



## A New Effective Speed and Distance Feature Descriptor Based on Optical Flow Approach in HAR

Hua Guang Hui<sup>1\*</sup>, Hemantha Kumar<sup>1</sup>, Manjunath Aradhya<sup>2</sup>, M.S. Maheshan<sup>3</sup>

<sup>1</sup> Department of Studies in Computer Science, University of Mysore, Mysuru 570006, Karnataka, India

<sup>2</sup> Department of Computer Applications, JSS Science and Technology University, Mysuru 570006, Karnataka, India

<sup>3</sup> Department of Information science and engineering, JSS Science and Technology University, Mysuru 570006, Karnataka, India

Corresponding Author Email: [huaguanghui@compsci.uni-mysore.ac.in](mailto:huaguanghui@compsci.uni-mysore.ac.in)

<https://doi.org/10.18280/ria.370114>

### ABSTRACT

**Received:** 23 November 2022

**Accepted:** 7 January 2023

#### Keywords:

human activity recognition, video surveillance, GMM, optical flow, multi-class SVM, feature descriptor

Nowadays, artificial intelligence and computer vision have been applied in various domains in the real world, such as Autonomous vehicles, video surveillance, human activity recognition, face recognition, smart home, and automated industry. Video-based human activity recognition is a big challenge yet. This paper proposes the Gaussian Mixture Model and Optical Flow approach to detect foreground and feature extraction for human activity recognition. The speed with a range of frames and radial distance from the Centroid to edge points of the human silhouette describe the feature vector of human activities. And then, the features are classified by multi-class SVM. The proposed system has been tested in the Weizmann datasets and KTH datasets. The experiment shows that our methods distinguish walking, bending, running, and wave-hand efficiently and accurately.

## 1. INTRODUCTION

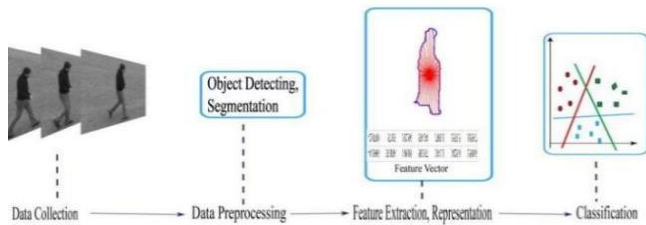
The main aim of video-based human activity recognition (HAR) is automatically classified human activity. It is for human activity classification based on computer vision and machine learning (ML). Especially in Smart homes, artificial remote monitoring, human-machine interaction, and publicly area abnormal activity detection have a huge significance in application and research. In recent years, human activities based on complexity are categorized into gestures, actions, human-human or human-object interaction, and group actions [1]. Gestures are considered an element of action from parts of the human body. Actions are movements to describe a motion or motivation of a person. Human-human or human-object interactions are activities that include two or more two persons or objects.

The traditional system of HAR is divided into data collection (Raw Data), pre-processing, feature extraction, representation and classification. The input data has two types of ways vision-based and sensors. Binary, Gray, and RGB-D are the vision-based raw data. Smartwatches are popular sensor devices. After data collection, the input data may be improper to access, and then the necessary operation is the data pre-processing, such as foreground detecting (background model, frame subtraction, Gaussian mixture model, and optical flow etc.), segmentation, re-sizing and smooth and so on. After data pre-processing, we can design a proper model that features extraction and representation to represent human activity by a handcrafted or trainable feature extractor. such as Space-time based approach (space-time volumes, space-time trajectories, and space-time features), appearance-based approaches (shape-based, motion-based, hybrid) and others approaches like fuzzy logic. Aradhya et al. [1] combine two types of techniques wavelet transform and Gabor filter to

extract sharpened edges and textural features for an image. Finally, the features vector is put into the classifier to recognize different activities to which the class belongs. Finally, the features vector is put into the classifier to recognize different activities to which the class belongs.

An intelligent system of HAR is more efficient and helpful for application in the real world than an artificial one. In HAR, the traditional and deep learning approaches are two main branches. In the past decade, most researchers staying focused on traditional methods. Feature representation methods are divided into global representation (Space-time volume, Frequency), local representation (Depth maps based, skeleton-based) and body modelling (Simple Blob, 2D/3D model) [2]. Three types of datasets single, multi-view and RGB-D and the state-of-the-art deep learning techniques were reviewed and discussed [3]. Due to high-performance computer development, there has been an increasing amount of literature on Deep Learning. Such as Convolutional Neural networks (CNN), Generalized Regression Neural Networks (GRNN), Probabilistic Neural network (PNN), Two-Stream networks, and some hybrid networks are applied widely in image or sequence images (video) detection, segmentation, feature extraction, representation and classification. Aradhya et al. [4] proposed a model based on an ensemble of Generalized Regression Neural Networks (GRNN) and Probabilistic Neural network (PNN) classifiers at the decision level. Xu et al. [5] using a CNN extracted six kinds of features to recognize physical activity. The result shows the CNN got a good accuracy, which outperforms the traditional method SVM from 3D datasets. Aradhya et al. [6] apply the Probabilistic Neural network for unconstrained handwritten characters for the south India language of Kannada on image datasets. In this paper, a traditional approach based on Optical Flow and Gaussian Mixture Model (GMM) to detect foreground and

silhouette information on a human body. Then the multi-class SVM as the classifier to identify each class of activities. Figure 1 shows the structure of the human activity recognition system. The structure of this paper is organized as follows: Section 2 gives the recent related work for human activity recognition. Section 3 is the proposed method in the HAR system. In section 4, the relevant datasets and the experiment result are displayed. Moreover, we also analyze the experiment result based on the optical flow approach to Weizmann and KTH datasets. Finally, section 5 gives the conclusion.



**Figure 1.** The architecture of the HAR system

## 2. RELATED WORK

It is a hot top that video-based human activity recognition systems with deep learning in recent years. The traditional approach that bases global or local feature research still has great significance in human activity recognition. Most of all, the conventional method to feature vector dimension, training sample size, amount of computation, and equipment demand is less than the deep learning approach.

Vrigkas et al. [7] provide a detailed review of the state of research advantages and limitations, human activity approaches, and existing benchmark datasets in human activity recognition. Wu et al. [3] introduced the different types of datasets in which single, multi-viewpoint, and RGB/RGB-D video datasets are in the human activity recognition system. Meanwhile, a variety of applications and challenges are introduced too. Sharma et al. [8] aim to present a comparative review of vision-based human activity recognition on a recent benchmark dataset and deep learning technique. Tripathi et al. [9] gave details about visual surveillance-based human activity recognition applications, issues, and challenges in the suspicious activity recognition system. Zhang et al. [10] highlight the advantage of the activity recognition method. Especially for the human activity (global/local and depth-based) representation and classification methods. Recently, with the development of deep learning techniques, deep learning algorithms have solved so many traditional problems with better results. At moment, the traditional approaches also have a big significance in the research domain. Sargano et al. [11], divided the representation into body (simple Blob, 2D/3D model), global (space-time volume (STV), frequency) and local (Depth map, skeleton joints) feature representation modelling. Compared traditional approach and deep learning method. Meanwhile discussed the recently open challenges and publicly available 2D/3D datasets in human activity recognition systems. Dhiman et al. [12] summarize the existing variation of the type of information (2D/3D, RGB-D and skeletal data) in abnormal activity recognition. Dang et al. [13] from the generated data type, classifying the raw data into two main groups, which are vision-based and sensor-based human activity recognition. Moreover, discussing the

challenges in the current human activity recognition topic. Horn and Schunck [14] provide a method for finding the optical flow pattern. Stauffer et al. [15-17] present a robust and efficient framework of the Gaussian Mixture Model (GMM) for detecting objects based on a background subtraction algorithm without extra computation cost. Boufama et al. [18] use a traditional simple shape descriptor in which trajectories as mid-level features and the standard bag of word algorithms for human activity recognition. Singh et al. [19] combine contour-based distance signal feature, optical flow-based motion feature and uniform rotation local binary patterns for Hidden Markov models (HMMs) to classify the view-invariant human activity. Dong et al. [20] proposed foreground trajectory and motion different descriptors for support vector machine (SVM) classifier to classify the action of UCF Sport and YouTube datasets. Blob detection and the Kalman filter object tracking algorithm by frames [21], Bi-dimensional Empirical Model Decomposition (BEMD), Scale Invariant feature transform (SIFT) and Discrete Wavelet Transform (DTW) applied on the pre-processing to extract efficient features for Convolution Neural Network classifier. Jalal et al. [22] proposed a depth video-based method to extract and represent robust multi-features for the Hidden Markov Model (HMMs) to recognize the daily activity of elder people. Hbali et al. [23] proposed a skeleton-based approach to describe the Spatial-temporal information of human activity. To use the Extremely Randomised Trees algorithm to train and validate the system on MSR 3D and MSR Daily Activity 3D datasets. Ehatisham et al. [24] utilize data from multi-sensors, which include RGB, depth map and wearable inertial sensors. To extract efficient features for training and testing the fusion model with K-nearest neighbour and SVM method.

In recent years, united traditional and deep learning approaches show that human activity recognition accuracy is better than the single method. Basly et al. [25] combined a recent CNN method of deep learning to extract the relevant features and fed them into the traditional classifier SVM. The hybrid model in which traditional and deep learning approaches is a hot technique than one model only. The result of the proposed technique proves that it has performed better. Yadav et al. [26] proposed a skeletal coordinate feature from the image/video frame into convolutional long short-term memory (ConvLSTM) network for activity recognition and fall detection. The experiment result shows that the deep learning techniques already got good accuracy and solved many challenges in the vision-based human activity recognition domain. The traditional approach still has superiority in feature dimension, computing speed, Interpretability and hardware device demand aspects.

## 3. METHODOLOGY AND IMPLEMENTATION

### 3.1 The structure of the HAR system

The main challenges of vision-based human activity recognition are intra-class variability, inter-class similarity, illumination changes, occlusion, shadow effect, and noise image/jitter image. Intra-class variability and Inter-class similarity especially are issues in human activities recognition, so it requires us to find an effective descriptor to represent the feature of human activities. In the HAR system, feature extraction and representation are the best signification steps. The fit information of features will present the essence of an

activity pattern. In traditional approaches, the global or local features from object context, silhouette, coordinate and geometry information are extracted to classify human activity. The deep learning approach that convolution neural networks are a very good method to extract features map from convolution networks.

In this paper, The Gaussian Mixture Model (GMM) Stauffer et al. [15-17], was used to detect, and segment foreground object information from background information of the video sequences. The optical flow techniques [14] extracted relevant features through boundaries for recognizing different human actions. Wang et al. [27-29] gives more details about the Support Vector Machines (SVM) and applications in different domains. Figure 2 displays the architecture of our proposed HAR system.

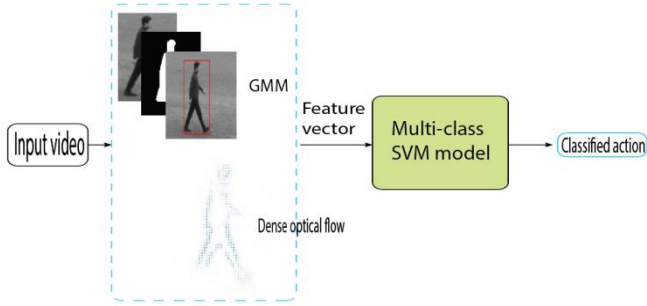


Figure 2. The HAR architecture of the proposed method

### 3.2 Preprocessing and feature extraction

In the human activity recognition system, raw data collection, Data pre-processing, feature extraction, representation and classification are primary steps. The video-based human activity recognition poses a lot of challenges. Nowadays, the main problem is to characterize the kinds of features of human activities for identification. The benchmark available datasets as raw data employ the GMM approach to detect the foreground object from the video sequences. The GMM technique was a typical probability model for detecting foreground objects from the background through learning update Mean, Variance, and weights. Compare each input pixel with the Mean of associated elements to match whether the value is closed enough to the chosen Mean. The difference between the pixels and Mean must be less than the element's standard deviation (Variance) considered as the matched element. Then update the Gaussian weight, Mean and Variance to demonstrate the obtained pixel value. For non-matched elements, the Gaussian weight decreases whereas the Mean and Variance remain constant. Finally, identify the foreground pixels which are not matched with any element determined as a background. A Gaussian mixed model can be formulated as below:

$$P(x_t) = \sum_{i=1}^k W_{i,t} \cdot \eta(x_t, u_{i,t}, \Sigma_{i,t}) \quad (1)$$

where obviously:

$$\sum_{i=1}^k (W_{i,t}) = 1 \quad (2)$$

The mean of such a mixture equals:

$$u_t = \sum_{i=1}^k W_{i,t} u_{i,t} \quad (3)$$

where,  $i$  is the current pixel in the frame,  $k$  is the number of distribution of Gaussian,  $t$  represents time (the frame index),  $w$  is an estimate of the weight of the  $i_{th}$  Gaussian in the mixture at time  $t$ ,  $u$  is the Mean value of the  $i_{th}$  Gaussian in the mixture at time  $t$ ,  $\Sigma_{i,t}$  is the covariance matrix of the  $i_{th}$  Gaussian in the mixture at time  $t$ .

Compared with the Frame difference technique, the Gaussian Mixed model is a more robust foreground detection technique in a video sequence with a little vibrated background. Based on the GMM to detect the foreground body box. For a body square box, we have two coordinate values (left top and right bottom),  $[(x_1, y_1), (x_2, y_2)]$ . The centroid point coordinate is  $(X_0, Y_0)$ .

$$X_0 = (x_1 + x_2) / 2 \quad Y_0 = (y_1 + y_2) / 2$$

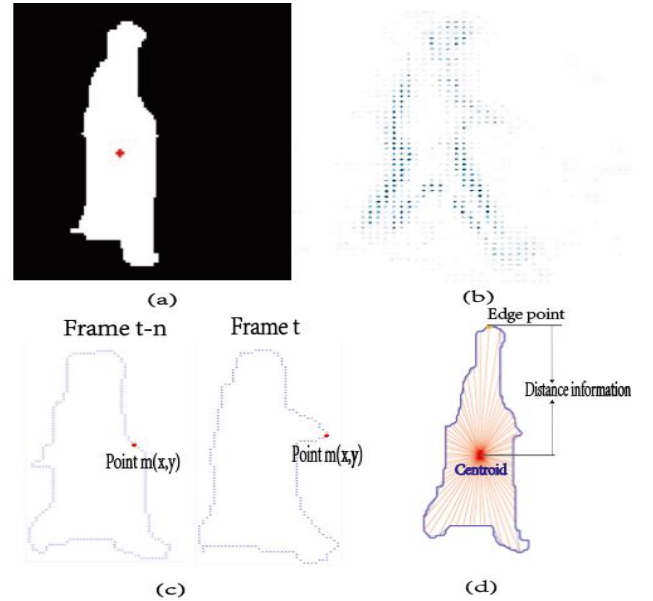


Figure 3. (a) Foreground detection and calculating the Centroid point on human activity silhouette. (b) Dense Optical flow features on a frame. (c) The speed information on frames difference of human activity. (d) The distance information on a frame from the Centroid to the edge point.

About the optical flow, it is the moving distance of the pixels value from one frame to next frame with the same object in a video sequence. The dense optical flow method calculates the silhouette points around the human activity silhouette. Around all silhouette points, we select one unique coordinate point by every 15 degrees,  $24(24 \cdot 15 = 360)$  degrees) paired values  $(x, y)$ . The GMM help us to extract precise foreground object with the centroid. The optical flow gives the silhouette coordinate value. The feature extraction process in which the human foreground object detect with the centroid shows in Figure 3(a). For the speed information, the current frame is  $72 \cdot 2(x, y)$ , 72 coordinate points (5 is divided by 360 degrees). The points of the moving object that 72 coordinates values were calculated with the current frame-N (here  $N=3$ ) frame. Figure 3(b) displays the optical flow points on the image frame. The 24-Distance features are extracted by 15 degrees (0, 15,

30, 45, 60, 75, 90, ..., 360) from the Centroid to the edge points of the silhouette. The speed and distance feature vector is described as  $\{Sx1, Sy1, \dots, Sx72, Sy72, D1, D2, \dots, D24\}$ . Figure 3(c) displays the speed information on the neighbour frames of the human silhouette. The distance feature from the Centroid to the edge points represent in Figure 3(d). The difference values of 72 coordinate points and 24 radial distance features are ordered into a row vector  $(1*(72*2)+24)$  to represent the optical flow-based feature vector. The speed information discriminates intra-class activities such as running, walking, jump, and the distance information discriminates the similarity. The hybrid speed and radial distance feature vector are put into the multi-class SVM classification model to train and test human activities.

### 3.3 Multi-class SVM

The SVM was originally a binary classification method. The main idea is to construct a hyperplane to separate two classes (Labeled 1, -1).

In this section, we discussed the classic multi-class SVM strategies, namely one against one (OAO) and one against all (OAA) Hsu et al. [28, 29]. Assume there is a three-class problem:

**OAO:** Here, three binary classifiers are trained, and each classifier is trained from a pair of classes. After training, the unknown sample is classified into the class with the largest number of votes.

In **OAA**, each class is trained against the remaining two categories. The  $i_{th}$  SVM is trained with all of the examples in the  $i_{th}$  class with positive labels, and all other examples with negative labels. After training, three decision functions are used to determine the class of an unknown sample  $x$  as follows:

$$class\ of\ x \cong \arg \max_{j=1,2,3} \left( \sum_{i=1}^n a^i_j y_i k(x, x_i) + b^j \right) \quad (4)$$

## 4. EXPERIMENTAL RESULT AND ANALYSIS

### 4.1 Human activity recognition datasets

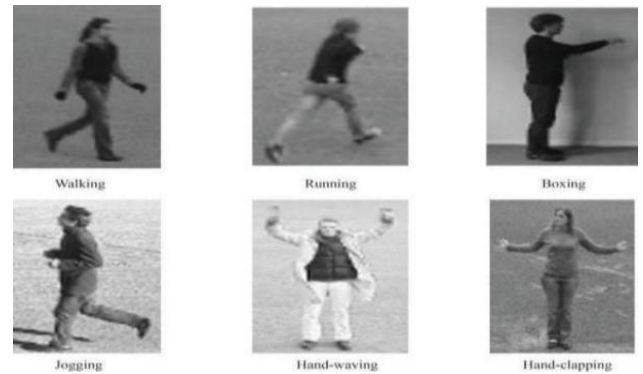


**Figure 4.** The sample images of the Weizmann datasets

The Weizmann datasets Gorelick et al. [30] are easily available. It was captured with a stationary camera and uncomplicated background in 2005. It consists of 90 video sequences (low-resolution  $180 * 144$ , deinterlaced 50 fps). It involves a total of 5701 frames, ten natural actions by nine different people performance. The actions are bend, jack, jump, p-jump (jump-in-place-on-two-legs), run, side (gallop

sideways), skip, wave1 and wave2. The datasets only contain a single action in each clip. This is one of the established benchmarks of the datasets for human activity recognition. The sample frame of the datasets is shown in Figure 4.

The KTH datasets Schuldt et al. [31] are more challenging than the Weizmann datasets. The current video sequences datasets contains 2391 sequences with a static camera (25fps, resolution of  $160*120$  pixels), including six types of human activity such as walking, jogging, running, boxing, hand waving and hand clapping. it is performed several times by 25 subjects in four different scenarios (outdoors, scale variation, outdoors with different clothes and illustrated change.) The sample images of the datasets are depicted in Figure 5.



**Figure 5.** The sample images of the KTH datasets

### 4.2 Experiment result and comparison analysis

**Table 1.** The results of the proposed method on Weizmann datasets

Method	Human action	Classification rate
Proposed Method	Walk	100%
	Bend	86.9%
	Jack	96%
	P jump	100%
	Run	91%
	Side	100%
	Wave1	92%
	Wave2	92%
	<b>Average</b>	<b>94.7%</b>

The proposed method is validated on the Weizmann and KTH datasets. The leave-one-out (LOO) approach has been used to test the hybrid features vector descriptor whose speed and radial distance base the optical flow method. Table 1 and Table 2 give the results of the proposed method on the Weizmann and KTH datasets. The average result is 94.7% on the Weizmann datasets. The accuracy of the walk, bend, jack, p-jump, run, side, wave1 and wave2 are 96.4%, 86.9%, 96%, 100%, 91%, 100%, 92% and 92%. The average accuracy is 87.2% on the KTH datasets. The result of each activity is 98.2% for walking, 84% for boxing, 86% for jogging, 81% for running and 93% for hand-waving. The proposed method distinguishes the activities well on the Weizmann datasets. The KTH datasets are more challenging as compared to the Weizmann datasets. The experiment result of the KTH datasets shows that it is still not ideal due to the different light conditions, uniform background, human activity shadow and camera jitters. In object detection, The GMM is one of the background model approaches. Sometimes the method detects the wrong foreground object on a frame. So it is not a

satisfactory method for complex conditions. Furthermore, the proposed method also has advantages in feature dimension and complements efficiency compared to existing work in Table 3. Table 4 and Table 5 give an accurate comparison of the Human activity recognition system on the Weizmann and KTH datasets.

**Table 2.** The results of the proposed method on KTH datasets

Method	Human action	Classification rate
Proposed Method	Walking	92%
	Boxing	84%
	Jogging	86%
	Running	81%
	Hand waving	93%
	<b>Average</b>	<b>87.2%</b>

**Table 3.** A comparison of feature dimension in the optical flow method

Method	Features dimension
Optical flow [32]	72*9
Optical Flow [33]	120*120
Optical Flow+Action Flow Net [34]	512
<b>Proposed method</b>	<b>72*2+24</b>

**Table 4.** Accuracy comparison of HAR system on the Weizmann datasets

Method	Activity accuracy
Hong and Kim [35]	76.6%
Dorin and Hurwitz [36]	81.9%
Gomathi and Santhanam [37]	82.0%
Nadeem et al. [38]	89.4%
<b>Proposed method</b>	<b>94.7%</b>

**Table 5.** Accuracy comparison of HAR system on KTH datasets

Method	Activity accuracy
Hong and Kim [35]	80.1%
Dorin and Hurwitz [36]	79.2%
Gomathi and Santhanam [37]	83.0%
Nadeem et al. [38]	86.67%
<b>Proposed method</b>	<b>87.2%</b>

## 5. CONCLUSION

In this paper, we have proposed a novel feature descriptor based on the optical flow approach. The speed and radial distance feature vector based on GMM and dense optical flow were used to classify activity by multi-class SVM. The proposed method has considered feature dimension and accuracy in human activity recognition. In especial, the proposed system distinguishes between the intra-class variability and inter-class similarity well based on speed information and radial distance from the centrefold to the edge points of the human silhouette. The hybrid feature descriptor of speed and radial distance was applied to the Weizmann and KTH benchmark datasets. The experiment result shows that our proposed method achieves better results for classifying different human activities when compared to other existing approaches. In future work, we wish to expand the proposed

method to more benchmark datasets of human activity recognition and also explore characterization and description approaches. Moreover, we plan to use object detection based on deep learning techniques to extract a precious foreground on each frame to improve the accuracy of the KTH datasets.

## REFERENCES

- [1] Aradhya, V.M., Pavithra, M.S., Naveena, C. (2012). A robust multilingual text detection approach based on transforms and wavelet entropy. *Procedia Technology*, 4: 232-237. <https://doi.org/10.1016/j.protcy.2012.05.035>
- [2] Boualia, S.N., Amara, N.E.B. (2019). Pose-based human activity recognition: a review. In 2019 15th international wireless communications mobile computing conference (IWCMC), Tangier, Morocco, pp. 1468-1475.
- [3] Wu, D., Sharma, N., Blumenstein, M. (2017). Recent advances in video-based human action recognition using deep learning: A review. In 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, pp. 2865-2872. <https://doi.org/10.1109/IJCNN.2017.7966210>
- [4] Aradhya, V.N., Mahmud, M., Guru, D.S., Agarwal, B., Kaiser, M.S. (2021). One-shot cluster-based approach for the detection of COVID-19 from chest X-ray images. *Cognitive Computation*, 13(4): 873-881. <https://doi.org/10.1007/s12559-020-09774-w>
- [5] Xu, W., Pang, Y., Yang, Y., Liu, Y. (2018). Human activity recognition based on convolutional neural network. In 2018 24th International Conference on Pattern Recognition (ICPR), pp. 165-170. <https://doi.org/10.1109/ICPR.2018.8545435>
- [6] Aradhya, V.M., Niranjan, S.K., Kumar, G.H. (2010). Probabilistic neural network based approach for handwritten character recognition. *Special issue of IJCTT*, 1(2): 3.
- [7] Vrigkas, M., Nikou, C., Kakadiaris, I.A. (2015). A review of human activity recognition methods. *Frontiers in Robotics and AI*, 2: 28. <https://doi.org/10.3389/frobt.2015.00028>
- [8] Sharma, V., Gupta, M., Pandey, A.K., Mishra, D., Kumar, A. (2022). A review of deep learning-based human activity recognition on benchmark video datasets. *Applied Artificial Intelligence*, 36(1): 2093705. <https://doi.org/10.1080/08839514.2022.2093705>
- [9] Tripathi, R.K., Jalal, A.S., Agrawal, S.C. (2018). Suspicious human activity recognition: A review. *Artificial Intelligence Review*, 50(2): 283-339. <https://doi.org/10.1007/s10462-017-9545-7>
- [10] Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., Li, Z. (2017). A review on human activity recognition using vision-based method. *Journal of Healthcare Engineering*, 2017: 3090343. <https://doi.org/10.1155/2017/3090343>
- [11] Sargano, A.B., Angelov, P., Habib, Z. (2017). A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition. *Applied Sciences*, 7(1): 110. <https://doi.org/10.3390/app7010110>
- [12] Dhiman, C., Vishwakarma, D.K. (2019). A review of state-of-the-art techniques for abnormal human activity recognition. *Engineering Applications of Artificial Intelligence*, 77: 21-45. <https://doi.org/10.1016/j.engappai.2018.08.014>

- [13] Dang, L.M., Min, K., Wang, H., Piran, M.J., Lee, C.H., Moon, H. (2020). Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition*, 108: 107561. <https://doi.org/10.1016/j.patcog.2020.107561>
- [14] Horn, B.K., Schunck, B.G. (1981). Determining optical flow. *Artificial Intelligence*, 17(1-3): 185-203. [https://doi.org/10.1016/0004-3702\(81\)90024-2](https://doi.org/10.1016/0004-3702(81)90024-2)
- [15] Stauffer, C., Grimson, W.E.L. (1999). Adaptive background mixture models for real-time tracking. In *Proceedings. In 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149)*, Fort Collins, CO, USA, pp. 246-252. <https://doi.org/10.1109/CVPR.1999.784637>
- [16] Zivkovic, Z. (2004). Improved adaptive Gaussian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*, Cambridge, UK, pp. 28-31. <https://doi.org/10.1109/ICPR.2004.1333992>
- [17] Tian, Y., Feris, R.S., Liu, H., Hampapur, A., Sun, M.T. (2010). Robust detection of abandoned and removed objects in complex surveillance videos. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(5): 565-576. <https://doi.org/10.1109/TSMCC.2010.2065803>
- [18] Boufama, B., Habashi, P., Ahmad, I.S. (2017). Trajectory-based human activity recognition from videos. In *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, Fez, Morocco, pp. 1-5. <https://doi.org/10.1109/ATSIP.2017.8075536>
- [19] Singh, R., Kushwaha, A.K.S., Srivastava, R. (2019). Multi-view recognition system for human activity based on multiple features for video surveillance system. *Multimedia Tools and Applications*, 78(12): 17165-17196. <https://doi.org/10.1007/s11042-018-7108-9>
- [20] Dong, S., Hu, D., Li, R., Ge, M. (2019). Human action recognition based on foreground trajectory and motion difference descriptors. *Applied Sciences*, 9(10): 2126. <https://doi.org/10.3390/app9102126>
- [21] Basavaiah, J., Patil, C., Patil, C. (2020). Robust feature extraction and classification based automated human action recognition system for multiple datasets. *International Journal of Intelligent Engineering and Systems*, 13(1): 13-24. <https://doi.org/10.22266/ijies2020.0229.02>
- [22] Jalal, A., Kamal, S., Kim, D. (2017). A depth video-based human detection and activity recognition using multi-features and embedded hidden Markov models for health care monitoring systems. *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(4): 54. <https://doi.org/10.9781/ijimai.2017.447>
- [23] Hbali, Y., Hbali, S., Ballihi, L., Sadgal, M. (2018). Skeleton-based human activity recognition for elderly monitoring systems. *IET Computer Vision*, 12(1): 16-26. <https://doi.org/10.1049/iet-cvi.2017.0062>
- [24] Ehatisham-Ul-Haq, M., Javed, A., Azam, M.A., Malik, H.M., Irtaza, A., Lee, I.H., Mahmood, M.T. (2019). Robust human activity recognition using multimodal feature-level fusion. *IEEE Access*, 7: 60736-60751. <https://doi.org/10.1109/ACCESS.2019.2913393>
- [25] Basly, H., Ouarda, W., Sayadi, F.E., Ouni, B., Alimi, A.M. (2020). CNN-SVM learning approach based human activity recognition. In: El Moataz, A., Mammas, D., Mansouri, A., Nouboud, F. (eds) *Image and Signal Processing. ICISP 2020. Lecture Notes in Computer Science*, vol 12119. Springer, Cham. [https://doi.org/10.1007/978-3-030-51935-3\\_29](https://doi.org/10.1007/978-3-030-51935-3_29)
- [26] Yadav, S.K., Tiwari, K., Pandey, H.M., Akbar, S.A. (2022). Skeleton-based human activity recognition using ConvLSTM and guided feature learning. *Soft Computing*, 26(2): 877-890. <https://doi.org/10.1007/s00500-021-06238-7>
- [27] Wang, Y., Li, Z., Feng, L., Bai, H., Wang, C. (2018). Hardware design of multiclass SVM classification for epilepsy and epileptic seizure detection. *IET Circuits, Devices Systems*, 12(1): 108-115. <https://doi.org/10.1049/iet-cds.2017.0216>
- [28] Hsu, C.W., Lin, C.J. (2002). A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, 13(2): 415-425. <https://doi.org/10.1109/72.991427>
- [29] Weston, J., Watkins, C. (1998). Multi-class support vector machines (pp. 98-04). Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May.
- [30] Gorelick, L., Blank, M., Shechtman, E., Irani, M., Basri, R. (2007). Actions as space-time shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12): 2247-2253. <https://doi.org/10.1109/TPAMI.2007.70711>
- [31] Schuldts, C., Laptev, I., Caputo, B. (2004, August). Recognizing human actions: A local SVM approach. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, Cambridge, UK, pp. 32-36. <https://doi.org/10.1109/ICPR.2004.1334462>
- [32] Kumar, S.S., John, M. (2016). Human activity recognition using optical flow based feature set. In *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*, Orlando, FL, USA, pp. 1-5. <https://doi.org/10.1109/CCST.2016.7815694>
- [33] Zhang, N., Hu, Z., Lee, S., Lee, E. (2017). Human action recognition based on global silhouette and local optical flow. In *International Symposium on Mechanical Engineering and Material Science (ISMEMS 2017)*, pp. 1-5. <https://doi.org/10.2991/ismems-17.2018.1>
- [34] Ng, J.Y.H., Choi, J., Neumann, J., Davis, L.S. (2018). Actionflownet: Learning motion representation for action recognition. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, USA, pp. 1616-1624. <https://doi.org/10.1109/WACV.2018.00179>
- [35] Hong, S., Kim, M. (2016). A framework for human body parts detection in RGB-D image. *Journal of Korea Multimedia Society*, 19(12): 1927-1935.
- [36] Dorin, C., Hurwitz, B. (2016). Automatic body part measurement of dressed humans using single RGB-D camera. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 3042-3048. <https://doi.org/10.1145/2851581.2892337>
- [37] Gomathi, S., Santhanam, T. (2018). Application of rectangular feature for detection of parts of human body. *Adv. Comput. Sci. Technol*, 11: 43-55.
- [38] Nadeem, A., Jalal, A., Kim, K. (2020). Human actions tracking and recognition based on body parts detection via Artificial neural network. In *2020 3<sup>rd</sup> International Conference on Advancements in Computational*

