Bulletin of Electrical Engineering and Informatics
Vol. 7, No. 3, September 2018, pp. 359~366
ISSN: 2302-9285, DOI: 10.11591/eei.v7i3.1184

A Finite State Machine Fall Detection Using Quadrilateral Shape Features

M. F. Abu Hassan, M. H. M. Saad, M. F. Ibrahimand A. Hussain
Centre for Integrated Systems Engineering and Advanced Technologies (INTEGRA),
Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia,
Bangi 43600, Selangor, Malaysia

Article Info

Article history:

Received March 28, 2018
Revised Aug 10, 2018
Accepted Aug 24, 2018

Keywords:

Fall detection
Finite state machine
Pose recognition
Quadrilateral shape features
Shape generalization

ABSTRACT

A video-based fall detection system was presented; which consists of data acquisition, image processing, feature extraction, feature selection, classification and finite state machine. A two-dimensional human posture image was represented by 12 features extracted from the generalisation of a silhouette shape to a quadrilateral. The corresponding feature vectors for three groups of human pose were statistically analysed by using a non-parametric Kruskal Wallis test to assess the different significance level between them. From the statistical test, non-significant features were discarded. Four selected kernel-based Support Vector Machine: linear, quadratics, cubic and Radial Basis Function classifiers were trained to classify three human posture groups. Among four classifiers, the last one performed the best in terms of performance matric on testing set. The classifier outperformed others with high achievement ofaverage sensitivity, precision and F-score of 99.19%, 99.25% and 99.22%, respectively. Such pose classification model output was further used in a simple finite state machine to trigger the falling event alarms. The fall detection system was tested on different fall video sets and able to detect the presence offalling events in a frame sequence of videos with accuracy of 97.32% and low computional time.

Corresponding Author:

M. H. M. Saad,
Centre for Integrated Systems Engineering and Advanced Technologies (INTEGRA),
Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia,
Bangi 43600, Selangor, Malaysia.
Email: hanifsaad@ukm.edu.my

1. INTRODUCTION

Falling event detection (FED) based on computer vision is one of the components in realising a smart home-based surveillance system. This feature is essential to ameliorate the existing smart surveillance system in tracking human activities. Various falling events and anomaly movement detection techniques were proposed by researchers for human activity monitoring and surveillance [1], [2], where high performance system is one ofthe key factors to be considered in buildinga smart system. Additionally, a low cost system development, an effective sensor selection and short processing time for algorithm execution are other factors to be considered in realisingan effective real-time tracking system.

Falling eventis an unusual anomaly event that often happens especially to seniors (> 60 years old) [3]. This event is defined as an accidental occurrence that causes a subject to relax at a lower place like on the floor or ground. This event can occur either due to intrinsic factors, such as self-inflicted health like fever, shortness of breath and weak joints or due to extrinsic factors, such as drug with drawal and obstruction of objects [4]. Although such anomaly eventsrarely happen in daily activities, however it can have adverse effects on health and safety in case of occurrence to the subject. Hence, early notification to the respective guardian is extremely important so that an appropriate action can be takento avoid a worse situation from occurring, such as severe injury, disability or death.

The World Health Organisation (WHO) projected that the percentage of senior citizensin 20 selected Western Pacific Region countries will increase by 2030; in which the percentage rates in China, Korea and Japan are the highest (> 30% of total population) [5]. This will bring the countries to an aging nation in the next 12 years. According to WHO, 87% of senior citizens have health problems with non-communicable diseases, such as heart disease, osteoarthritis, stroke, diabetes and Parkinson. These health factors can threaten their safety from falling apart from the extrinsic factors. In addition, this group is likely to suffer from ‘empty nest' syndrome which affects their emotional and health stability. These factors pose a challenge to the community, especially guardians in overseeing theirroutine activities and able to take appropriate fast actions in helping to minimise morbidity as well as cost of medical treatment and mortality.

Rapid growth of computer and software technology gives positive impact on human life, especially in health and safety. As an example, the use of video surveillance systems (VSS) to monitor human activities in particular area. This computer-based system has proven to be helpful in providing useful information for abnormality movements tracking in public areas and workplaces, even in residential areas. However, most VSS, especially for home-based surveillance system are not fully automated and less efficient in detecting anomalies activity, where the supervision and assessment of the activities are typically closely monitored by the guardian or human operator. These tasks require a high level of visual focus and time consuming while they also in volve high remuneration costs. With increase in the number of surveillance cameras in the house or nursing homes, these tasks will not only be more challenging, but it will also increase the cost of development and maintenance. As such the VSS is more likely to record human activiesfor the post-event investigative material purposes. Therefore, a paradigm shift in the use of VSS is essential instead of using it as post-event investigations to prevent the worse occurrence of the unexpected event.

Nowadays, camera technology spreads with extraordinary rapidity. The camera with high resolution with three-dimensional feature is capable toextract high meaningful features for the purpose of classification [6]. While in [2], they proposed a set of motion features using bio-inspired approach (GaussH-BFFNN-PD) in detecting an event into fall and non-fall states. However, the complexity of these high dimensional features is a great challenge in minimising computational time and development cost. Thus, an efficient VSS is indispensable to monitor human activities, particularly in the house in addressing the problem of falling events amongst senior citizens. Therefore, an efficient finite state machine-based FED system by using low-dimensional quadrilateral shape-based featuresis proposed in this article.

2. RESEARCH METHOD

At the first stage, a pose recognition system (PRS) was developed to detect and classify the human pose in an image. The diversities of human postures were categorised in to three groups (denoted as A1, A2 and A3). The first posture group, A1 consists of human performing normal activities images, such as walking and standing. While the second posture group, A2 includes the anomaly actions, such as bending, squatting, crawling, kneeling, sitting and crawling. The last group, A3 consists of second anomaly action images; for example, lying on side, lying down in afacing downward and upward state. The images were acquired from two different databases: CASIA Gait database [7] and Laboratoired ’Electronique, Informatiqueet Image (Le2i) [8]. The first database contributes the A1 set and the second database is used for the anomaly action groups; A2 and A3 as shown in Figure 1. The quadrilateral shape-based features were extracted from the silhouette images and four different types of kernel for Support Vector Machine (KSVM) classifiers were tested to classify the posture groups. The best classifier in terms of performance will be selected as PRS. Then, the PRS output will be fed to the finite state machine (FSM) of falling event detection.

Group of human posture

A1 **A2** **A3**

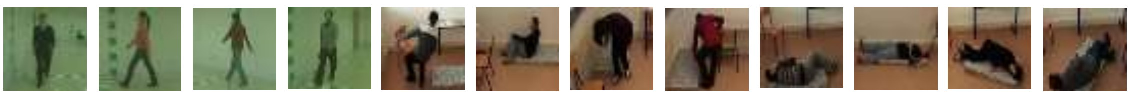


Figure 1. Example of human pose images in three posture groups: A1, A2 and A3

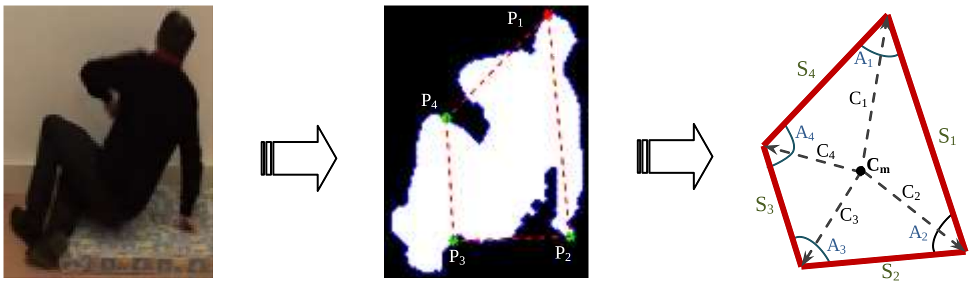
2.1. Pre-processing

The detection of moving objects in a video sequence is a primary step in vision-based systems. Unfortunately, the task becomes difficult due to dynamic changes in natural environment. Thus, various new methodswere proposed to improve the detection of moving objects towords the robustness to shadows, noise and illumination changes [9]. In this work, the two-dimensional images received from the camera will undergo the background subtraction process to detect the moving object. Thecurrent foreground image, F(t) can be extractedfrom the image by comparing every pixel of the current image, I(t) to the background model image, I_b; $F(t)=I(t) - I_b$. This will result inseclusion of the interest object (silhouette) from the background. Then, the image will go through the image treatment process to reducenoise caused by several factors, such as scattered backgrounds and changes in illumination which may affect the formation of silhouette. Therefore, median filter and morphological technique are applied on the F(t) to improve the silhouette image. This non-linear median filter technique does not only produce noise-free images, but it is also able to preserve the edge boundary of a shape in the image rather than the linear filter [10]. Then, the morphology of the image is applied to reduce the imperfections ofshape and structure of the silhouette [11].

The human activities in video datasets were recorded by using an uncalibrated single and multi-stationed camera. During the shooting session, the subjects were directed to freely perform normaland anomaly actionsin a provided room space. Hence, the silhouette size changes in the frame will occur due to the variation of distance and view angle in between the object and camera during the simulation. Therefore, the normalisation of silhouette size is important to ensure every feature vectors extracted from a uniform silhouette size images. The vertical dimension of silhouette, Y will be scaled to a constant dimension, Y’(i.e.100 pixels),whereby the horizontal dimension of silhouette, X will be scaled tothe proportional of variable ratio, *n* between the selected Y’ and Y; where $n=Y'/Y$. Hence, the scaled image; $X'=nX$.

2.2. Quadrilateral Shape Features

The silhouette shapes will be generalised to quadrilateral shape for the purpose of minimising the complexity of posture. The polygonal type shape was chosen by considering the optimum form to represent the human posture for classification. Generally, the quadrilateral shape is derived from fourpoints (vertices) connection located on silhouette boundary. The boundary’s distance was equally partitioned into four parts, where the locations of these ended-parts (points) represent the vertices of quadrilateral shape. The starting point, $P_1$was located at the highest *y*-axis onsilhouette's boundary and the searching order of next point location, P_{i+1} until P_4 was according to clockwise rotationas shown in Figure 2(b). This shape generalisation process will form asimple and common form to represent various silhouette shapes but with unique and distinct features. Three main feature groups were extracted from this quadrilateral shape: centroidal distance (C_i), side length (S_i) and angular angle between vertexes (A_i) as shown in Figure 2(c).



(a) Raw image (b) Segmentation and shape generalization (c) Quadrilateral shape-based features

Figure 2. The shape generalization of silhouette to quadrilateral and features extraction

Overall, 12 feature vectors are extracted from the quadrilateral shape and the feature vectors are defined as follows:

C_i=Distance in between center of mass, C_m and vertex;

S_i=Side length;

A_i=Inner vertex angle.

2.3. Feature Selection

Feature selection is intended to further improve the performance of classification and reduce the processing time [12]. Thus, the entire feature vector setsof each posture group were analysed to identify whether there is statistical significant evidence that each of these quadrilateral-based features is capable of distinguishing the three groups of human posture. The Shapiro Wilk (SW) and the Levene’s (LV) tests were conducted to asses the normality of distribution and homogeneity of variance, respectively. These testswere considered as apre-requisite of determining appropriate statistical test toinvestigate the above hypothesis [13]. The dataset which was neither normally distributed nor had equal variances amongst the groups was subjected to non-parametric Kruskal Wallis (KW) test. Whilst the dataset which was normally distributed with equal variances among groups was subjected to one-way ANOVA (parametric test).

2.4. Pose Classification

The Support Vector Machine (SVM) is a supervised discriminative classifier defined by a separating hyperplane and finding themaximum-margin hyperplane from a given data set. Multiple improvements on the traditional SVM wereproposed specially to classify non-linear data; among which the kernel SVM (KSVM) is the most effective [14]. The extended SVM algorithm allows us to fit the maximum-margin hyperplane in a transformed feature space. Four KSVM classifiers with various selected kernel functions: linear (lin-KSVM), quadratic (quad-KSVM), cubic (cub-KSVM) and Radial Basis Function (RBF-KSVM) were considered to test the effectiveness of the quadrilateral features in differentiating the human poses and classifying in to three different groups of human posture. These kernels can be attained by the following models:

$$\text{lin-KSVM:} K(x_m, x_n) = x_m^T x_n \tag{1}$$

$$\text{quad/cub-KSVM:} K(x_m, x_n) = (x_m^T x_n + c)^d \tag{2}$$

$$\text{RBF-KSVM:} K(x_m, x_n) = exp\left(-\frac{\|x_m - x_n\|}{2\sigma^2}\right) \tag{3}$$

where: K=Kernel function
σ =Scaling factor
x_m, x_n=Vectors in the input space
d=Degree of polynomial (quadratic: *d*=2; cubic: *d*=3)
C=Soft margin constant

2.4. Falling Event Detection

As explained in previous section, each feature vectors sets in A1, A2 and A3 represent different human postures. The PRS output may represents a simple event and the sequence of simple events may composes a complex event, such as falling event. Therefore, the model of FED are characterised as a FSM as shown in Figure 3, where it consists of three event states: Normal Event 1; NE(1), Normal Event 2; NE(2) and Falling Event; FE. The current state depends on the past states of the system and the transition takes place based on the outputs provided by the PRS model.

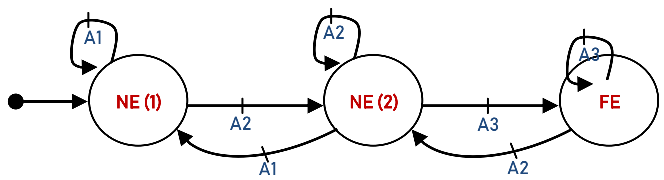


Figure 3. A 3-state machine for falling event detection

The FSM for detecting falls wastested on human activities video set provided by Milegroup based at the University of Vigo in Spain. The dataset consists of 224 videos of seven actions and it was clustered in to two groups; namely normal and falling events as shown in Table 1. Each action was performed for several times by eight subjects of different physiques and gender. The lateral movement actions with cleanblack background were captured by using a single stationary camera as shown in Figure 4.

Table 1. Normal and Falling Events Video Sets

Event Group	Action	Number of video set
Normal Event (NE)	1. Normal walking	40
	2. Exaggerated walking	40
	3. Jogging	40
	4. Bending over	32
	5. Sitting on the chair	40
Falling Event (FE)	6. Falling	16
	7. Lying down	16

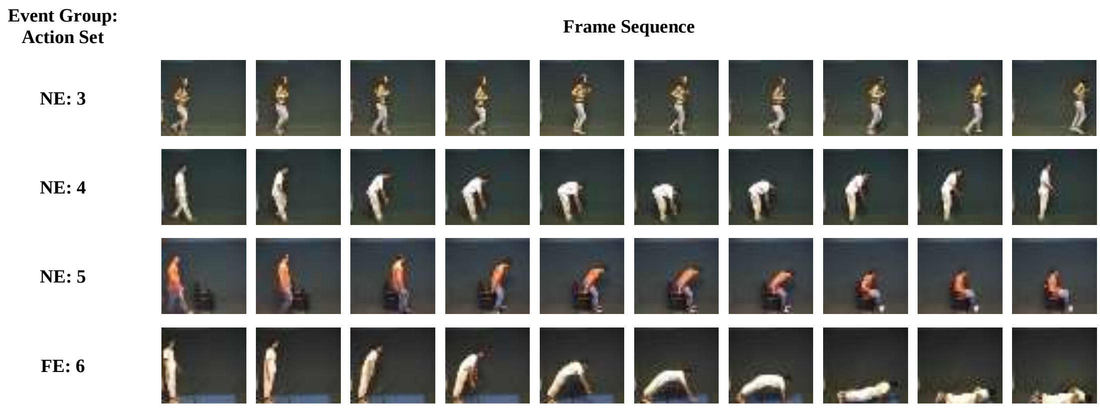


Figure 4. Example of human actions in video frame sequence in MILE dataset

2.5. Performance Evaluation

The classification performance assessment is based on the correct and incorrect predictions numbers for each class, which is encoded in a confusion matrix table.From the matrix table, several global estimation measurements of binary and multi-classification performances can be derived as proposed in [15]. To get a sense of effectiveness on this small multiple classes, two performance measures: macro-averaging sensitivity a.k.a. recall ($sens_M$), macro-averaging precision ($prec_M$) and macro-averaging F-Score ($Fscore_M$) were considered to estimate the quality of overall classification performance [16]. The evaluation of $sens_M$ focusing on average per-class effectiveness of a classifier to identify class labels and it may formulated as:

$$sens_M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l} \tag{4}$$

Where by tp_i, fp_i, fn_i and tn_i are true positive, false positive, false negative and true negative for l classes counts, respectively. While the $prec_M$ evaluation focusing on average per-class agreement of the data class labels with those of a classifier and $prec_M$ calculated as:

$$prec_M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l} \tag{5}$$

Futher evaluation of classifer accuracy is measured by observing the relations between data's positive labels and those given by a classifier based on a per-class average and the harmonic mean of precision and recall; $Fscore_M$ formulate as:

$$Fscore_M = 2 \frac{sens_M . prec_M}{sens_M + prec_M} \tag{6}$$

The best performance of classifier will be chosen as PRS model. Subsequently, the output (simple event) of recognition system will be fed into FSM to detect complex events; falling events. Finally, the performance of FED is measured to determine the extent of their effectiveness of the system in detecting the presence of falling event.

3. RESULTS AND ANALYSIS

All tasks were done in MATLAB® R2017a and Statistical Package for the Social Science (SPSS) V22 software, which are embedded in a notebook computer: Intel i7 processor, running Windows 10 OS, with 16GB of RAM. The numbers of A1, A2 and A3 samplesused for training and cross validation are 10,000, 6910 and 10,000, respectively. All extracted feature vectors were normalised before it was statistically analysed and used as training and validation datasets.

3.1. Statistical Analysis

Two thousand samples were randomly selected from the total of sample number in each group by using free online random sampling software; namely *Research Randomizer*. It eliminates the source of bias in samples sets and permits the use of appropriate probability theory to express the likelihood of chance as a source for the difference of end outcome [17]. The SW and LV tests were conducted for assessing normality and homogeneity of variances, respectively. The SW test summarised that all probability result, *p*-value to correspond features were less than 0.001 (N=2,000). Thus, the test rejected the hypothesis of normality for all features due to the conducted test resulting *p*-value is less than 0.05 (data significantly deviates from a normal distribution with 95% confidence level). While the LV test summarised the variances over all posture groups for each feature were not equal. These tests resulting violation of normality and homogeity of variance assumptionsof the parametric test, ANOVA. Therefore, the non-parametric KW test was chosen to determine if there are statistically significant differences between the three independent groups.

The statistical relevance of theresults have been verified by means ofKW test, which does not assume gaussianity in the dataunder study. The selected test analysed all corresponding features extracted from the generalised quadrilateral shape of human posture. The test shows all probabilities values, *p*for corresponding feature were below 0.001; rejecting the null hypotheses for all features (α<0.05). Thus, the mean rank between the groups for all 12 features were statistically associated and were significantly different median latencies in A1, A2 and A3 (N=2,000). This concludes that non-significant features were discarded and the 12 features will be utilised as the attributes for classification process.

3.2. Classification

The *k*-fold cross validation was applied on each classifiers in which the datasets were randomly divided into *k* approximately equal size subsets (i.e *k*=10). Each training and validation sets were comprised of *k*-1 subsets and the remaining subset, respectively. This procedure was repeated *k* times and single estimation of the whole dataset was calculated from the combination of *k*-fold result. The performances of each classifier from 12 features are summarised in Table 2.

Table 2. The performance of KSVM classifiers

Performance	Classifier			
	lin-KSVM	quad-KSVM	cub-KSVM	RBF-KSVM
$sens_M$	96.71%	98.31%	98.78%	99.19%
$prec_M$	96.79%	98.31%	98.86%	99.25%
$Fscore_M$	96.75%	98.31%	98.82%	99.22%

Table 2 presents the performance evaluation: macro-averaging sensitivity and precision and F-score of four selected classification models. In general, all KSVM models performed very well (>96%) in term of mean sensitivity and precision rates. The minimum and maximum $sens_M$ rates were 96.71% (lin-KSVM) and 99.19% (RBF-KSVM), respectively. Where by the minimum and maximum $prec_M$ rates were 96.79% (lin-KSVM) and 99.25% (RBF-KSVM), respectively. While the harmonic means of $sens_M$ and $prec_M$ for four classifiers were 96.75%, 98.31%, 98.82% and 99.22%, respectively. Where the RBF-KSVM model out performed other type of SVMs' kernels models. Globally, we observed that all performances were proportionally increased to the kernel complexity level (linear, quadratic: polynomial of degree-2, cubic: polynomial of degree-3 and Gaussian). As a result, the highest performance classifier; RBF-KSVM was chosen as the model of PRS.

Figure 4(a) and Figure 4(b) show the detail of RBF-KSVM's precision and sensitivity performance for each class, respectively. From these matrix tables, we observed that the positive prediction rate and true positive rate for all classes is higher (>98%). It means the model was able to identify >98% correctness classes with prediction probability rate>98% for each class. In addition, the model was incorrectly labelled A2 for the majority of the mislabelled cases. This is due to some of the human postures in group A2 is almost the same with postures in A1 and A3; precisely the pose during action changes transition, such asbending-standing and crawling-lying down; vice versa. Consequently, this minor deficiency is expected to affect the performance of FED.

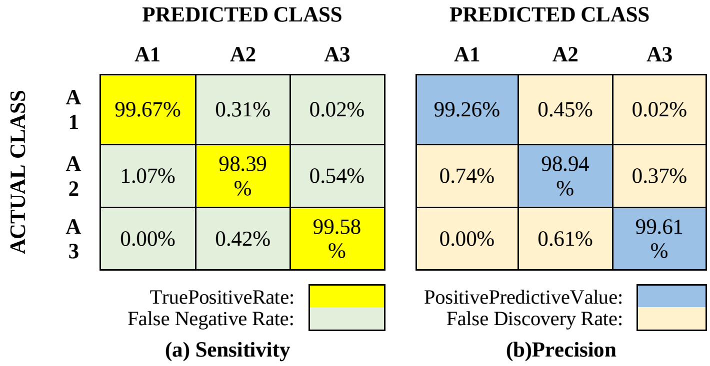


Figure 4. The RBF-KSVM performance

3.3. Fall Detection Performance

Our FED algorithm was evaluated on 224 videos from MILE dataset; comprising two groups of events (NE and FE). Table 3 tabulates results of the proposed algorithm against results ofthe state-of-the-artGaussH-BFFNN-PD fall detection algorithm in [2].

Table 3. The performance of FEDs

FED model	Accuracy	Error	Sensitivity	Specificity	F-score	CT
GaussH-BFFNN-PD[2]	99.30%	0.7%	98.47%	98.50%	-	-
Proposed method	97.32%	2.68%	98.95%	88.24%	94.70%	198.24 ms

Surprisingly, the proposed method was able to detect the normal and falling states with only six misclassified among 224 predictions (97.32%) with error rate of 2.68%. Specifically, two out of 32 fall detection tasks were wrongly predicted, and four FPs were detected out of 192 normal events. Whereby, the sensitivity and specificity rates were about 98.95% and 88.24%, respectively. Whereas, the macro-averaging F-score is about 94.70%. These classification performances impliedthat the overall measure of exactness or quality, completeness or quantity and the classifer accuracy from the fall detector were high. The overall performance of the proposed method was slightly low compared with [2]; however, both models were considered performing well in detecting the binary events with accuracy, sensitivity and specificity greater than 88%. Besides, the proposed algorithm computional time (CT) for each prediction process is quite fast; approximately 198.24 ms only. This simple feature extraction process gives an advantage on time execution compared to [2] which is higher due to the complexity of the motion-based features extraction process.

4. CONCLUSION

We have proposed a PRS based on quadrilateral shape features of silhouette. The KW test was conducted to asses all corresponding 12 feature vectors between three groups of human poses. Statistically, all proposed features were significantly different (significance level of p<0.05). In detecting and classifying the human poses into three posture groups, the RBF-KSVM classifier outperformed the other type of SVMs' kernels, namely, lin-KSVM, quad-KSVM and cub-KSVM with $sens_M$ =99.19%, $prec_M$ =99.25% and $Fscore_M$ =99.22%. Overall, all KSVMs performed very well with performance rates above 96%. Such pose classification model output was further used in the FSM to trigger the falling event alarms. The FSM model performed well (with accuracy of 97.32%) in detecting the presence of falling events in a frame sequence of videosand involved lowcomputional time. Nevertheless, we are keen to assess our proposed falldetection model on other online falling event databases which consists of dynamic angle movement towards real-time application; particularly in a surveillance system.

ACKNOWLEDGEMENT

The authors acknowledge the financial support from the following: MOSTI (01-01-02-SF1386) and DIP-2015-012.

REFERENCES

[1] Ranasinghe S, Al Machot F, Mayr HC. A review on applications of activity recognition systems with regard to performance and evaluation. *Int J Distrib Sens Networks.* 2016 Aug 24;12(8): 155014771666552.
[2] Suriani NS. Sudden Fall Event Recognition using Motion Features and Biological Inspired Classifier of Visual System. PhD Thesis. Universiti Kebangsaan Malaysia; 2015.
[3] World Health Organization. *The health of older people in selected countries of the Western Pacific Region*. Ageing and Health. 2014.
[4] Phelan EA, Mahoney JE, Voit JC, Stevens JA. Assessment and Management of Fall Risk in Primary Care Settings. *Med Clin North Am*. 2015 Mar; 99(2): 281–93.
[5] Abu Hassan MF, Hussain A, Md Saad MH, Win K. 3D distance measurement accuracy on low-cost stereo camera. *Sci Int*. 2017; 29(3): 599–605.
[6] Zheng S, Zhang J, Huang K, He R, Tan T. *Robust view transformation model for gait recognition*. Proceedings of the 18th IEEE International Conference on Image Processing. IEEE; 2011. p. 2073–6.
[7] Charfi I, Miteran J, Dubois J, Atri M, Tourki R. Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and Adaboost-based classification. *Journal of Electronic Imaging*. 2013; 22(4): 041106.
[8] Hdioud B, El M, Tirari H, Oulad R, Thami H, Faizi R. Detecting and Shadows in the HSV Color Space Using Dynamic Thresholds. *Bulletin of Electrical Engineering and Informatics (BEEI)*. 2018; 7(1): 70–9.
[9] Arias-Castro E, Donoho DL. Does median filtering truly preserve edges better than linear filtering? *Ann Stat*. 2009 Jun; 37(3): 1172–206.
[10] Yudhana A, Sunardi, Saifullah S. Segmentation comparing eggs watermarking image and original image. *Bulletin of Electrical Engineering and Informatics (BEEI)*. 2017; 6(1): 47–53.
[11] Zhu P, Xu Q, Hu Q, Zhang C, Zhao H. Multi-label feature selection with missing labels. *Pattern Recognition*. 2018; 74: 488-502.
[12] Ghasemi A, Zahediasl S. Normality tests for statistical analysis: A guide for non-statisticians. *Int J Endocrinol Metab*. 2012; 10(2): 486–9.
[13] Cristianini N, Schölkopf B. Support Vector Machines and Kernel Methods: The New Generation of Learning Machines. *Artif Intell Mag*. 2002; 23(3): 31–42.
[14] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management.* 2009; 45(4): 427-437.
[15] Manning CD, Raghavan P, Schutze H. Introduction to Information Retrieval. Cambridge University Press. 2009. 3-3 p.
[16] Suresh K. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *J Hum Reprod Sci*. 2011; 4(1): 8.
[17] Lájer K. Statistical tests as inappropriate tools for data analysis performed on non-random samples of plant communities. *Folia Geobot*. 2007 Jun; 42(2): 115–22.