Moving object detection via TV-L1 optical flow in fall-down videos

Nur Ayuni Mohamed, Mohd Asyraf Zulkifley
Center for Integrated Systems Engineering and Advanced Technologies, Faculty of Engineering and Built Environment,
Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor Malaysia

Article Info

ABSTRACT

There is a growing demand for surveillance systems that can detect fall-down events because of the increased number of surveillance cameras being installed in many public indoor and outdoor locations. Fall-down event detection has been vigorously and extensively researched for safety purposes, particularly to monitor elderly peoples, patients, and toddlers. This computer vision detector has become more affordable with the development of high-speed computer networks and low-cost video cameras. This paper proposes moving object detection method based on human motion analysis for human fall-down events. The method comprises of three parts, which are preprocessing part to reduce image noises, motion detection part by using TV-L1 optical flow algorithm, and performance measure part. The last part will analyze the results of the object detection part in term of the bounding boxes, which are compared with the given ground truth. The proposed method is tested on Fall Down Detection (FDD) dataset and compared with Gunnar-Farneback optical flow by measuring intersection over union (IoU) of the output with respect to the ground truth bounding box. The experimental results show that the proposed method achieves an average IoU of 0.92524.

Corresponding Author:

Nur Ayuni Mohamed,
Center for Integrated Systems Engineering and Advanced Technologies,
Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia,
43600 Bangi, Selangor Malaysia.
E-mail: ayuni@siswa.ukm.edu.my

1. INTRODUCTION

Object detection is one of the research topics in computer vision which detects the presence of objects of interest and locates their positions. The main task of object detection method is to segment any moving objects from the background [1]. Thus, the segmented object of the interests is then labeled as 'foreground' and the rest of the pixels are labeled as 'background'. With the advancement of image and video fields, many automatic complex systems can be designed with the help of the high-resolution cameras and high-speed computer networks. Hence, the role of automatic detection algorithm has become more important, especially for the applications that focus on in daily life such as behavioral analysis [2-3], urban surveillance [4-5], and object recognition [6-8].

Whilst, a human fall-down event is defined as an incident in which the body of the person of interest halts or rests unintentionally laying on the ground or any other lower surfaces [9]. A fall-down event also takes place when a person slips unexpectedly while walking or standing. The World Health Organization (WHO) [10] has also reported that fall-down event is a major public health concern, which is the second leading cause of unintentional injury death after road accident worldwide. In addition, WHO also reported that an estimated of 646 000 individuals died globally each year because of fall-down event related incidents. Figure 1 shows some samples of fall-down events for various fall postures.

Figure 1. Image samples of fall-down events

Therefore, automatic fall-down event sensor is very crucial for the applications in hospitals, elderly houses, and other public places. Early detection of the event is crucial to reduce the "long lie" situation which is the time period where a subject remains lying on the floor after the fall incident. It is the key factor that determines the health impact severity of a fall-down event. Clinical studies also have shown that a long lie case usually leads to dehydration, hypothermia and pressure sores [11, 12]. In addition, it might also lead to psychological consequences such as loss of independence ability, fear f falling and trauma [13].

The proposed work is based on a vision approach that does not require too much complex processing of the videos. The basis of fall detector is TV-L1 optical flow algorithm. The main motivation behind this work is due to the fact that there is a relatively large number of peoples who die owing to the late awareness and treatment of fall-down event after the incidents have occurred. Hence, an automatic and efficient system based on motion analysis is much needed to mitigate this problem. This paper is organized as follows: Section 2 discusses related works on object detection in fall-down events. Section 3 explains the proposed method of object detection, Section 4 shows the experimental results and performance comparison, and a conclusion is provided in Section 5.

2. RELATED WORK

Object detection techniques for fall-down events have become a big subtopic under computer vision and image processing field. Generally, these techniques are tremendously explored because of its less intrusive behavior as well as robust and easy to be implemented in various environments. In addition, a vast amount of information can also be extracted concurrently from the surveillance cameras such as motions, locations or actions of the monitored person of interest [14]. Typically, object detection methods in the fall-down event can be divided into two approaches; background subtraction and optical flow.

2.1. Background subtraction

Background subtraction is the most frequently used method in fall-down event detection that finds the differences between the established background model and the current image so that the foreground object can be segregated from background. It is used mainly for static cameras set up. Basically, background subtraction techniques provide fast computation algorithm with good accuracy. Poonsri & Chiracharit [15] used a mixture of Gaussian model (MoG), which is a statistical approach to extract the foreground objects. Then, they merged the results of MoG method with the mean filter. They also implemented some morphological operations to remove the noise to improve foreground detection accuracy. However, MOG is sensitive to detect all the moving objects which usually lead to false foreground detection.

Yu et al. [16, 17] applied background subtraction method using codebook algorithm to extract the foreground silhouette for a single camera system. They argued that codebook algorithm can achieve better performance by utilizing more comprehensive information from the color space. They also stated that their approach is capable to cope with illumination changes and adaptive parametric variation since no assumption is made as compared to other background methods such as MoG and single-mode background subtraction method in [18, 19].

Wang et al. [20] implemented background subtraction VIBE+ method to extract the foreground object. Then, they performed connected component analysis to combine and label the components as the foreground. Besides that, Yun et al. [21] performed background subtraction method using Gaussian Mixture

Models (GMM) to depth images instead of RGB images. The foreground is extracted such that the mixture models can determine the most probable pixel's class by modelling the intensity values. They further implemented a series of morphological operations to remove the noise to obtain a clean silhouette of the foreground object.

2.2. Optical flow

Optical flow [22] is defined as the movement pattern of the object's motion contained in a video. The motions which are usually represented by velocity are estimated based on similar points of the two consecutive frames. Optical flow can give complete movement information of the whole frame and suitable to be implemented as a moving object detector. Bhandari et al. [23] used Lukas-Kanade optical flow for the motion estimation of the foreground. Their algorithm first finds points of interest using Shi-Tomasi method applied to the output of Harris-Stephens corner detection with several threshold parameters. The points of interest are kept if the flow of any respective points is matched with a corner point and their distance difference is less than a small number. Otherwise, the points will be discarded. The same process is repeated until the end of a video.

Paper in [21] utilized Horn-Schunk and Lucas-Kanade optical flow methods to detect foreground objects motions in RGB depth videos. Then, they extracted histogram-based features of the optical flow. This histogram is a useful measure to describe motions of the foreground pixels, which later used to classify the fall-down event.

Belshaw et al. [24] applied Farneback optical flow to represent motion of the foreground blob between consecutive video frames. Optical flow is also incorporated into background adaptation approach so that background model can be updated to cater multiple active region blobs. Specifically, magnitude of the optical flows is used to control background adaptation rates. This method is devised based on assumption that motion cues can be used to remove stationary blobs as well as to identify lighting changes.

Alaoui et al. [25] proposed a combination method between Farneback optical flow algorithm with Von Mises distribution to determine and identify the moving object. Background subtraction and morphological operations are performed first before extracting the foreground pixels. Then, optical flow is used to calculate motion vectors of the foreground object. Later, Von Mises distribution is implemented to calculate the mean direction of the vectors. Apart from foreground extraction, the proposed method is able to determine the orientation vectors of the object before and after the event.

3. RESEARCH METHOD

Figure 2 shows flowchart of the proposed framework of moving object detection that comprising of three stages; preprocessing, object detection and performance measure. The proposed method utilizes optical flow approach to detect the moving object, in which a single person in the case of walk and fall videos.
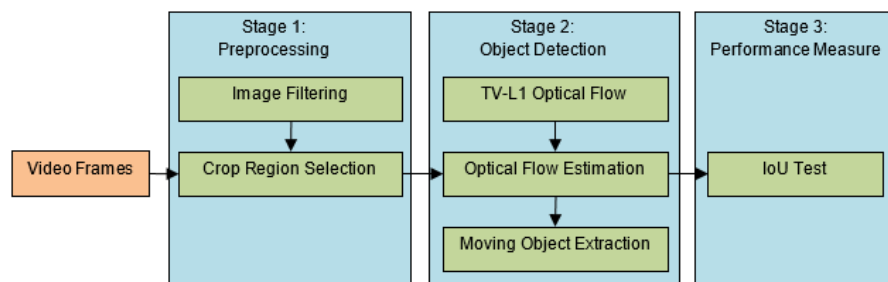

Figure 2. Flowchart of the proposed framework

3.1. Preprocessing

The preprocessing part focuses on producing better input image to the moving detection part. Image filtering using 5×5 averaging filter as in (1) is implemented to smooth out the noise in the input videos. The underlying assumption used is any raw input videos are not suitable to be directly processed because of the noisy pixels and illumination variations that present in the video. Therefore, some preprocessing techniques need to be performed first to reduce the effect of the previously mentioned issues. Then, each video frame is cropped to obtain region that consists only the moving objects.

$$K = \frac{1}{25}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{1}$$

3.2. Object detection

Optical flow is one of the moving object detection methods which approximate the apparent motion of the pixel's brightness between two consecutive frames. Optical flow can also be used to represent the velocity of the pixels. The TV-L1 optical flow [26] is implemented in this work to determine and detect the moving object based on OpenCV implementation. The TV-L1 optical flow is chosen due to better performance under various lighting conditions compared to other optical flow algorithms [13]. It is also a fast computation algorithm with comparable accuracy and ability to deal with occluded areas so that flow distortion can be prevented [27].

In general, TV-L1 optical flow is one of the variational methods for the optical flow estimation and it has become more popular and extensively researched because of its robustness and accuracy. Basically, the underlying idea of this optical flow is the brightness between two images remains similar under motion and sometimes coined as brightness constancy assumption [28]. Thus, TV-L1 optical flow is defined as a combination of the brightness and gradient constancy assumptions but with varying the weight under the Chambolle function in the regularization term [29] with respect to classical Horn-Schunk approach [30]. Moreover, the previous studies have shown that the combination between both brightness and gradient have led to a robust flow estimation under various illumination changes [31] and image noises. The average magnitude of optical flow vector is computed in each video frame to infer the predicted bounding box. These boxes are then used in the next module for the performance measure purpose of identifying the fall-down event.

Let $\mathbf{u} = \{u, v\}$ be the displacement field at pixel coordinate $\mathbf{x} = \{x, y\}$. The optical flow can be written in the non-linear formulation as (2) with I_i and I_{i+1} are the current and future frames, respectively. The equation can be linearized using Taylor expansion as in (3) with $\mathbf{u^0}$ as an approximation to $\mathbf{u}$.

$$I_{i+1}(\mathbf{x} + \mathbf{u}) - I_i(\mathbf{x}) = 0 \tag{2}$$

$$\rho(\mathbf{u}) = \nabla I_{i+1}(\mathbf{x} + \mathbf{u^0}) \cdot (\mathbf{u} - \mathbf{u^0}) - I_{i+1}(\mathbf{x} + \mathbf{u^0}) - I_i(\mathbf{x}) \tag{3}$$

(3) assumes that pixel intensities are constant over time which is not practical in the real-life scenario. Thus, this equation can be modeled with an additional function ω with weight γ as in (4).

$$\rho(\mathbf{u}) = (\nabla I_{i+1})^T (\mathbf{u} - \mathbf{u^0}) + I_t + \gamma \omega \tag{4}$$

The L^1 penalization for both regularization and data term can be optimized by minimizing the energy function as in (5), where λ is the trade-off between regularization and data term. However, (5) is not trivial to be solved as an optimization problem. Thus, (6) is used to solve the problem by introducing the convex relaxation term with $\mathbf{p}$ is another auxiliary variable as $\mathbf{u}$ and θ is a constant, in which the goals is to minimize the mentioned energy function.

$$E = min_{(u,v)} \left\{ \lambda \int_\Omega \|\rho(u,v)\|_1 + \int_\Omega \|u\|_1 + \|v\|_1 \right\} \tag{5}$$

$$E = min_{(u,p)} \left\{ \lambda \int_\Omega \rho(\boldsymbol{p}) + \int_\Omega \|u\| + \|v\| + \frac{1}{2\theta} \|\boldsymbol{u} - \boldsymbol{p}\|^2 \right\} \tag{6}$$

3.3. Performance measure

The ground truth localization of the moving object in each frame is annotated manually by using a bounding box that surrounds the object. These boxes are then used to evaluate the Intersection over Union (IoU) between the ground truth boxes and the output boxes of the proposed method. The IoU or also known as the Jaccard index is defined in (7) and illustrated in Figure 3.

$$IoU = \frac{\text{Area of overlap}}{\text{Area of union}} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{7}$$
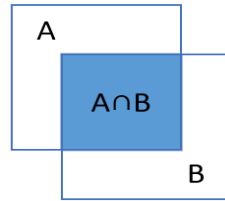
Figure 3. Illustration of the intersection over union (IoU)

4. RESULTS AND ANALYSIS

The proposed framework is tested using Fall Detection Dataset (FDD) [32], which is an online fall-down database that comprises a total of 126 annotated videos. These videos contain four different stages with nine different persons acting the fall incident. The frame rate is 25 frames/sec with a frame size of 320×240 pixels. The proposed framework is also employed using Python 3.6 on Intel Core i7 3.4 GHz 12 GB RAM desktop computer.

Figure 4 shows the sample results of the proposed method using FDD database. The performance of the proposed method is compared with the Gunnar-Farneback optical flow [33], which is another type of dense optical flow. Figure 4(a) shows the sequential frames of the FDD, while Figure 4 (b) and Figure 4 (c) show the corresponding optical flow images of the TV-L1 and Gunnar-Farneback optical flow.

(a)
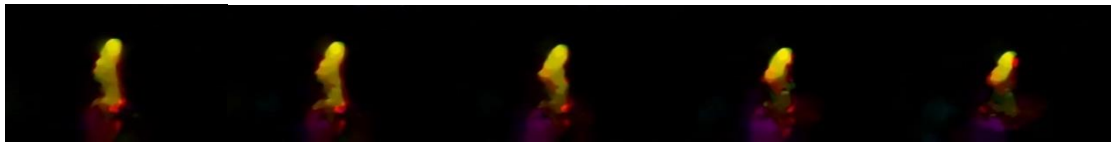
(b)
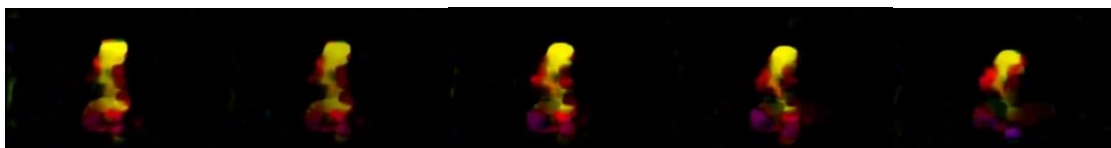
(c)

Figure 4. (a) Samples of sequential frames of Fall Detection Dataset, (b) Corresponding optical flow images of TV-L1 algorithm, (c) Corresponding optical flow images of Gunnar-Farneback algorithm

Figure 5 and 6 show the graphs of the average flow magnitude of the TV-L1 and Gunnar-Farneback algorithms for all sequential frames in Figure 4. From the graphs, the Gunnar-Farneback optical flow method produces higher average magnitude compared to TV-L1 optical flow. This is because Gunnar-Farneback cannot cope well with the noise as it considers them as moving pixels and thus higher moving magnitude.
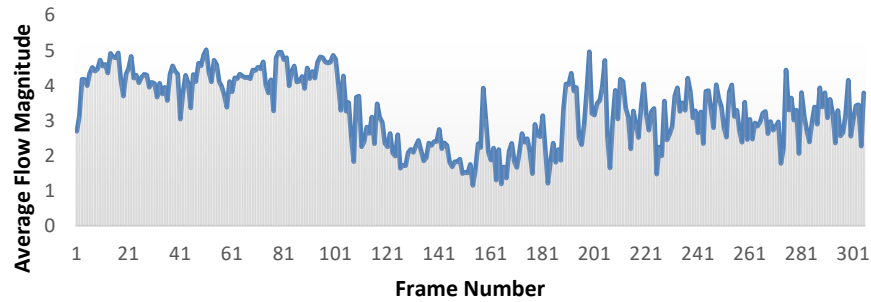
Figure 5. Average flow magnitude of TV-L1 optical flow for all sequential frames in Figure 4
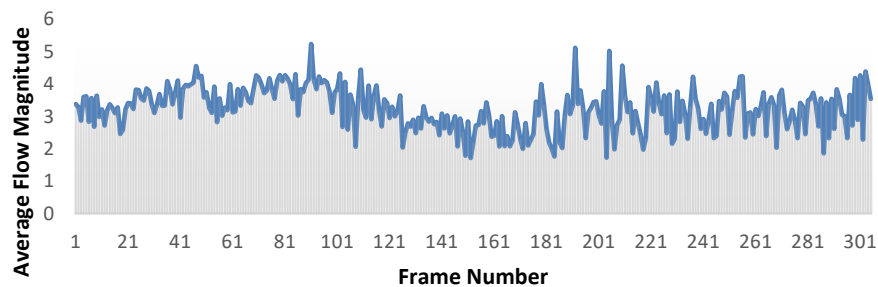

Figure 6. Average flow magnitude of Gunnar-Farneback optical flow for all sequential frames in Figure 4

The IoU test is then performed on both outputs of the optical flow algorithms. The illustration of the IoU is shown in Figure 7 with green bounding box is the ground truth of the FDD and the red bounding box is the optical flow output bounding box. Table 1 lists the computation IoU results for both TV-L1 and Gunnar-Farneback optical flow for 14 videos. In average, the IoU of TV-L1 optical flow is higher compared to Gunnar-Farneback method with an average of 0.92524 compared to 0.92346. However, Gunnar-Farneback method produces higher IoU results for Video 3, 8 and 9 because of the low noise videos. Even though the IoU differences are not too big, it still gives a big impact for fall-down detection, especially during the transition period between just before and after the fall-down incident.
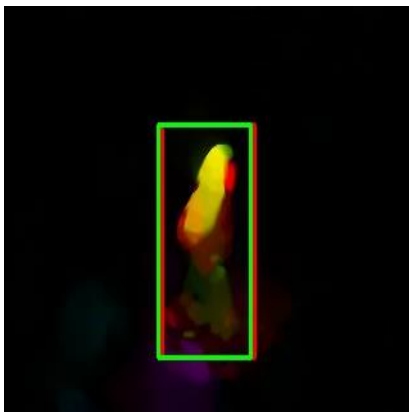

Figure 7. The illustration of IoU

Table 1. IoU results for TV-L1 and Gunnar-Farneback optical flow

| #Video | Intersection over union (IoU | |
	TV-L1	Gunnar-Farneback
1	0.92227	0.92189
2	0.91823	0.91685
3	0.91786	0.91998
4	0.91289	0.91086
5	0.91776	0.91516
6	0.94445	0.93651
7	0.92682	0.92467
8	0.93954	0.94199
9	0.93460	0.93694
10	0.91080	0.90927
11	0.93145	0.92838
12	0.92236	0.91979
13	0.94605	0.93929
14	0.90834	0.90697
Average	0.92524	0.92346

5. CONCLUSION

In conclusion, a moving object detection using TV-L1 optical flow for fall-down videos has been proposed and tested. The proposed framework starts with the preprocessing step followed by the computation of the optical flow algorithm which is TV-L1 optical flow. The average flow magnitude is then computed for each frame to obtain the output bounding box. Then, this box is compared with the ground truth data using IoU test. The performance of the proposed method is benchmarked with the Gunnar-Farneback optical flow. Based on the experimental results, TV-L1 optical flow achieved an average IoU with 0.92524 which outperforms the Gunnar-Farneback optical flow. For future work, the detector can be further by using additional features such as titled angle, middle-points, and motion speed.

REFERENCES

[1]   Parekh HS, Thakore DG, Jaliya UK. "A survey on object detection and tracking methods". *International Journal of Innovative Research in Computer and Communication Engineering*; 2(2): 2970-2978. 2014.
[2]   Yang CJ, Chou T, Chang FA, Ssu-Yuan C, Guo JI. "A smart surveillance system with multiple people detection, tracking, and behavior analysis". *International Symposium on VLSI Design, Automation and Test (VLSI-DAT)*. pp. 1-4. 2016.
[3]   Merad D, Aziz KE, Iguernaissi R, Fertil B, Drap P. "Tracking multiple persons under partial and global occlusions: Application to customers' behavior analysis". *Pattern Recognition Letters*; 81: 11-20. 2016.
[4]   B. Tian, Y. Li, B. Li and D. Wen, "Rear-View Vehicle Detection and Tracking by Combining Multiple Parts for Complex Urban Surveillance," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 597-606, April 2014.
[5]   Song JF. "Vehicle Detection Using Spatial Relationship GMM for Complex Urban Surveillance in Daytime and Nighttime". *International Journal of Parallel Programming*.; 46(5): 859-872. 2018.
[6]   Phalke K, Hegadi R. "Object tracking using Mutiple Hypothesis". *International Conference on Signal Processing and Integrated Networks (SPIN)*.: 707-710. 2016.
[7]   Ramalingam SP, Nadesh RK, SenthilKumar NC. "Robust Face Recognition Using Enhanced Local Binary Pattern". *Bulletin of Electrical Engineering and Informatics (BEEI)*.; 7(1): 96-101. 2018.
[8]   Sudhakar K, Nithyanandam P. "An Accurate Facial Component Detection Using Gabor Filter". *Bulletin of Electrical Engineering and Informatics(BEEI)* ; 6(3): 287-294. 2017.
[9]   Ahmed M, Mehmood N, Nadeem A, Mehmood A, Rizwan K. "Fall detection system for the elderly based on the classification of shimmer sensor prototype data". *Healthcare informatics research*.; 23(3): 147-158. 2017.
[10] World Health Organization. Falls. Geneva: World Health Organization, 2018.
[11] Dorgham O, Rass SA, Alkhraisat H. "Improved Elderly Fall Detection by Surveillance Video using Real-Time Human Motion Analysis". *International Journal of Soft Computing*.; 12(4): 253-262. 2017.
[12] Igual R, Medrano C, Plaza I. "Challenges, issues and trends in fall detection systems". *Biomedical engineering online*.; 12(1): 66. 2013.
[13] Núñez-Marcos A, Azkune G, Arganda-Carreras I. "Vision-based fall detection with convolutional neural networks". *Wireless Communications and Mobile Computing*.; 1-16. 2017.
[14] C. Rougier, J. Meunier, A. St-Arnaud and J. Rousseau, "Robust Video Surveillance for Fall Detection Based on Human Shape Deformation," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 5, pp. 611-622, May 2011.
[15] Poonsri A, Chiracharit, W. "Fall detection using Gaussian mixture model and principle component analysis". *International Conference on "Information Technology and Electrical Engineering (ICITEE)*.: 1-4. 2017.
[16] Yu M, Gong L, Kollias S. "Computer vision-based fall detection by a convolutional neural network". *International Conference on Multimodal Interaction*.: 416-420. 2017.
[17] M. Yu, A. Rhuma, S. M. Naqvi, L. Wang and J. Chambers, "A Posture Recognition-Based Fall Detection System for Monitoring an Elderly Person in a Smart Home Environment," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1274-1286, Nov. 2012.
[18] C. R. Wren, A. Azarbayejani, T. Darrell and A. P. Pentland, "Pfinder: real-time tracking of the human body," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780-785, July 1997.
[19] Horprasert T, Harwood D, Davis, LS. "A statistical approach for real-time robust background subtraction and shadow detection". *IEEE International Conference on Computer Vision*.; 99: 1-19. 1999.
[20] Wang K, Cao G, Meng D, Chen W, Cao W. "Automatic fall detection of human in video using combination of features". *International Conference on Bioinformatics and Biomedicine (BIBM)*.: 1228-1233. 2016.
[21] Yun Y, Innocenti C, Nero G, Gu IY. "Fall detection in RGB-D videos for elderly care". *International Conference on E-health Networking, Application & Services (HealthCom)*.: 422-427. 2015.
[22] Chauhan A. K., & Krishan, P. "Moving object tracking using gaussian mixture model and optical flow". *International Journal of Advanced Research in Computer Science and Software Engineering*.; 3(4): 243-246. 2013.

[23] Bhandari S, Babar N, Gupta P, Shah N, Pujari S. "A novel approach for fall detection in home environment". *Global Conference on Consumer Electronics (GCCE)*,: 1-5. 2017.

[24] Belshaw M, Taati B, Snoek J, Mihailidis A. "Towards a single sensor passive solution for automated fall detection". *International Conference on Engineering in Medicine and Biology Society*.: 1773-1776. 2011.

[25] Alaoui AY, El Hassouny A, Thami RO, Tairi H. "Video based human fall detection using von mises distribution of motion vectors". *Intelligent Systems and Computer Vision (ISCV)*.: 1-5. 2017.

[26] Pérez JS, Meinhardt-Llopis E, Facciolo G. "TV-L1 optical flow estimation". *Image Processing On Line*.: 137-150. 2013.

[27] He L, Liu X, Pei L. "A high-precision optical flow method based on TV-L1 model and neighborhood patch". *International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII)*.: 321-324. 2017.

[28] Hofinger M, Pock T, Moosbrugger T. "Robust Deformation Estimation in Wood-Composite Materials using Variational Optical Flow". arXiv preprint arXiv:1802.04546.: 1-8. 2018.

[29] C. Zhang, Z. Chen, M. Wang, M. Li and S. Jiang, "Robust Non-Local TV-L1 Optical Flow Estimation With Occlusion Detection," in *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 4055-4067, Aug. 2017.

[30] Horn, BK, Schunck BG. "Determining optical flow". *Artificial intelligence*.; 17(1-3): 185-203. 1981.

[31] Papenberg N, Bruhn A, Brox T, Didas S, Weickert J. "Highly accurate optic flow computation with theoretically justified warping". *International Journal of Computer Vision*.; 67(2): 141-158. 2006.

[32] Charfi I, Miteran J, Dubois J, Atri M, Tourki R. "Optimised spatio-temporal descriptors for real-time fall detection: comparison of SVM and Adaboost based classification". *Journal of Electronic Imaging (JEI)*.; 22(4): 041106. 2013.

[33] Farnebäck G. "Two-frame motion estimation based on polynomial expansion". *Scandinavian Conference on Image Analysis*.: 363-370. 2003.