An Image-based Fall Detection System using You Only Look Once (YOLO) Algorithm to Monitor Elders' Fall Events

Koh Zhen Long[1], Hazlina Haron[1], Maizura Ibrahim[2] and Zeti Darleena Eri[3]

[1]*SEGI College Subang Jaya, Selangor, Malaysia, {KohZhenLong@segi.edu.my, hazlinaharon@segi.edu.my}*
[2]*Malaysian Nuclear Agency, Kajang, Selangor, Malaysia {maizura@nuclearmalaysia.gov.my}*
[3]*Universiti Teknologi MARA Cawangan Terengganu, Kampus Kuala Terengganu, Malaysia, {zetid415@uitm.edu.my}*

ABSTRACT

Fall is one of the primary causes of fatal as well as non-fatal injuries in the elderly community. The falls in the elderly may cause different consequences and in serious cases, it may cause death. Timely treatment is critical where immediate treatment may reduce the risk of serious injuries. The detection should be taken out in an automated way and detect fall events accurately. This paper presented the image-based fall detection system which integrated the YOLO object detection algorithm with the Image-based Fall Detection system algorithm in detecting fall events. The system will first get track of the person in the video frame with the object detection algorithm and the fall detection algorithm will be used to get track of the person's height and to detect fall events immediately and accurately to notify the caregivers. The system was evaluated with different use cases and conditions. The result shows the system can detect fall events with the accuracy of 92% under the daylight condition and 60% under the low light condition. An email notification will be sent as an alarm to notify the caregivers when any fall events were detected by the system. The quick fall detection and notification of system able to ensure the safety of the elderly were well monitored and timely treatment can take place when fall events were detected by the system.

Keywords: Fall Detection, You Only Look Once, Image-based fall detection, Object-based detection.

I INTRODUCTION

Fall is one of the main life-threaten factors for humans, especially the elderly who live alone. It is caused by the inability of their muscle to support and balance their body due to the aging process. Fall events may cause serious injuries especially in the elderly community and some may be fatal. Several fall prevention solutions had been deployed by different manufactures and industries, but they are still some falls that are unpreventable. Following a fall event, immediate help and treatment are extremely critical. Therefore, fall should be noticed immediately to prevent life-threatening risk (Milat et al., 2011). The outcomes of fall events are far beyond physical injuries as they may also lead to psychological, medical, and social consequences (El-Bendary et al., 2013).

Currently, the improvement of healthcare technology had led to an increase in the average life expectancy of the world population to 80 years old (OECD, 2011). In other words, the percentage of the population with disabilities also growth linearly as people grow older (Iliev et al., 2011). Therefore, due to aging problems, falling events in the elderly increase dramatically. Based on the Public Health Agency of Canada, 12.5% of the population was 65 years old and above in 2001. In the year 2026, the percentage of elderly in the population estimated to increase to 20% (Rougier et al., 2006). The risk of falling for the elderly with age around 65 years old had increased because of the loss of nutrition in their body. Based on the study by Auvinet et al. (2011) and Tinetti (2003), each year, 30% of the elderly community experience a fall event, and half of them fall repeatedly. Immediate aid and treatment are critical and important. Fall without timely treatment could be fatal. To resolve that, fall events should be detected as soon as possible to provide timely treatment for the patient. The immediate treatment could be life rescue for the elderly (Yu, 2008). Therefore, detection of fall events is important for the elderly and people who care for them.

This project aims in developing an automated image-based fall detection system utilizing the YOLOv3 algorithm that can help monitor elderly activity. The fall events will be detected and notified upon detection. Our project proposed to integrate the YOLOv3 object detection algorithm with the IFADS fall detection algorithm to achieve low cost, high accuracy, and real-time computing requirements.

II RELATED WORKS

There are few methods to detect fall events as depicted in Table 1. The wearable device method is impractical for the elderly due to the need for frequent battery and sensor replacement as well as other maintenance work. Besides, the elderly is more likely to be forgetful to wear the device. While the environmental sensor-based method is impractical and not recommended due to the high cost in implementation and maintenance.

Table 1. Different methods for fall detection

Type	Description	Draw-back
Wearable device method	Utilize sensors (i.e., accelerometers, posture sensors) to detect the activity of the users' body (e.g., wrist band) (Estudillo-Valderrama et al., 2009; Doukas et al., 2007)	
Environ-mental sensor or ambient device method	Detect fall by applying environmental sensors such as pressure sensors and radar sensors (Feng et al., 2016)	
Image-based methods	Employing real-time video images and image processing technology to get track of human's activities. (Lu & Chu, 2018)	
Posture recogniti on with surveilla-nce camera	Using the greatest number of moving pixels in the video or images to track a moving person, integrated with the application of posture recognition and classification to classify the fall event (Yu et al., 2012)	False alarm when lying down
Mathema-tical formulati on	Adopting the calculation of aspect ratio and orientation of the person in detecting fall events, then combine with background and multiple person subtractions in a scene to increase the accuracy (Agrawal et al., 2017)	False alarm when lying down
Wearable camera-device with image processin g	Utilize a pattern recognition algorithm namely Histograms of Oriented Gradients (HOG) and gradient local binary patterns, involves extracting density grid from an input image and passed to the support vector machine to classify the combined features (Ozcan et al., 2017; Chua et al., 2013; Jamshed et al., 2015).	Large computati-onal time is required; inappropri ate while doing real-time tracking. Extra battery maintenan ce.

A. Object Detection Algorithm

Nowadays, deep learning had gained its popularity with the introduction of the speech recognition approach in 2006 (Deng L. et al., 2010 and Hinton & Salakhutdinov, 2006), growth of large-scale training data (Deng J. et al., 2010), and improvement of computation power as well as enhancement in network structures design (Zhao et al., 2019). Haar Cascade algorithm was proposed to detect an object with simple Haar-like features. The algorithm implies three main features, i.e., integral images, Adaboost learning algorithm, and cascade classifier (Mohsen Abdul Hossein et al., 2017). Those features had contributed to achieving a high detection rate and rapid capability of image processing for the algorithm (Viola & Jones, 2001).

YOLO object detection algorithm is an algorithm introduced by Redmon et al. (2016). This algorithm uses the one-step scanning approach which differs from the traditional multiple correlated stages approach that is time-consuming and requires high computing power. It is devised to use a single regression method and will only look once at the image. Therefore, the name was short formed as YOLO. This algorithm started by setting the input image into a fixed 2-dimension grid. For each of the grid that contains an object, the grid cell will then predict the bounding box and count the confidence value of the object in the predicted bounding box. While in the bounding box, the predicted bounding box will contain the x and y coordinate, height, and width of the bounding box, and lastly the confidence value of the object. The confidence of the object is calculated by calculating the intersection over the union of the bounding box with the grid cell. While the number of bounding boxes was set while testing, each grid cell will be responsible for predicting only one object hence only the object with the highest confidence value will be shown (Redmon et al., 2016).

B. Fall Detection Algorithm

The fall detection algorithm is introduced by Lu et al. (2018). The Image-Based Fall Detection System (IFADS) algorithm is designed to detect fall events based on the frames captured by a camera. It focuses on tracking the posture state of the person in every frame and fall events will be declared when any suspicious posture changes were detected. It is designed for the detection of fall in real-time video and can be integrated into any of the surveillance cameras. It involved a combination of object detection and a fall detection algorithm. IFADS compares the human's posture states frame by frame and get track of the posture states in every frame. The IFADS algorithm included a process of person detection and fall detection and carried out by a different algorithm (Lu & Chu, 2018).

III METHODOLOGY

The scrum framework (Mellor, 2001) is applied during the development of this system. The main reason is due to the short period of development time. Beside that, it also able to simplify a complicated system into smaller sprint, thus simplified the development process. Figure 1 shows the scrum framework applied in this work

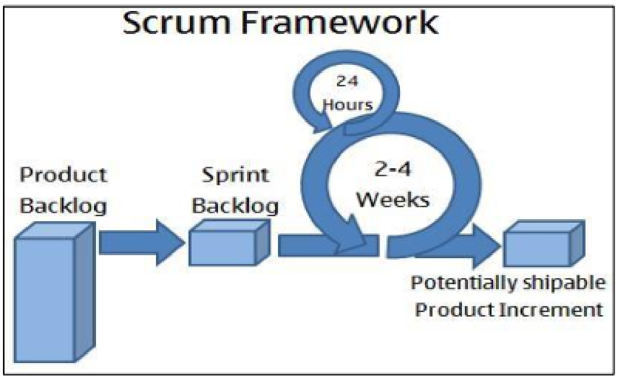

Figure 1. Scrum Framework

The scrum framework will first start by identify the product backlog which carries the used stories about system requirement and each of the requirement will be divided into sprint based on the priority. And once the sprint was developed and reviewed, the product increment can be delivered to the user or client.

IV RESULT AND FINDINGS

Figure 2 illustrates the system flow diagram for the proposed system. The system started with the input camera feed. The video frames will be captured by the camera in real-time. The captured frame will then be extracted into a single frame and to be sent for person detection in the YOLOv3 algorithm. After that, the frame will then be sent to be processed with the fall detection algorithm. If any fall events are detected by the system, an email notification attach with the captured frame will being sent to the caregivers as an alarm to notify them.

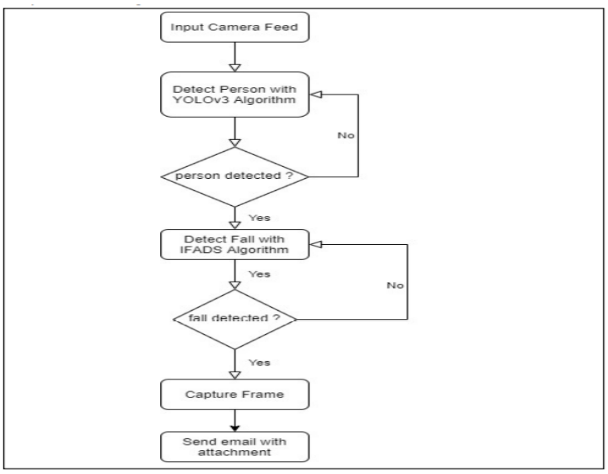

Figure 2. Image-Based Fall Detection System Flow Diagram

The YOLO algorithm used in this system is to detect a moving person and to draw a bounding box surrounding the detected person to get track of the person's movement. The fall detection algorithm will then be calculated based on the tracked person's bounding box. The YOLOv3 which is the enhanced version of the YOLO algorithm was chosen for the proposed system due to the higher accuracy and higher fps to detect an object in real-time (Redmon et al., 2016).

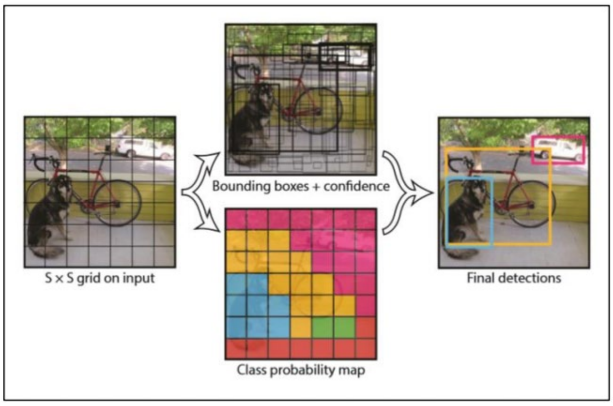

Figure 3. YOLO Algorithm

Figure 3 demonstrates the process of the YOLO object detection algorithm. The input image is first set with a fixed S x S grid. An example of 7 x 7 grid cells is used in Figure 2. Once the grid cell was set, each grid cell then predicts a number of the bounding box. As for this case, 2 bounding boxes are used for each grid cell. Within each of the predicted bounding boxes, there are 5 values which are the x,y,w,h, and the confidence of the detected object. Hence, the non-max suppression will be taken out based on a calculation to refine the bounding box and to localize the object. The IFADS algorithm needs to obtain information detected by the YOLOv3 algorithm to detect fall events.

Figures 4 and 5 reveal the fall detection algorithm using the data generated from an object detection algorithm. The frame loaded from video capture will first appended into a data array. The objective is to store the current frame that is being compared with the data in the frame of 1.5 seconds before because human fall occurred within the maximum time of 1.5 seconds. Fall that happened within 1.5 s are critical and are most likely to affect injuries. Therefore, if the bounding box of that person in the current frame is smaller than 5.5/8, which means lower than 5.5 head of the previous frame, a fall event was identified and detected (Lu & Chu, 2018).

Notation	Definition
Fid	Frame ID of each frame from video
F	Current Frame from video
F^{fid}	Previous Frame from video
FPS	Frame rate per second
H	Current height of the object
H^{fid}	Previous height of object
W	Current weight of the object
W^{fid}	Previous weight in selected frame

Figure 4. Notation List

Fall Detection Algorithm
Input: Webcam Feed
Output: FID, F, H, W
1. Get Video Feed
2. while video Capture:
3. Load object detection model
4. Object detected >> FID, H, W
5. Store FID, H, W in array with FID as index value
6. compare H with $H^{fid-(1.5 * FPS)}$
7. if (H < ((5.5/8) * $H^{(fid-(1.5*FPS))}$))
8. = fall detected
9. Else
10. Loop again

Figure 5. Fall Detection Algorithm

Evaluation on the system accuracy were conducted with different test cases and condition. The proposed system was tested with different scenarios and test cases to evaluate on the accuracy of fall event detection. The different cases that were carried out to be evaluated are: (1) fall events that occurs while standing under a daylight condition, (2) fall events that occurs while sitting under a daylight condition, (3) fall events that occurred under a low light condition. Beside of different lighting condition, different camera distance had also been used to evaluate the detection accuracy. The evaluation results will be reveals in next sections.

A. Fall events occurred while standing under day light condition.

As to ensure the fall detection system can work under different situations and use cases, different test cases had been carried out in evaluating the accuracy and performance of the system. Table 2 depicts the evaluation result for the fall events that occurred while standing under a daylight condition. It shows 28 fall scenes were evaluated with the system under the day light condition. For the 28 fall events, 23 was detected accurately and notification was sent by the system. The accuracy was 82% in detecting the fall events.

Table 2. Fall events while standing under daylight

Test Cases	Actual Fall	Fall Detected	Accuracy
Total Cases	28	23	82%

Table 3. Fall events while sitting under daylight

Test Cases	Actual Fall	Fall Detected	Accuracy
Total Cases	25	23	92%

Table 4. Fall events under lowlight condition

Test Cases	Actual Fall	Fall Detected	Accuracy
Total Cases	10	6	60%

B. Fall events occurred while siting under day light condition

Besides of fall events while standing, fall events occurred while sitting were also evaluated. As shown in Table 3, 25 fall scenes were being evaluated with the fall detection system. Out of 25 actual fall scenes, 23 of them managed to be detected by the system. Hence the accuracy of fall detection for fall events occurred while sitting under day light condition can reach up to 92%. The other 2 fall events that were not able to detect by the system was due to the view angle of the fall events was no obvious and hence the fall events were not being detected by the system.

C. Fall events occurred under lowlight condition

Furthermore, low light condition will also need to be evaluated as fall events might happens during nighttime or low light condition as well. Fall occurred at night is more critical and dangerous as it may happen when everyone is at sleep and could cause the fall events not being noticed by the caregivers. Therefore, 10 fall scenes under low light condition had been evaluated by the fall detection system as shown in Table 4. Out of 10 fall events, only 6 fall events were detected by the system. The accuracy of fall detection only reaches up to 60% under low light condition. This is because the low light condition caused a loss detection on the scenes due to the noise of pixels captured under low light condition. The accuracy however could be improved by implementing a night-vision camera which can capture clear scene event under low light condition.

D. Efficiency Evaluation

The efficiency testing is done to test the frame rate per second and the confidence value of the YOLOv3 algorithm in detecting person. The input image is resized into different resolution before sending into the YOLOv3 algorithm. The resolution may affect the confidence and performance of the real-time tracking. Therefore, this test was carried out to identify the resolution with highest efficiency. Table 5 and Figure 6 shows the evaluation result for the confidence of the YOLOv3 object detection algorithm in detecting person in different resolution.

The results revealed the image that is resized to 224 x 224 resolution has the highest confidence in classifying object. This means that the YOLOv3 object can detect object in this resolution more accurately. The efficiency rate of the overall object detection performance was calculated by multiplying the average confidence rate with the frame rate per second during run time. The efficiency with highest rate is more efficiency and will be chosen to be implement into the fall detection system. Equation to calculate the efficiency rate: Efficiency rate = confidence * fps.

Table 5. Fall events occurred under low light condition

	Confidence of object				
Resolution	**90^2**	**190^2**	**224^2**	**320^2**	**700^2**
Average	0.980417	0.991568	0.999103	0.998290	0.994267

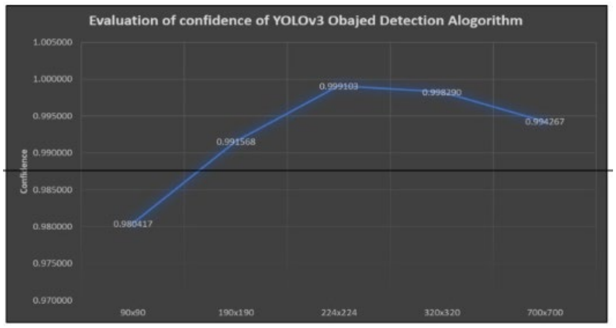

Figure 6. Evaluation graph

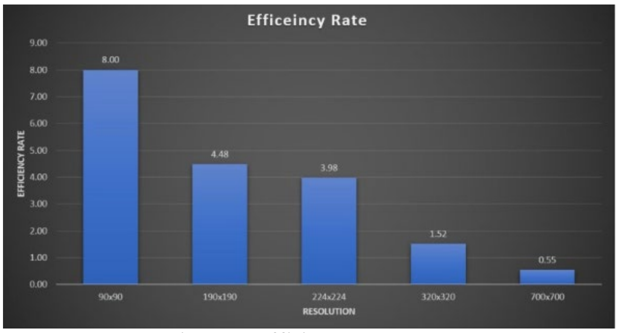

Figure 7. Efficiency rate graph

Figure 7 presents the result for the efficiency evaluation of the YOLOv3 object detection algorithm with different resolution. The result revealed that the detection with 90 x 90 has the highest efficiency rate. Therefore, the proposed system resized the input video stream to the 90 x 90 resolution to be fitted into the YOLOv3 object detection algorithm.

E. Accuracy evaluation with different camera distance

Besides of lighting condition, the location of the camera is also one of the main factors that may affect the accuracy of the fall detection system. Therefore, the system was also being evaluated by placing camera in different location with difference distance to test on the optimum distance to implement the camera with highest accuracy of fall detection.

Table 6. Camera Distance Evaluation

Distance	Fall Detection Accuracy
1.5 Meters	93%
3 Meters	87%
4 Meters	67%
>4.5 Meters	53%
Average	75%

Table 6 depicted the fall scene tested by implementing the camera in different distances from the victims. There are 4 distance ranged from 1.5 meters to 4.5 meters and above. This evaluation is important because the distance of the camera will affect the fall detection accuracy due to the pixels which represented the person in the captured frame will be smaller as a person move further from the camera. The optimum distance would be 1.5 meters as the accuracy in this distance can reach up to 93% while the accuracy decrease as the camera goes further and the maximum distance that can be captured by the system will be less than 4.5 meters. Any fall events occurred further than 4.5 meters will not be detected accurately. Figure 8 presents the overall accuracy for the fall detection system running in different conditions while Figure 9 reveals the test graph for different camera distance.

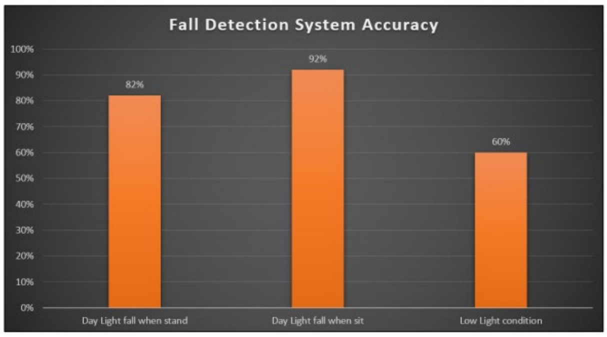

Figure 8. Evaluation Result 1

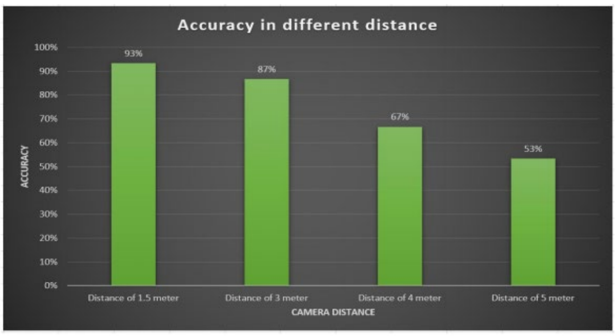

Figure 9. Evaluation Result 2

V CONCLUSION AND FUTURE WORK

The combination of YOLOv3 and the Image-based fall detection system proposed in this paper is one of the sub-modules created to detect swift movement for a knowledge-based and integrated security system. The proposed fall detection system is easy to integrate to any camera. One of the examples of implementation is to integrate it with a surveillance camera. Besides, the propose system is very cost efficient because no additional device are needed. Only a camera and a processor unit are needed. The accuracy of the detection can reach up to 92% during daylight condition. However, there are several limitations are identified, such as the performance issue and false detection when there are too many people captured by the camera. These limitations will be resolved in the future work.

ACKNOWLEDGMENT

The authors would like to express their gratitude to SEGI College Subang Jaya for the support and facilities provided. This research paper is supported by the International Atomic Energy Agency (IAEA) research grant under the IAEA Coordinated Research Project CRP J02011 (Contract No. 22165).

REFERENCES

Agrawal, S. C., Tripathi, R. K., & Jalal, A. S. (2017). Human-fall detection from an indoor video surveillance. 8th International Conference on Computing, Communications and Networking Technologies (ICCCNT 2017).

Auvinet, E., Multon, F., Saint-Arnaud, A., Rousseau, J., & Meunier, J. (2011). Fall detection with multiple cameras: An occlusion-resistant method based on 3-D silhouette vertical distribution. IEEE Transactions on Information Technology in Biomedicine, 15(2), 290–300.

Mellor, S., Schwaber, K., Sutherland, J., & Thomas, D. (2001). Manifesto for Agile Software Development. The Agile Alliance.

Chua, M., Aha, D. W., Auslander, B., Gupta, K., & Morris, B. (2013). Comparison of object detection algorithms on maritime vessels. Technical Report AIC-14-041, Naval Research Laboratory, Navy Center for Applied Research in Artificial Intelligence.

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2010). ImageNet: A large-scale hierarchical image database.

Deng, L., Seltzer, M., Yu, D., Acero, A., Mohamed, A., & Hinton, G. (2010). Binary coding of speech spectrograms using a deep auto-encoder. Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010.

Doukas, C., Maglogiannis, I., Tragas, P., Liapis, D., & Yovanof, G. (2007). Patient fall detection using support Vector Machines. IFIP International Federation for Information Processing.

El-Bendary, N., Tan, Q., Pivot, F. C., & Lam, A. (2013). Fall detection and prevention for the elderly: A review of trends and challenges. International Journal on Smart Sensing and Intelligent Systems, 6(3), 1230–1266.

Estudillo-Valderrama, M. Á., Roa, L. M., Reina-Tosina, J., & Naranjo-Hernández, D. (2009). Design and implementation of a distributed fall detection system-personal server. IEEE Transactions on Information Technology in Biomedicine.

Feng, G., Mai, J., Ban, Z., Guo, X., & Wang, G. (2016). Floor Pressure Imaging for Fall Detection with Fiber-Optic Sensors. IEEE Pervasive Computing.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science.

Iliev, I. T., Tabakov, S. D., & Dotsinsky, I. A. (2011). Two Steps Approach for Falls Detection in the Elderly. Annual Journal of Electronics, 5(2), 46–48.

Jamshed, M., Parvin, S., & Akter, S. (2015). Significant HOG-Histogram of Oriented Gradient Feature Selection for Human Detection. International Journal of Computer Applications, 132(17), 20–24.

Lu, K. L., & Chu, E. T. H. (2018). An image-based fall detection system for the elderly. Applied Sciences (Switzerland), 8(10).

Milat, A. J., Watson, W. L., Monger, C., Barr, M., Giffin, M., & Reid, M. (2011). Prevalence, circumstances and consequences of falls among community-dwelling older people: results of the 2009 NSW Falls Prevention Baseline Survey. New South Wales Public Health Bulletin, 22(3–4), 43–48.

Mohsen Abdul Hossen, A., Abd Alsaheb Ogla, R., & Mahmood Ali, M. (2017). Face Detection by Using OpenCV's Viola-Jones Algorithm based on coding eyes للرؤية مفتوحة برمجية مكتبة باستخدام الوجه كشف اساس على جونز ال فيو ارزمية خو في الحاسوبية العيون شفرة. Iraqi Journal of Science, 58(2A), 735–745.

OECD. (2011). Health indicators: Life expectancy. Society At A Glance 2011: OECD Social Indicators, 78–79.

Ozcan, K., Velipasalar, S., & Varshney, P. K. (2017). Autonomous Fall Detection with Wearable Cameras by Using Relative Entropy Distance Measure. IEEE Transactions on Human Machine Systems.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, realtime object detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Rougier, C., Meunier, J., St-Arnaud, A., & Rousseau, J. (2006). Monocular 3D head tracking to detect falls of elderly people. Annual International Conference of the IEEE Engineering in Medicine and Biology Proceedings, February, 6384–6387.

Tinetti, M. E. (2003). Preventing falls in elderly persons. New England Journal of Medicine, 348(1), 42–49.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1(February).

Yu, M., Rhuma, A., Naqvi, S. M., Wang, L., & Chambers, J. (2012). A posture recognition based fall detection system for monitoring an elderly person in a smart home environment. IEEE Transactions on Information Technology in Biomedicine.

Yu, X. (2008). Approaches and Principles of Fall Detection. Healthcom, 10(1), 42–47.

Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object Detection with Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems, 30(11), 3212–3232.