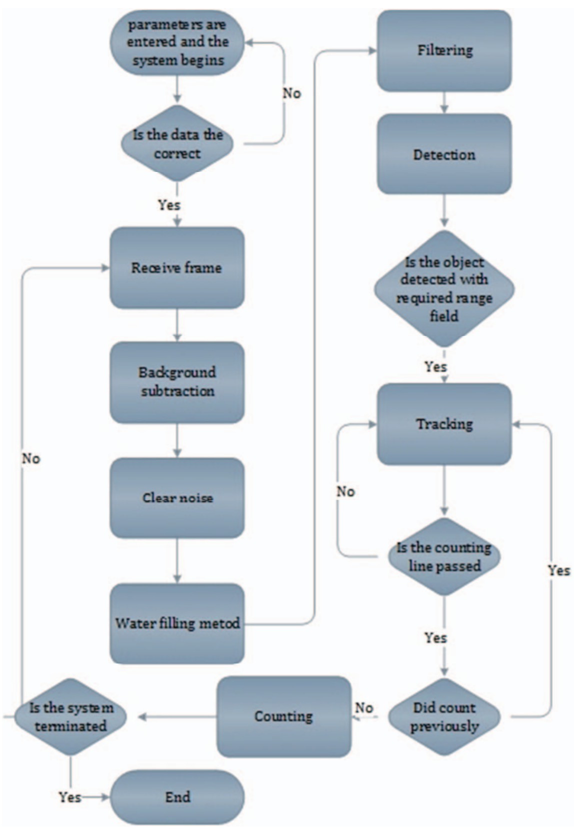


Fig. 2. Flow chart of people counting system

A. *Capturing and Refinement of Depth Data*

Basically, the system captures the depth data from Kinect sensor. Fig. 3(a) shows a depth image which consists of two people. The image is taken from the sensor by aid of ROS (Robot Operating System) which makes capturing image easy. Environmental conditions like bright ground or objects, corners, etc. cause noise in the depth data. To eliminate noise, some filters, such as median filter, morphological operations, etc., are used. For the filtering of the images, the OpenCV (Open Computer Vision Library) is used. The filtered image is given in Fig. 3(b).

B. *Background Subtraction*

The background subtraction process facilitates the people detection and increases the system performance. After the process, the image just handles the regions of interest which contains most probably the head and shoulders of people. However, gray level of the the image must be protected while applying the background subtraction. For this purpose, the image $A(i,j)$ obtained by filtering with a median filter is converted to a binary image $B(i,j)$ using (1).

$$\mathrm{B(i,j)} = \begin{cases} 1, \mathrm{minthresh} < \mathrm{A(i,j)} < \mathrm{maxthresh} \\ 0, \mathrm{otherwise} \end{cases} \quad (1)$$

Then, using (2) the background is removed and gray level is preserved.

$$\mathrm{C(i,j)} = (\mathrm{AoB})_{\mathrm{ij}} = \mathrm{A_{ij} B_{ij}} \quad (2)$$

where $C(i,j)$ is the background subtracted image. In Fig. 3(c), the background subtracted image is given.

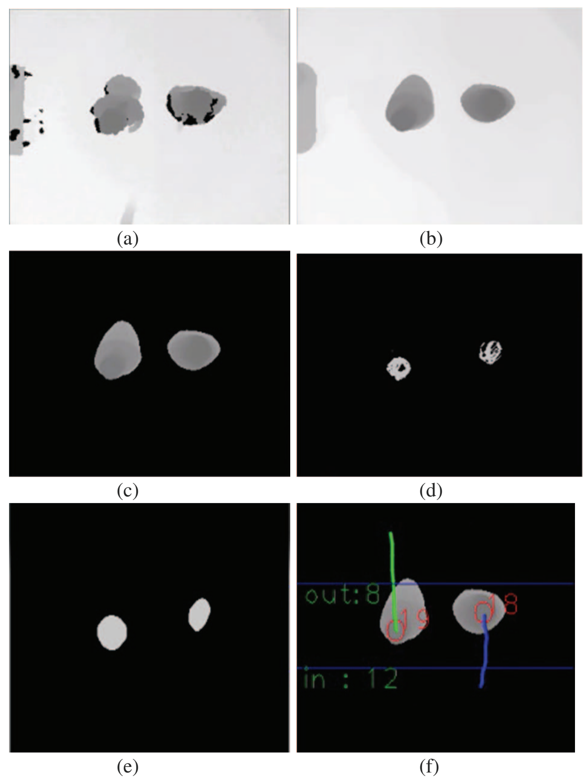
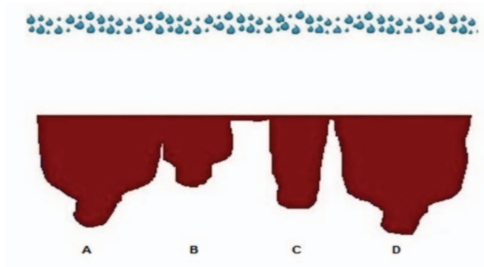


Fig. 3. Image Processing: (a) Original depth data, (b) Cleaned depth image, (c) Background subtraction, (d) Implementation of water filling method, (e) Filtered image, (f) Counting process

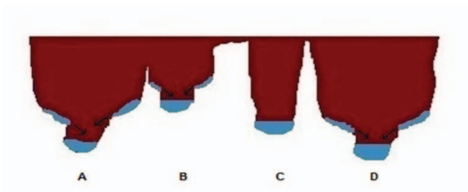
C. *Determination of the Local Maximum Points*

In the system, the important point is finding the local maximum points correctly. For this purpose, an efficient algorithm called the water-filling algorithm [4] is used. This method is motivated by the water filling process, that the water moves away from the heave and out to the nearby hollow under the force of gravity until the gravitational potential energy can't be reduced any more. This method simulates the rain by generating the raindrop according to a uniform distribution. Once a raindrop arrives, its landing spot is compared with its neighborhood to find the descent direction until it can't descend any more, then the number of raindrops at the balance spot increases. Since most of the raindrops at nearby landing spot tend to flow to the same hollow, a fast algorithm that speeds up the process is proposed. In this way, local minimum points in the negative image and local maximum points in the original image are found. The pseudo-codes of the algorithm are given in [4]. Fig. 4 shows the illustration of the water filling algorithm. A, B, D corresponding to three people respectively and C is a noise region. Region B has smaller scale compared with A and D, so one can take it as a child. C has no layers since it is considered as noise. After the water filling process, the local maximum points of A, B and D is found but the local maximum points of C is not found..

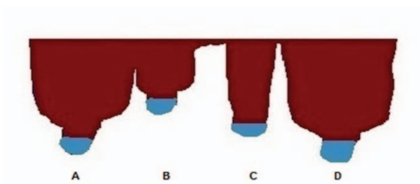
In Fig. 3(d), the image which is obtained after the implementation of the water-filling algorithm is given.



(a)



(b)



(c)

(d)

Fig. 4. Illustration of Water Filling method: (a) Original depth image, (b) duration of raindrops, (c) flowing raindrops from the layers, (d) result of water filling

D. Filtering

Some morphological operations are applied by using erosion and dilation to the image which consist of the local maximum points for taking the form of typical head, so errors are reduced as seen in Fig. 3(e).

E. Detection and Tracking

In this stage, the people are detected by applying threshold to the image obtained with water-filling algorithm. Then, an ID is assigned to each person in the image.

The important part of the system is tracking process. For tracking, a combinatorial optimization algorithm called as Hungarian algorithm [10] and Kalman filter [11] are used. The Hungarian algorithm solves the assignment problem in polynomial time. The Kalman filter uses a series of measurements over time, in the presence of noise (random variations) and other inaccuracies and produces estimates of unknown variables that tend to be more precise than those based on a single measurement alone. The head tracking is handled by sequential frames. The ID assignments to heads are performed by using the Hungarian algorithm. On the other hand, the Kalman filter determines the head positions having the same id in each frame.

F. Counting Process

Bidirectional counting system serves in a specific gateway area. Therefore, counting lines are placed perpendicular to the gateway direction to realize the entry and exit procedures. Head motions are tracked and counting was performed according to the transition of the lines. The counting process is performed when the person passes from input or output lines. Counted head is not counted again if it passes the counting line without leaving the screen. The image including counting line, head IDs, and tracked path is given in Fig. 3(f). If a head gets lost for a short time while it is in the region of vision, the system continues to track it with the same ID whenever it reappears on the region of vision.

G. Software Design of the System

The software for the system was designed by using UML class diagram (Fig. 5) and was implemented using C ++ programming language

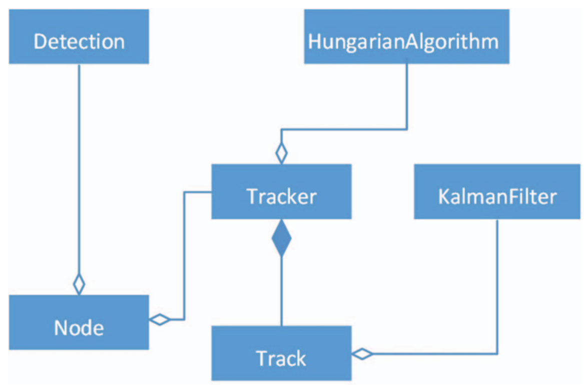


Fig. 5. UML class diagram for the People counting system.

In this diagram, there are six classes: Node, Track, Tracker, KalmanFilter, HungarianAlgorithm and Detection.

- *Node:* Basic class. By calling detection and tracker classes, the system cycle is executed in this class.
- *Track:* A point is detected in a frame.
- *Tracker:* Point sequence is detected in a frame.
- *KalmanFilter:* Estimates the coordinates of the next point sequence.
- *HungarianAlgorithm:* Mapping of the points is done by possible cost calculations between sequential frames.
- *Detection:* detects the object having the specified area.

III. APPLICATIONS

Some real applications are performed to show the performance of the proposed system in the building of the Electrical and Computer Engineering Laboratory at the Eskisehir Osmangazi University. For the applications, some possible conditions are considered and results are commented.

A. *Normal Passing*

The sequential frames and system output are given in Fig. 6. In this application, there is no crowd in the region of vision. People do not exhibit behaviors such as leaning, opening arms, etc. People walk at a normal tempo.

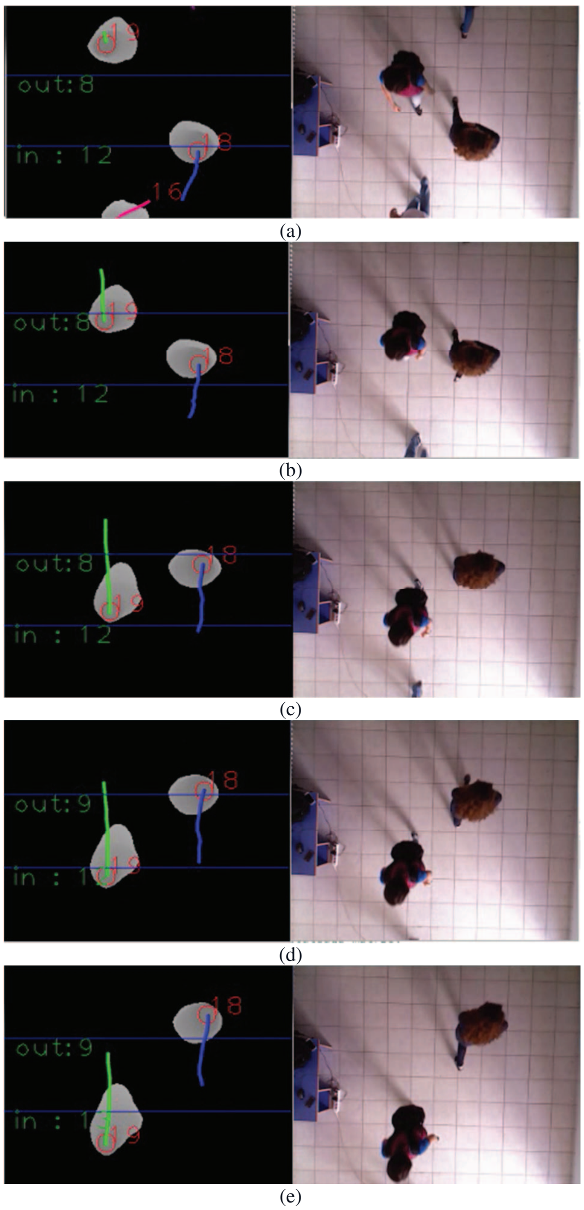


Fig. 6. Normal passing of two people

As seen in Fig. 6, there are two people passing in the opposite directions, and the system successfully tracks and counts them.

B. *Passing with Arm and Hand Movements*

The sequential frames and system output are given in Fig. 7 and 8. A person is passing by opening his arms to the sides (Fig. 7). In Fig 8, person is placing his hands to the top of his head while passing. Thus, the frames are disturbed by the behavior of the person. Person walks at a normal tempo.

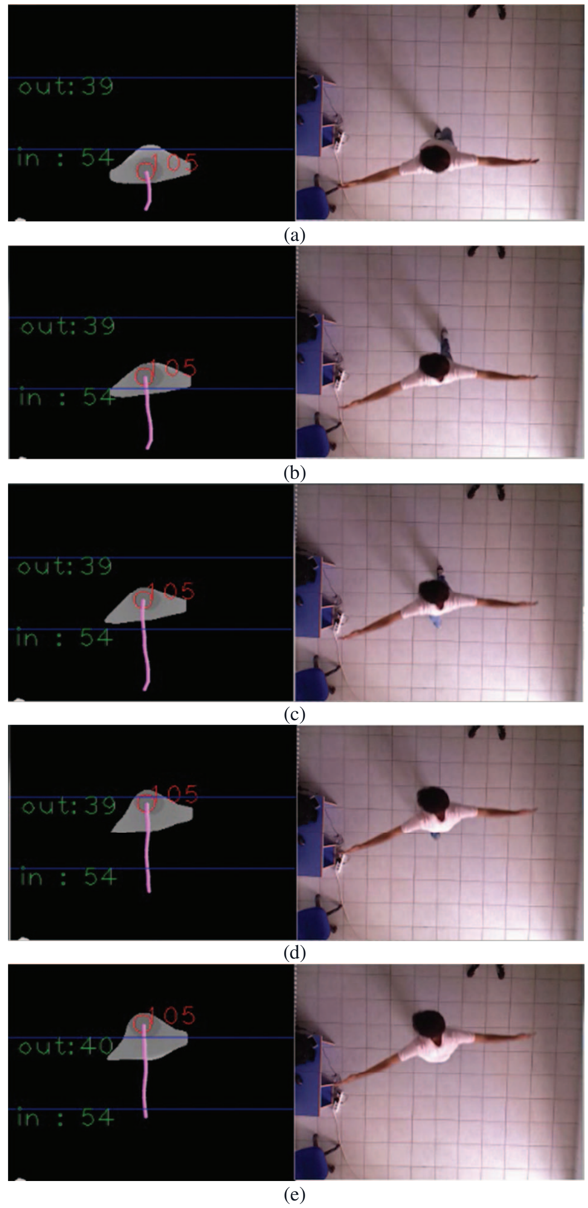
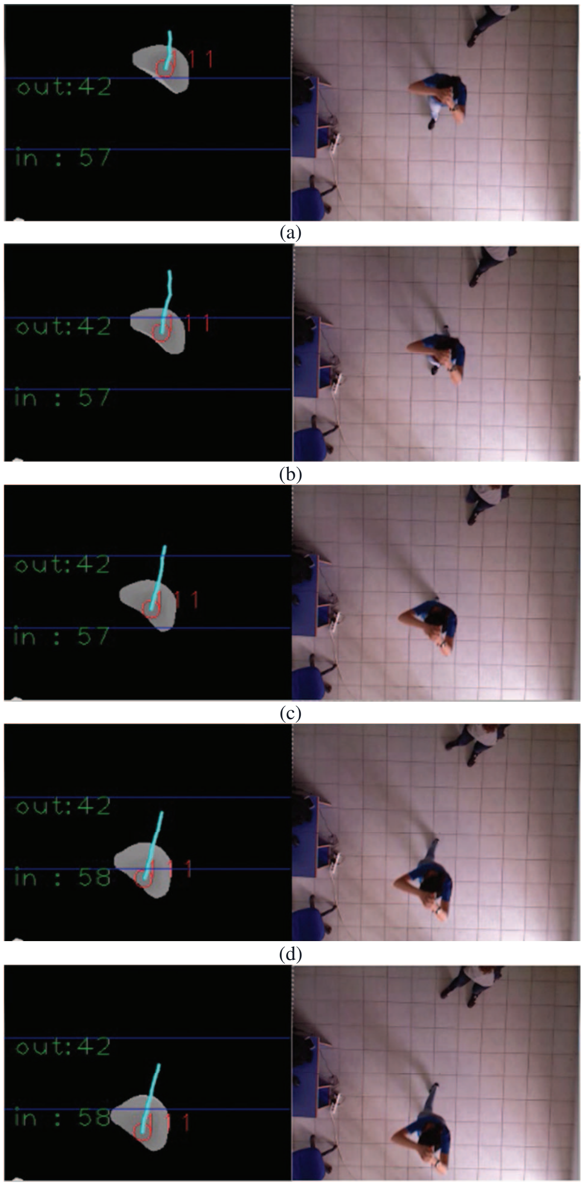


Fig. 7. Passing while opening arms to the sides

As seen in Fig. 7, the person is tracked and counted successfully. Opening arms to the sides has no effect on the system performance.

In Fig. 8, the person is placing his hands to the top of his head. This behavior is guessed to have negative effect for detection of the head since the algorithm considers heads for detection of a person. However, the test results show that there is no effect on the detection process.



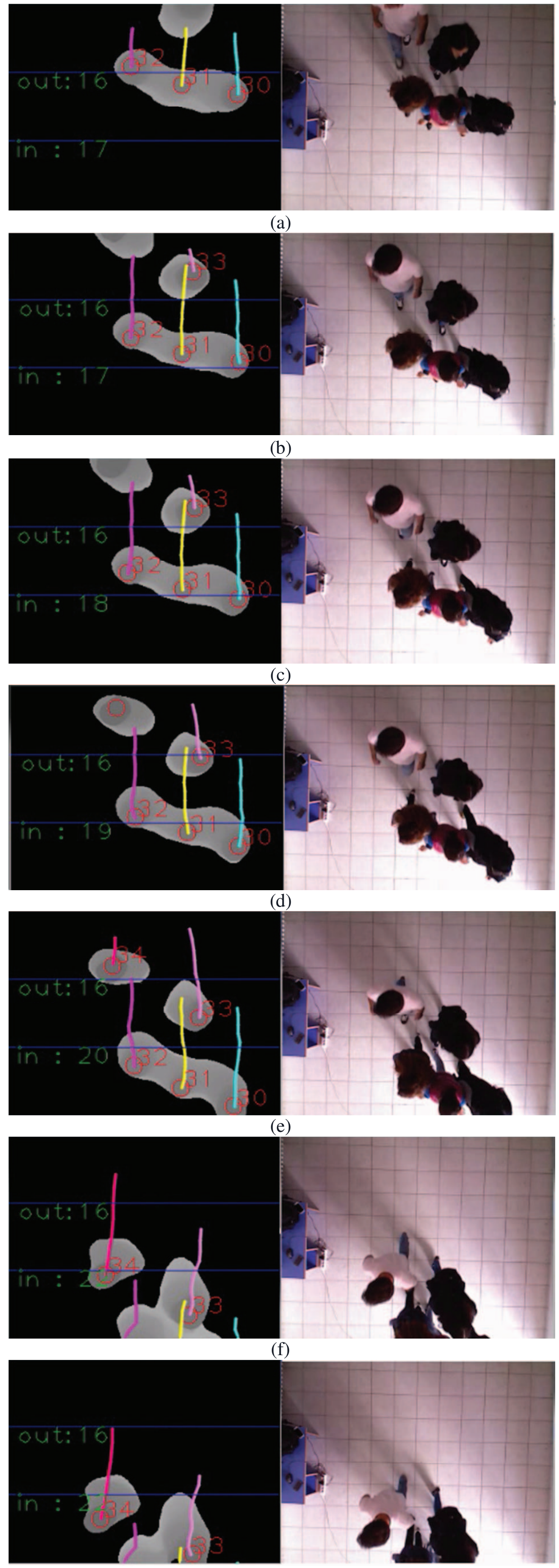
(a)
(b)
(c)
(d)
(e)

Fig. 8. Passing with hands on the head.

C. *Passing with the Crowd and Close Contact*

The sequential frames and system output are given in Fig. 9. There are five people passing in the region of vision. The people are very close to each other. The people are walking with close contact at a normal tempo. The heights of the people differ from each other.

As seen in Fig. 9, a unique ID is assigned to each person in the frame and each one is tracked and counted successfully. The crowd and close contact in the region of vision causes generally difficulties for counting and tracking. However, the proposed system is not affected from this situation.



(a)
(b)
(c)
(d)
(e)
(f)
(g)

Fig. 9. Passing with the crowd and close contact

D. Quick Passing with the Crowd and Close Contact

The sequential frames and system output are given in Fig. 10. In this application, we measure the performance of the system for fast passing with an example where five people pass in the region of vision by running. The people are very close to each other.

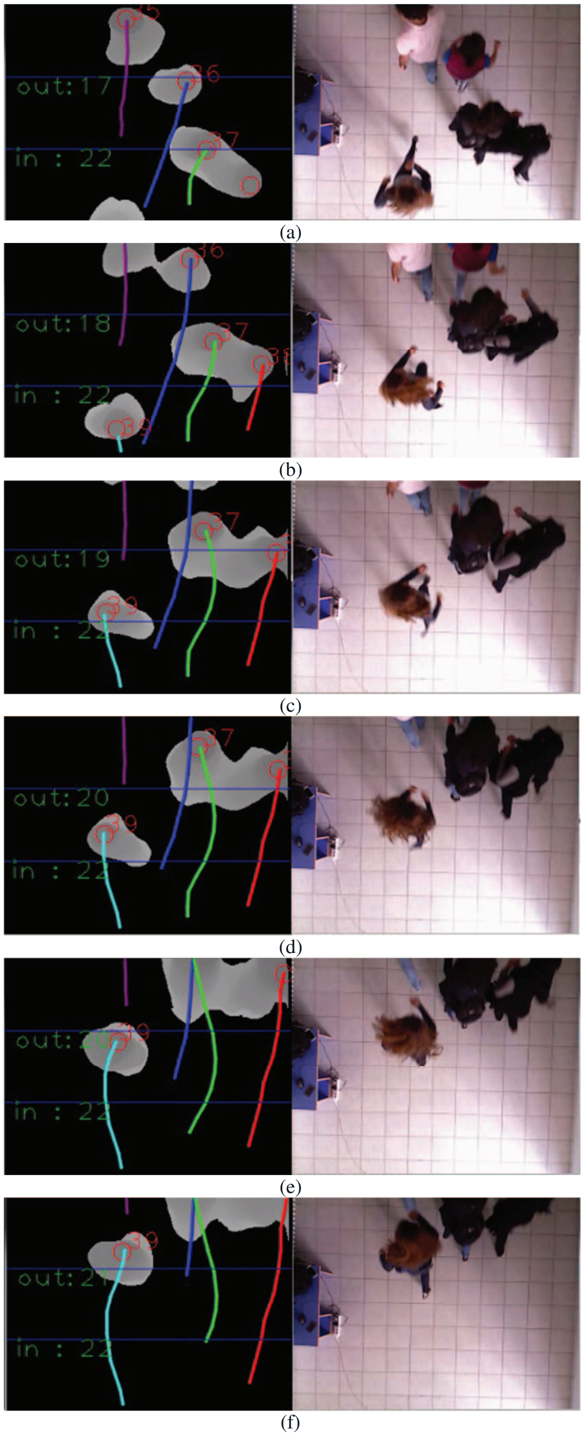


Fig. 10. Quick passing with the crowd and close contact

As seen in Fig. 10, five people are passing from the region of vision by running. The system successfully tracks and counts each person. Thus, fast or slow passing with/without the crowd does not affect the system performance despite the presence of the calculation loads of Kalman Filter, water-filling, and Hungarian Algorithms.

E. Elimination of Chattering while Passing Counting Line

The sequential frames and system output are given in Fig. 11. In this application, we show that the false counts caused by passing the counting line more than once without getting out of the region of vision are eliminated. There are two people talking each other while standing on the region of vision. One of the people moves from one side to other on the counting line more than once.

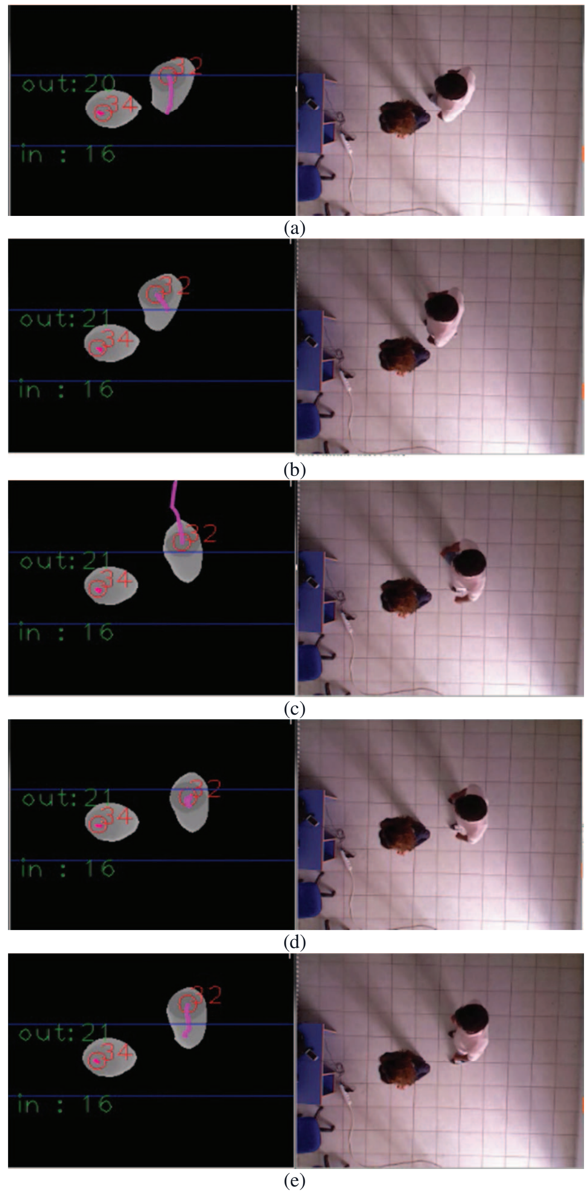


Fig. 11. Elimination of chattering while passing counting line

As seen in Fig. 11, the white dressed person is swinging in and out of the region. Therefore, his head passes from one side to other side of the out counting line more than once. However, the system counts this person just one time. If a person in the region of vision passes one of the counting line more than once without passing the other counting line, he is counted just one time. The same solution is also available for the in counting line.

IV. Conclusion

In this study, we have presented a method for people counting. Although there are similar studies in the literature, we proposed some minor enhancements as a contribution, such as elimination of chattering, tracking the person lost in the image for a short time. Also, we have inspected the counting problem by giving various real applications from different perspectives.

The real applications have approved the good performance of the proposed system. Nowadays, the authors are working on obtaining a compact product including the sensor, single board computer, and modular embedded software with the capability of remote access from Wi-Fi.

References

[1] K.-Y. Yam, W.-C. Siu, N.-F. Law, and C.-K Chan, “Effective bidirectional people flow counting for real time surveillance system,” 2011 IEEE Int. Conf. on Consumer Electron. (ICCE), pp.863-864, Jan. 2011.

[2] A. Mecoci, F. Bartolini, and V. Cappellini, “Image sequence analysis for counting in real time people getting in and out of a bus,” Signal Processing, vol. 35, no. 2, pp. 105-116, Jan. 1994.

[3] X. Liu, P. H. Tu, J. Rittscher, A. Perera, and N. Krahnstoever, “Detecting and counting people in surveillance applications,” IEEE Conf. on Advanced Video and Signal Based Surveillance(AVSS) 2005, pp. 306- 311, Sept. 2005.

[4] X. Zhang, J. Yan, S. Feng, Z. Lei, D. Yi, and S.Z. Li, “Water Filling: Unsupervised People Counting via Vertical Kinect Sensor”, 2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance, pp.2015-220, 2012.

[5] B. Dan, Y. Kim, S., J. Jung, and S. Ko, “Robust People Counting System Based on Sensor Fusion”, IEEE Transactions on Consumer Electronics, Vol. 58 (3), pp.1013-1021, 2012.

[6] T. Teixeira and A. Savvides, “Lightweight people counting and localizing for easily deployable indoors WSNs,” IEEE Journal of Selected Topics in Signal Processing, vol.2, no.4, pp.493-502, Aug. 2008.

[7] J. Segen and S. Pingali, “A camera-based system for tracking people in real time,” in IEEE Proc. Of Int. Conf. Pattern Recognition, vol. 3, pp. 63-67, 1996.

[8] J.-W. Kim, K.-S. Choi, B.-D. Choi, and S.-J. Ko, “Real-time system for counting the number of passing people using a single camera,” Lecture Notes in Computer Science, vol. 2781, pp. 466-473, Sep. 2003.

[9] K. Terada, D. Yoshida, S. Oe, and J. Yamaguchi, “A method of counting the passing people by using stereo images,” in Proc. IEEE Int. Conf. Image Processing, vol. 2, pp. 338–342, Oct. 1999.

[10] H.W. Kuhn, "The Hungarian Method for the assignment problem", Naval Research Logistics Quarterly, 2: 83–97, 1955.

[11] R.E. Kalman, “A New Approach to Linear Filtering and Prediction Problems”, Journal of Basic Engineering, Vol.82, pp. 35-45, 1960.