Hindawi
Applied Bionics and Biomechanics
Volume 2022, Article ID 7931729, 14 pages
https://doi.org/10.1155/2022/7931729

Research Article

An Effective Approach for Human Activity Classification Using Feature Fusion and Machine Learning Methods

Muhammad Junaid Ibrahim,[1] Jaweria Kainat,[2] Hussain AlSalman,[3] Syed Sajid Ullah,[4] Suheer Al-Hadhrami,[5] and Saddam Hussain[6]

[1]*Department of Computer Science, University of Wah, 47040, Pakistan*
[2]*Department of Computer Science, COMSATS University Islamabad, Wah Cantt, Pakistan*
[3]*Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia*
[4]*Department of Information and Communication Technology, University of Agder (UiA), N-4898 Grimstad, Norway*
[5]*Computer Engineering Department, Engineering College, Hadhramout University, Hadhramout, Yemen*
[6]*School of Digital Science, Universiti Brunei Darussalam, Jalan Tungku Link, Gadong BE1410, Brunei Darussalam*

Correspondence should be addressed to Syed Sajid Ullah; syed.s.ullah@uia.no
and Suheer Al-Hadhrami; s.alhadhrami1@gmail.com

Received 30 November 2021; Revised 29 December 2021; Accepted 3 January 2022; Published 2 February 2022

Academic Editor: Fahd Abd Algalil

Recent advances in image processing and machine learning methods have greatly enhanced the ability of object classification from images and videos in different applications. Classification of human activities is one of the emerging research areas in the field of computer vision. It can be used in several applications including medical informatics, surveillance, human computer interaction, and task monitoring. In the medical and healthcare field, the classification of patients' activities is important for providing the required information to doctors and physicians for medication reactions and diagnosis. Nowadays, some research approaches to recognize human activity from videos and images have been proposed using machine learning (ML) and soft computational algorithms. However, advanced computer vision methods are still considered promising development directions for developing human activity classification approach from a sequence of video frames. This paper proposes an effective automated approach using feature fusion and ML methods. It consists of five steps, which are the preprocessing, feature extraction, feature selection, feature fusion, and classification steps. Two available public benchmark datasets are utilized to train, validate, and test ML classifiers of the developed approach. The experimental results of this research work show that the accuracies achieved are 99.5% and 99.9% on the first and second datasets, respectively. Compared with many existing related approaches, the proposed approach attained high performance results in terms of sensitivity, accuracy, precision, and specificity evaluation metric.

1. Introduction

In recent years, the e-vision community has focused largely on recognizing human activities. This is mainly because of a large number of industrial applications including human-computer interaction [1], antiterrorist applications [2], traffic surveillance [3], automotive safety [4], pedestrian detection [5], video surveillance [6], real-time tracking [7], rescue missions [8], and human-robot interaction [9]. This research work focuses on efficient recognition of human activity from recorded videos. Design of an efficient and optimal cost algorithm to detect a person from a video or an image is a challenging task. It is challenging in terms of variations of appearance, color, and movements [10]. Few other detection issues are also noticed like light and background variations [11]. Recently, numerous approaches have been proposed to detect a human from a video or an image. These approaches focused on the distinct use of classifiers, segmentation techniques, and feature extraction methods. Segmentation methods for human detection mainly contain

foreground detection [12] and template matching [13]. Existing approaches do not yield optimal results with several humans in an image or a scene. Furthermore, there are many techniques used to detect humans like the Histogram of Gradients (HOG) [14], Haar-like features [15], adaptive contour features (ACF) [16], Hybrid Wind Farm (HWF) [17], Image Source Method (ISM) [18], edge detection [19], and movement characteristics [20]. These extraction methods do not clearly show the mark when people are unclear or have significant fluctuations in their positions. However, selections of relevant characteristics significantly improve human activity recognition. This research implemented a hybrid approach to overcome accuracy challenge of human activity identification. This is done by enhancing the quality of frames extracted from videos and later categorizing the regions on the basis of specified feature vectors. The approach proposed in this paper comprises of five major stages including (a) normalization, (b) feature extraction, (c) feature selection, (d) feature fusion, and (e) classification. Normalization is a preprocessing stage in which several techniques like background subtraction, noise removal, and object extraction are implemented. Three types of features are extracted which are HOG, Gabor, and chromatic features. Principal Component Analysis (PCA) is separately implemented on three feature vectors to get optimal features. Later, the serial feature fusion is incorporated on the selected features. Lastly, five versatile classifiers are applied to evaluate better accuracy.

1.1. Major Contributions. Inefficient and lengthy preprocessing procedures decline the optimality of any algorithm. This work focuses on the efficient and accurate use of preprocessing and feature extraction steps. Thus, main contributions in this work include the following:

(i) Morphological operations are applied after background subtraction to get the exact region of interest

(ii) Separate principal component-based scoring for feature subset selection

(iii) Optimal results are obtained by the application of multiple classification techniques

The chronological order of this manuscript is as follows: Section 1 provides domain introduction, Section 2 describes past work related to the recognition of the human activities, Section 3 describes the proposed method, and in Section 4, results are compared with other existing techniques.

2. Related Work

So much work has been done and is ongoing in human activity recognition. All of the approaches proposed lie under two main categories: (a) the traditional handcrafted feature extraction methods [21] and (b) the automatic features (deep learning) [22] which employ automatic feature extraction methods. Some major existing works performed in human activity recognition are discussed as follows: An activity recognition system based on streaming data is presented by Yala et al. [23]. The proposed technique efficiently detects significant human activities. Nunes et al. [24] presented a framework for daily human action recognition. The proposed technique firstly extracts various features. Later, every human activity frame is encircled by two consecutive automatically recognized key positions, in which maximum static and dynamic characteristics are extracted. Kantorov and Laptev [25] discovered feature encoding by Fisher vectors and determined accurate action recognition utilizing linear classifiers. Liu et al. [26] presented a framework in which multiple features are fused to make action recognition better. The proposed approach captures the silhouette deformation of the performer after considering activities as 3D objects. Azary and Savakis [27] use sporadic demonstrations of spatial and temporal aggregate movements with abnormal size and location characteristics. Oreifej and Liu [28] defined the depth order incorporating histogram that records the physical dispersion of the surface in the 4 dimensions including spatial, coordinates, depth, and time. Conde et al. [29] introduced a human crawling technique to watch videos that work in a dynamic environment. This approach used the combined function of HOG and Gabor [30].

In deep learning features, Wang et al. [31] proposed an algorithm which is useful to mine deep features from small video fragments. Additionally, depiction features of neighboring nodes of the secreted layer were considered according to similar activation states. Zhang et al. [32] introduced a less complex descriptor called 3D histogram texture in order to mine unique features from a given set of depth maps. On 3 orthogonal Cartesian planes, a three-dimensional histogram is formed. In [33], Lan et al. proposed an approach to influence operational methods from data-independent and data-driven methods to make action recognition systems better. Sargano et al. [34] proposed a new technique for recognizing human activity on the basis of the pretrained structure of the deep Convolutional Neural Network (CNN) for extraction and depiction of features in which the support vector machine (SVM) and K-Nearest Neighbor (KNN) are fused to recognize activity. In [35], the authors offered a small radial feature based on imaginary contour points and adapted to reactive real-time processing. Imaging-based features are useful for RGB-D images because of the shape, which is easily viewed as a bit mask based on the depth data provided by Microsoft Connect. Another common feature is presented by Tran and Sorokin [36]. It combines visual flow and silhouette into a single vector of attributes. With radial graphs, the silhouette and optical flow are encoded in X, Y dimensions and linked to a frame of fifteen adjacent frames. Lv and Nevatia [37] suggested a graph of the polynomial calculated by the selection of modified cell beams based on the logarithmic scale. Different kinds of human action include the abnormal activities by using the wireless connection. Support vector machine and the kernel nonlinear regression are used for reduction of the false positive rate. This can be done in the unsupervised learning. The proposed system performs the great function by using the real data [38]. Several techniques are used for finding the human activity in the

videos. The authors worked on the feature correlation and frame differencing [39].

3. Proposed Methodology

In the proposed algorithm, a novel technique for human activity recognition is proposed. The proposed approach comprises of five basic steps, namely, these approaches are (a) detecting moving objects from the video sequence; (b) extracting the HOG, Gabor, and color features of the moving object; (c) selecting the best characteristics; (d) fusing the selected features serially; and (e) classifying the moving object. Figure 1 shows the complete flow of the proposed technique.

3.1. Preprocessing. In the preprocessing stage, region-wise sliding window is implemented by considering variation in each consecutive frame. The needless regions including the background are ignored. A binary image is achieved after background subtraction on which the noise removal technique is applied. The binary image is transformed to RGB color space and later RGB is converted into Hue Saturation Intensity (HSI). In the next phase, a person is detected by drawing a bounding box around the person. The aim of preprocessing is to enhance the quality of video-extracted frames. The input image or frame extracted from a video is in RGB format. Preprocessing is applied to improve the foreground features for further processing. The steps of preprocessing are described below.

3.1.1. Background Subtraction. Background subtraction is the first step in which RGB frames are extracted from the background and foreground videos for frame-by-frame processing. This is done because every processed frame gives us diverse results. With the presence of variations in every processed frame, the authenticity of the proposed algorithm is judged in a better way. In some cases, moving background or the presence of multiple objects is challenging to handle. The background frames are subtracted from the frames having a person doing a particular activity. Before taking the difference, both images are transformed from RGB color to HSI color. After that, the background image is subtracted from the actual image after converting from RGB to HSI. The conversion equation from RGB space and HIS space is given as [40]

$$I = \sqrt{3}(R + G + B), \tag{1}$$

$$\theta = \arccos \frac{(1/2)[(R-G)+(R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B}}, \tag{2}$$

$$H = \{\theta \quad G \geq B\, 2x - \theta \quad G \leq B\}, \tag{3}$$

$$S = 1 - \frac{3\ \min\ (R, G, B)}{R + G + B}, \tag{4}$$

where I, S, and H are intensity, saturation, and hue, respectively. After this step, the binary image is produced. Some samples of the resulted images of this step can be seen in Figure 2(b).

3.1.2. Morphological Operation. Images yielded after general background subtraction steps are noisy as shown in Figure 2(b). Morphological operation is applied to minimize the noise present in the image because the noisy image is not used for further processing. For this purpose, the preprocessing steps can be performed. The operation is known as opening by reestablishment of erosion, and it conserves the underlying shape of the object [41]. Regions having the least number of pixels are removed. The aim of this step is to detect the person in the image easily. After applying the opening morphological operation using the structuring element of 12-pixel-wide circular, the resulting images are much enhanced, and an individual is easily detected from the frame. The outcome of the enhanced image is shown in Figure 2(c). Hence, it is obvious that the opening morphological operation is the necessary step before object detection in the preprocessing stage.

3.1.3. Image Cropping. Pixels from the white region in an image are counted to identify an object. The area which has more than 300 pixels is considered as the required object or human. All the white regions having less than 300 pixels are eliminated which are not required. When the object is detected, the bounding box is drawn around the person and the unnecessary part of the image is removed. The purpose of drawing the bounding box is to get the required part of the image by neglecting the unnecessary part as shown in Figure 2(d).

3.2. Feature Extraction. In the second stage of the proposed algorithm, three different types of extractors including HOG, Gabor, and cooccurrence matrices and chromatic features are employed to get the features of each frame. HOG, Gabor, and cooccurrence matrices and chromatic feature vectors are formed with 1×3780, 1×60, and 1×9 standard dimensions, respectively. Each feature is described as follows.

3.2.1. HOG Features. In HOG feature extraction [14], the image is separated into small segments for individual processing. These segments are joined later. To achieve G_x and G_y directions, the Sobel kernel function is used on processed images. Mathematically, the process is depicted in the following equations.

$$F_{\text{seg}_{|G(i,j)|}} = \sqrt{G_x(i,j)^2 + G_y(i,j)^2}, \tag{5}$$

$$F_{\text{seg}_{\varnothing_G}}(i,j) = \tan - 1\left(\frac{G_y(i,j)}{G_x(i,j)}\right), \tag{6}$$

where $|G|$ represents magnitude, $\varnothing_G$ donates the angle of gradient, and i and j represent rows and columns simultaneously. The angle allocates the cell votes to bins based on the gradient. Later, the standardized vector is being achieved by using every block of the histogram. On the segmented image, the HOG feature descriptor is

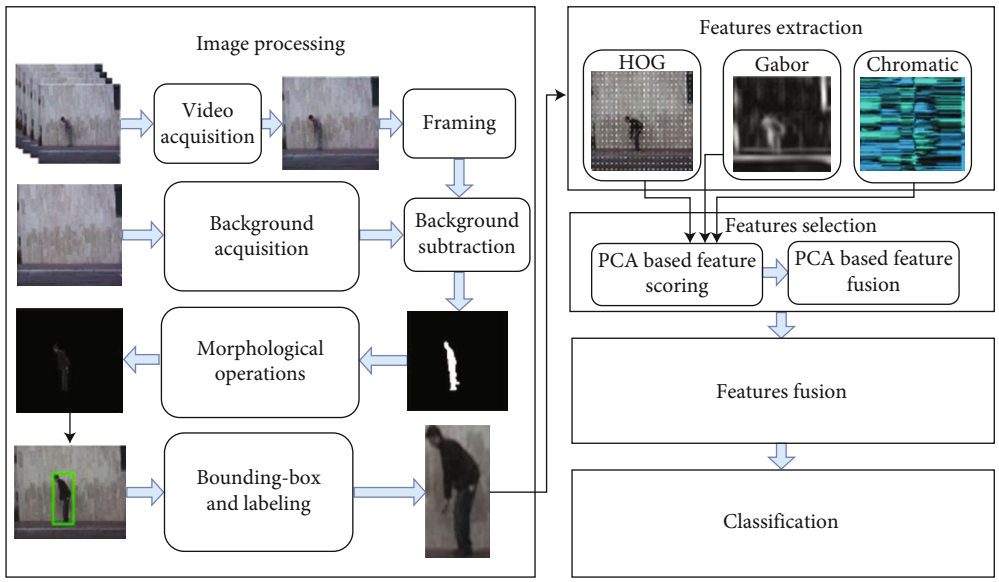


Figure 1: Detailed description of proposed model based on the machine learning methods.

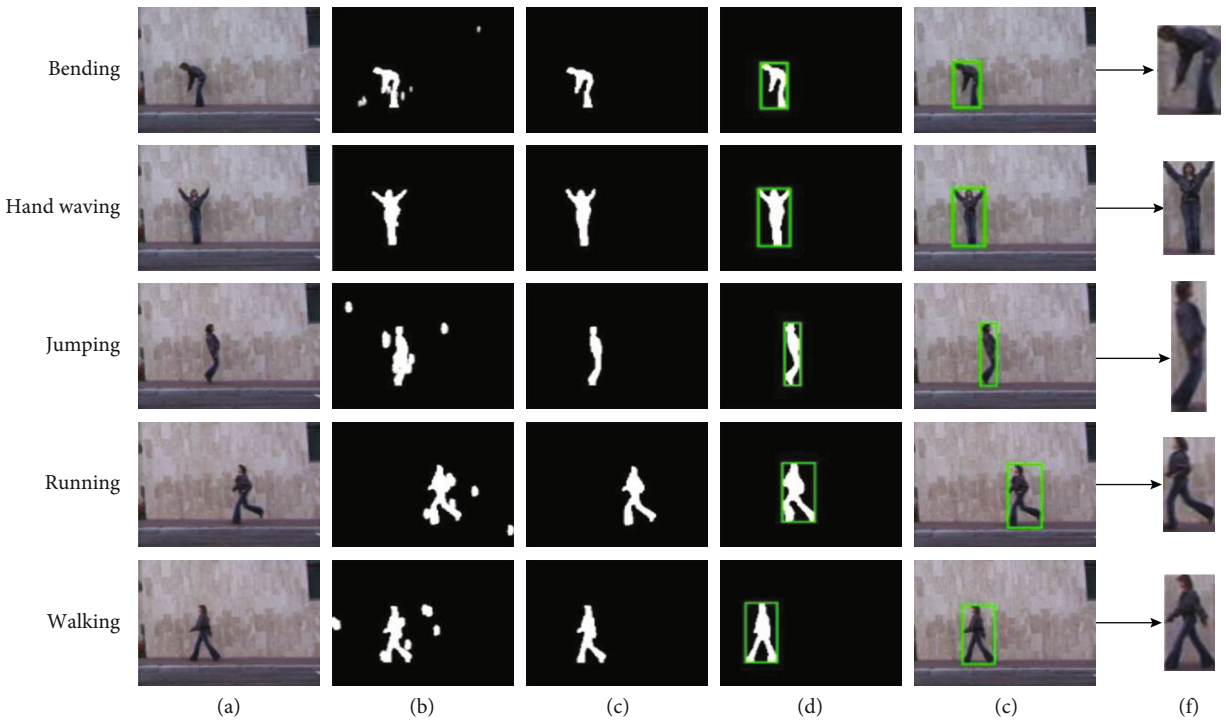


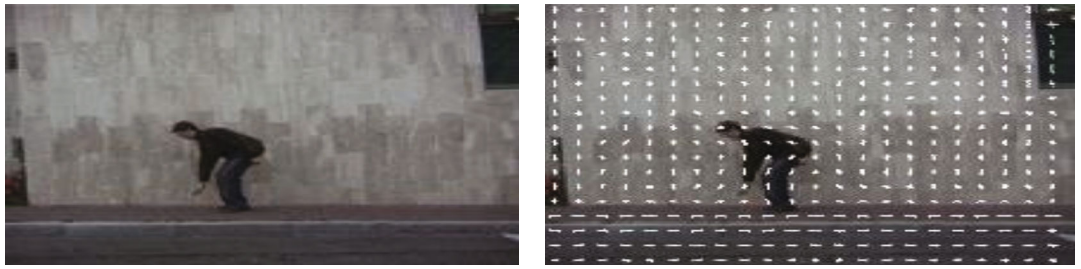
Figure 2: Preprocessing stages: (a) original images; (b) background subtraction images; (c) image enhancement; (d) object detection; (e) binary to RGB conversion; (f) image cropping.

being implemented with 8 bin cells which are represented in the following equation.

$$F_{segV^N{}_i} = \frac{V_i}{\sqrt{(V_2{}^2 + \epsilon^2)}}, \tag{7}$$

where "ϵ" is a minor constant which does not divide by zero and V indicates the vector which is not normalized by containing all histograms in a block. When all of these vectors are combined in a single block, the HOG feature vector is achieved. Furthermore, mean variance and range through each feature are measured. Graphical representation of HOG features is shown in Figure 3.

3.2.2. Gabor Features. In the spatial area, modified 2D-Gabor filter [42] is utilized using the "Gaussian Kernel" feature by a complex sinusoidal wave as shown in the following



(a) (b)

Figure 3: Visualization of histogram of oriented gradient features: (a) original image; (b) HOG features.

equation.

$$F_{\text{seg}} = \frac{fs^2}{\pi Y \eta'} \exp\left(-\frac{p' + Y^2 q'}{2\sigma^2}\right) \exp\left(2\pi fsx' + \varnothing\right). \quad (8)$$

Here fs shows sinusoidal frequency, θ represents band similarity direction of an activity described by Gabor, $\varnothing$ indicates the phase offset, σ indicates the Standard Deviation (SD) of the Gaussian wrapper, and Y shows the characteristics regarding space proportion in which the elliptic support of the function described by Gabor is designated; p' and q' are described in the following equations.

$$x' = x\text{coscos}\,\theta + y\text{sinsin}\,\theta, \quad (9)$$

$$y' = x\text{sinsin}\,\theta + y\text{coscos}\,\theta. \quad (10)$$

Gabor feature [43] is implemented in six directions and five scales. Gabor feature measurement is chosen as 1×30. The variance and mean through the Gabor feature are measured. Graphical representation of HOG features is described in Figure 4.

3.2.3. Cooccurrence Matrices and Chromatic Features. Grey tone spatial dependence is linked with cooccurrence technique. This approach works with the approximation. The function of the second-order density probability $h(i, j \mid d, \theta)$ is approximated. Each combined density function of the second order is calculated by measuring all pairs of pixels which are separated by distance d having gray levels i and j in the direction of the angle. The angular displacement θ is generally understood in the following interval: $\theta = \{0, \pi/4, \pi/2, 3\pi/4\}$. The correlation table records a considerable amount of textual information. For a rough texture, these matrices usually have high values near the main diameter, while the costs are split into a soft texture. The cooccurrence matrices are summarized from the different directions to obtain a rotational invariant characteristic. This technique has become a reference point because of its intensive use [44], while other researchers relied on a smaller number of functions, such as entropy (H), correlation (COR), energy (E), and local homogeneity (LH).

$$E = \sum_i \sum_j [h(i, j \mid d, \theta)]^2, \quad (11)$$

$$H = -\sum_i \sum_j [h(d, \theta)\ \log\ h(d, \theta)], \quad (12)$$

$$I = \sum_i \sum_j \left[(i - j)^2\, h(d, \theta\right], \quad (13)$$

$$\text{LH} = \sum_i \sum_j \frac{(i, j \mid d, \theta)}{1 + (i + j)^2}, \quad (14)$$

$$\text{COR} = \sum_i \sum_j \frac{(i - \mu_x)\left(j - \mu_y\right) h(i, j \mid d, \theta)}{\sigma_x \sigma_y}, \quad (15)$$

where μ_x is the horizontal mean, σ_x is the variance, and both μ_y and σ_y are the vertical statistics.

This technique records second-degree grayscale statistics related to human perception and texture discrimination which are used with various disadvantages [45]. The disadvantage of the given technique is that it does not explain the aspects of the shape and type of texture. In addition, this technique involves choosing an appropriate level of quantification. Text information may be lost due to the reduced number of antenna sizes at the quantization level. And a relatively large number of compartments can lead to irrelevant text features.

3.3. Feature Selection. The sensitivity of various machine defect features differs meaningfully in dissimilar working circumstances. It becomes vital to develop an organize feature selection structure. This provides the basis for organization of descriptive structures [46]. In the proposed technique, PCA is used for feature selection to select the prominent features separately from the results of HOG, Gabor, and cooccurrence matrices and chromatic feature vectors.

Generally, the PCA method converts from d-dimensional space of n vectors to another space of d'-dimensions having n vectors $(x_1', x_2', \cdots, x_i', \cdots, x_n')$ as given by the following equation [47].

$$x_r' = \sum_{n=1}^{d'} a_{n,r} e_n, \quad d' \le d, \quad (16)$$

where e_n shows the eigenvectors relating with d'-dimensional space and largest eigenvalues of the disseminated

(a) (b)

Figure 4: Gabor feature visualization: (a) original image; (b) Gabor features.

matrix S. On the other hand, $a_{n,r}$ are forecasts of the vectors xr on the eigenvectors. These are the main constituents of true datasets. The d and d' are both positive integers such that d' cannot be greater than d in any of the cases. $d \times d$ matrix S represents the original dataset $(x_1, x_2, \cdots, x_i, \cdots, x_n)$ which is defined as

$$S = E\left[x_i x_i^T\right], \quad \text{for } i = 1 \text{ to } n, \tag{17}$$

where $E[x_i x_i^T]$ is the "statistical expectancy operator" implemented on the external multiplicative product of x_i and x_i^T. The depiction illustrated in Equation (26) decreases the occurrence of error between the converted vectors and the original. If the variance of principal components such as $(a_{n,r})$ is considered, the problem is simplified.

3.4. Feature Fusion. The purpose of feature fusion makes the action recognition algorithm efficient. This also enhances the action classification rate of human in complicated scenarios. In this method, feature fusion produces considerably improved results not only in the dark background but also in the high brightness environment as compared to original Gabor and HOG features. Hand-crafted features are combined with the deep learning models. The model is named as the posteriori algorithm [48].

For fusing the features, a unique method is deployed which depends on the vector dimension size. The size of these feature vectors are 1×60, 1×3780, and 1×9 in HOG, Gabor, and cooccurrence matrices and chromatic features, respectively. For feature fusion, let $C_1, C_2, C_3, \cdots, C_n$ be the human activity classes, which need to be classified. Let $\Delta = \{\varnothing \text{v} \varnothing \in \text{RN}\}$ indicate the number of training samples. $\{\Upsilon_{\text{HOG}}, \Upsilon_{\text{Gab}}, \Upsilon_{\text{Chrom}}\} \in R^{N_{HOG+Gab+Chrom}}$ are the three feature vectors extracted. The size is defined as

$$\text{FV}_1 = \{j_1, \cdots j_k\}, \text{FV}_2 = \{y_1, \cdots y_k\}, \text{FV}_3 = \{d_1, \cdots d_k\}, \tag{18}$$

where FV_1, FV_2, and FV_3 indicate the size of HOG, Gabor, and cooccurrence matrices and chromatic features, respectively. The sizes of the feature vectors are characterized through set k, where $k \in \{60, 3780, 9\}$. As discussed earlier, the sizes of extracted feature sets are $(' \Upsilon_{\text{HOG}} \longrightarrow 1 \times 3780, ' \Upsilon \text{Gab} \longrightarrow 1 \times 60, ' \Upsilon \text{Chrom} \longrightarrow 1 \times 9)$. The final extracted vector is indicated as

$$F(\varnothing) = \sum_{j}^{\text{FV}_1} \Upsilon_{\text{HOG}} + \sum_{t}^{\text{FV}_2} \Upsilon_{\text{Gab}} + \sum_{d}^{\text{FV}_3} \Upsilon_{\text{Chrom}}, \tag{19}$$

$$F(\varnothing) = \{(1 \times 3780) + (1 \times 60) + (1 \times 9)\}, \tag{20}$$

$$\text{Final}\,(\varnothing) = \{1 \times 3849\}. \tag{21}$$

3.5. Classification. Five different classifiers including linear-SVM, cubic-SVM, complex tree, fine-KNN, and subspace-KNN are used for result comparisons. Figure 5 depicts the detailed view of feature selection, fusion, and selection.

The accuracy achieved by subspace-KNN is highest among all the classifiers on the KTH dataset, while cubic-SVM has achieved higher accuracy than other classifiers on the Weizmann dataset. The random subspace approach depends on a stochastic procedure which selects the components of the particular feature vector randomly to construct every classifier. In the KNN classifier, when a testing sample is compared to the original, only the chosen features will get the nonzero contributions [49]. On the other hand, State Vector Machines construct models which are complicated and contain radial basis function (RBF), polynomial classifiers, and large neural nets. It is easy to examine mathematically; it resembles a linear method in a multidimensional feature space nonlinearly associated with the input space [50].

4. Results and Analysis of Experiment

The experimental setups, datasets, and results based on the performance measures are discussed in this section.

4.1. Experimental Setup. The time elapsed during activity classification depends on resources such as memory, Central Processing Unit (CPU) speed, power supplies, disk storage, and cooling systems. This can precisely describe a linear relationship between elapsed time and CPU usage. The tested system (DELL Latitude E5520) to run the proposed algorithm consists of a Microsoft Windows 10 Pro environment with Intel Core-i5 2540M @ 2.60 GHz processor. The system RAM is 4.00 GB with a 64-bit operating system and an x64 processor. All the results presented in this section are the results obtained in this system.

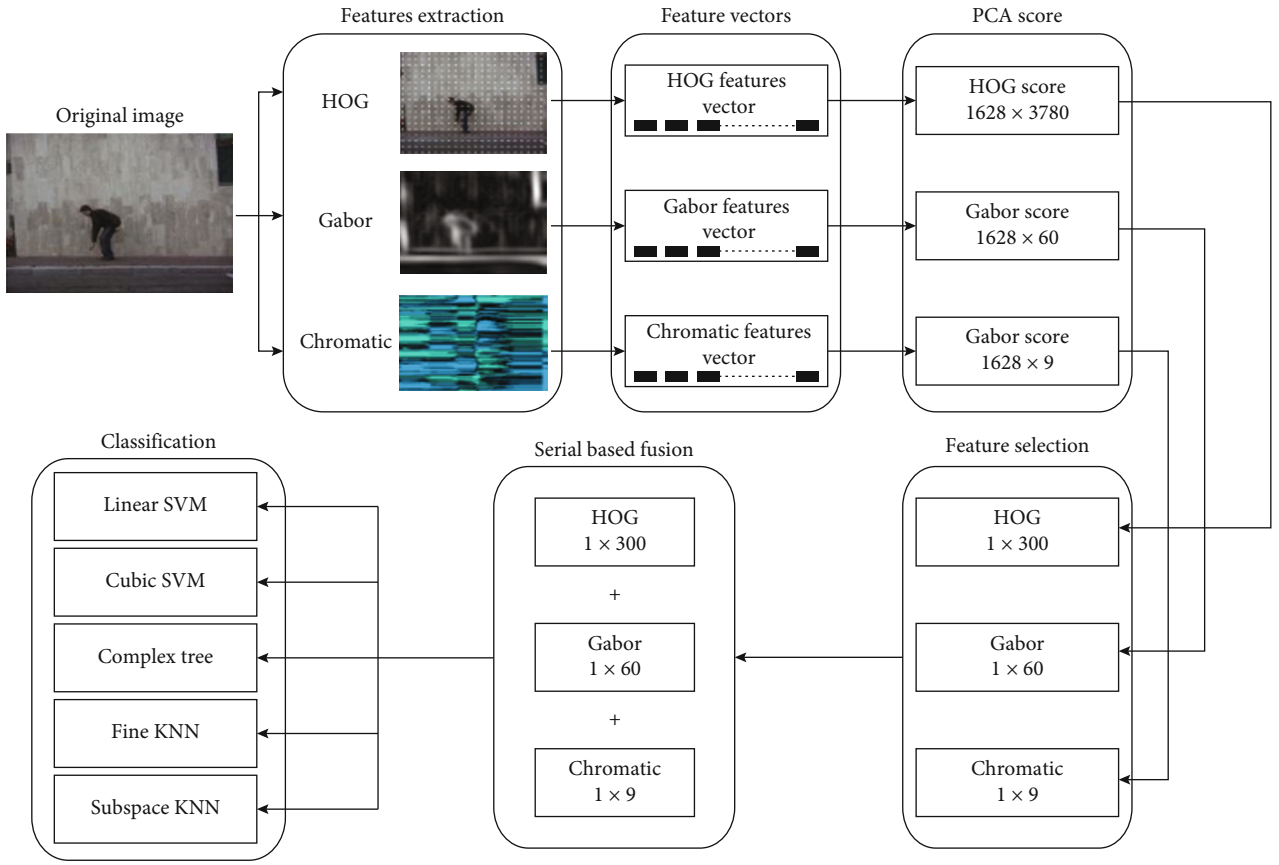


Figure 5: Feature vector selection, fusion, and classification.

4.2. Datasets. To validate the results, two different types of datasets are used in this research. This includes KTH and Weizmann datasets. Both of the datasets are described briefly hereunder.

4.2.1. Weizmann Dataset. The Weizmann dataset [51] contains 2513 human activity images. It covers five types of human behavior performed by nine different actors. To verify the proposed algorithm, a 50 : 50 method is run. This means that half of the images are used for training the classifier and the remaining half are used to test the performance of the algorithm. After the selection and fusion of features, classification techniques are applied to evaluate the results. Figure 6 shows some images as a sample from the Weizmann dataset.

4.2.2. KTH Datasets. KTH datasets consists of 1628 images of six different types of human activities. Images with different variations are chosen for the authentication of the proposed method. Half of the images are used in the training of the classifier while the remaining 50% are used for the purpose of testing. Figure 7 shows some of the images from the Weizmann datasets. These datasets include boxing, clapping, hand waving, running, and walking.

4.3. Performance Measures. Performance of the proposed algorithm is assessed on the basis of performance measures such as specificity (SPE), area under the curve (AUC), precision (PRE), sensitivity (SEN), and accuracy (ACU). Mathematically, it is represented by the following equations.

$$\mathrm{PRE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{22}$$

$$\mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{23}$$

$$\mathrm{SPE} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \tag{24}$$

$$\mathrm{ACU} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{FP} + \mathrm{TP} + \mathrm{FN} + \mathrm{TN}}, \tag{25}$$

$$\mathrm{AUC} = \int_{\infty}^{-\infty} \frac{\mathrm{TPR}(T)\mathrm{FPR}}{(T)dT}. \tag{26}$$

In the above equations, FP represents false positive, TN represents true negative, TP represents true positive, and FN represents false negative.

All of the performance measures mentioned in Equation (20) to (25) are calculated from confusion matrices. These matrices have the finest results of the Weizmann and KTH datasets.

4.4. Experiments. For quantitative outcomes, six different experiments are implemented using a separate number of

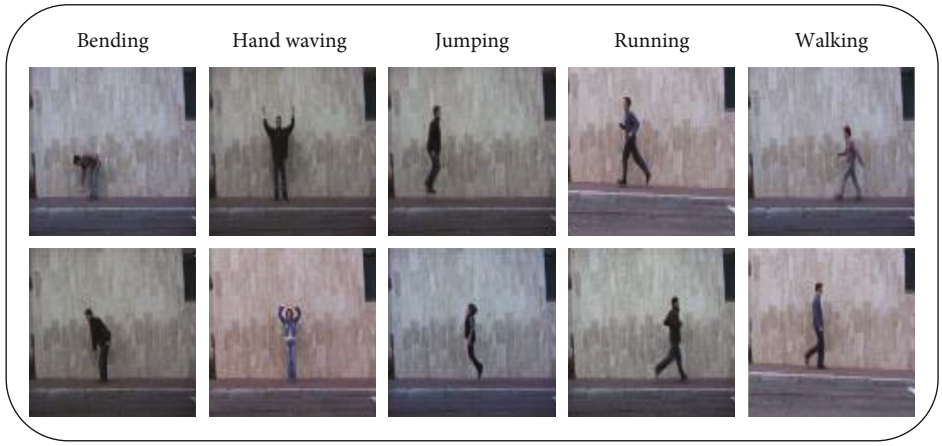


Figure 6: Sample images of Weizmann dataset.

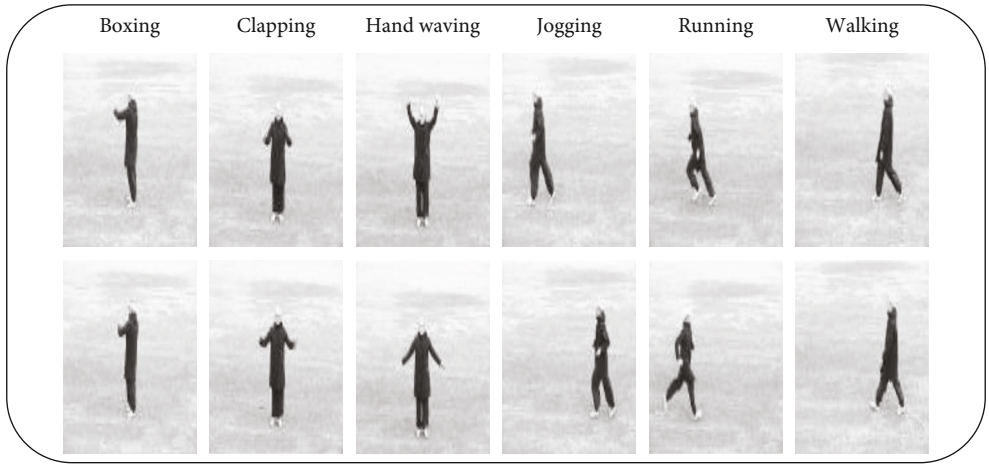


Figure 7: Sample images of KTH dataset.

features. The comprehensive description of all experiments with a numeral classes, numeral folds, and features can be seen in Table 1. The comprehensive analyses of experiments performed on 316 bend, 624 hand waving, 457 jumping, 405 run, and 711 walk images are described in the upcoming sections.

4.4.1. Experiment 1: Shape Features—100, Texture Features—60, and Color Features—9. In experiment 1, a total of 2513 and 1628 images are collected from the Weizmann and KTH datasets, respectively. The Weizmann dataset consists of 5 categories of manual bending, jumping, running, and walking images, while the KTH dataset includes 6 classes which are clapping, boxing, running, hand waving, and walking. To get the experimental results, 50% of images are used for the purpose of training and the remaining 50% of them are used for testing. For assessment of the results, the "5-fold" validation is used. For experiment 1, the maximum classification rate is 99.3% for the Weizmann dataset obtained with cubic-SVM. The linear-SVM and subspace-KNN obtained 99.8% accuracy simultaneously on the KTH dataset as shown in Table 2. Cubic-SVM obtained a better sensitivity rate of 98.84, specificity of 99.81 and accuracy of 98.98 as compared to other classification methods using the Weizmann dataset. On the other hand, linear-SVM and subspace-KNN obtained a sensitivity rate of 99.86, specificity of 99.96, and precision of 98.74 which is better in comparison with other classification methods using the KTH dataset.

4.4.2. Experiment 2: Shape Features—300, Texture Features—60, and Color Features—9. In experiment 2, 2513 and 1628 images are taken from the Weizmann and KTH datasets, respectively. The Weizmann dataset includes five categories. These five categories are bending, handshaking, jumping, running, and walking, while the KTH dataset includes 6 classes which are boxing, clapping, handshake, jogging, running, and walking. For experimental results, half of the images from each dataset are used for training and the remaining half are used for the purpose of testing. For assessment of the results, "10-fold" validation is used. The 10-fold validation is known as the evaluation method. For experiment 2, the maximum classification frequency is 99.5% for the Weizmann dataset on cubic-SVM, while for the KTH dataset, 99.9% is achieved in the subspace-KNN, as shown in Table 3. The cubic-SVM applied to the

Table 1: Summary of experiments setting for Weizmann and KTH datasets.

Exp no.	No. of classes		Shape	Texture	Color	Folds
	KTH	Weizmann				
1	6	5	100	60	9	5
2	6	5	300	60	9	10
3	6	5	500	60	9	8
4	6	5	800	58	9	5
5	6	5	1100	55	9	7

Weizmann dataset is better in terms of sensitivity of 99.34, specificity of 99.89, and precision of 99.5 than other approaches, whereas the KNN subdomain applied to the KTH dataset is better in terms of sensitivity of 99.97, specification of 99.99, and accuracy of 99.95 than the other classification approaches.

The experiment 2 produces the best results among all the five experiments implemented during this research process. The proposed algorithm produced the best results on the conditions provided in the experiment 2. The best results calculated on the basis of performance measures of the KTH and Weizmann datasets using confusion matrices are shown in Tables 4 and 5, respectively. The KTH datasets give 99.9% accuracy using the subspace-KNN classifier and the Weizmann dataset produced 99.5% accuracy using cubic-SVM which is best among all other classifiers.

4.4.3. Experiment 3: Shape Features—500, Texture Features—60, and Color Features—9. In experiment 3, a total of 2513 images of the Weizmann dataset and 1628 images of the KTH dataset are collected. Five classes from the Weizmann dataset are selected which includes bending, hand waving, jumping, running, and walking. Six classes from the KTH dataset including clapping, boxing, running, hand waving, walking, and hand waiving are selected.

For experimental results, half of images from both datasets are selected for training purposes and the remaining half are used for testing. For assessment of the results, "8-fold" validation is used. Maximum classification frequency attained on cubic-SVM is 98.7% for the Weizmann datasets. For the KTH datasets, 99.9% accuracy is attained on subspace-KNN as given in Table 6. The cubic-SVM implemented for the Weizmann datasets is better in terms of sensitivity of 97.86, specificity of 99.67, and precision of 98.54 as compared to other approaches. On the other hand, the subspace-KNN implemented for the "KTH" datasets is better in terms of sensitivity of 99.97, specificity of 99.99, and precision of 99.95 as compared to other approaches.

4.4.4. Experiment 4: Shape Features—800, Texture Features—59, and Color Features—9. In experiment 4, total of 2513 images of Weizmann datasets and 1628 images of KTH datasets are collected. The Weizmann datasets are comprised of 5 classes including bending, hand waving, jumping, running, and walking images. The KTH datasets are comprised of 6 classes including clapping, boxing, running, hand waving, walking, and hand waiving. For experimental results, half of images from both the datasets are selected for training and the other half of them are used for testing. For assessment of the results, "5-fold" validation is used. Maximum classification frequency of 95.9% for the Weizmann datasets is attained on linear-SVM while 99.9% for the KTH dataset on subspace-KNN as given in Table 7. The linear-SVM implemented for the Weizmann dataset is better in terms of sensitivity of 93.38, specificity of 98.96, and precision of 96.09 from other approaches. On the other hand, the subspace-KNN implemented for the "KTH" dataset is better in terms of sensitivity of 99.97, specificity of 99.99, and precision of 99.45 from other approaches.

4.4.5. Experiment 5: Shape Features—1100, Texture Features—55, and Color Features—9. In experiment 5, a total of 2513 images of the Weizmann dataset and 1628 images of the KTH dataset are collected. The Weizmann dataset comprising of 5 classes including bending, hand waving, jumping, running, and walking images is selected. The KTH dataset comprising of 6 classes including clapping, boxing, running, hand waving, and walking is selected. For experimental results, half of the images from both datasets is selected for training the algorithm and the remaining half is used for testing. For appraisal of the results, "7-fold" validation is used. Maximum classification frequency is 92.0% for the Weizmann datasets on subspace-KNN while 99.9% for the KTH dataset on subspace-KNN as given in Table 8. The class-wise AUCs are mentioned in Table 9. The subspace-KNN implemented for the Weizmann datasets is better in terms of sensitivity of 88.72, specificity of 97.96, precision of 90.81, and AUC. We can say that subspace-KNN gives the better results. On the other hand, the subspace-KNN implemented for "KTH" datasets is better in terms of sensitivity of 99.97, specificity of 99.99, and precision of 99.95. The results of experiment 5 are presented in Table 8.

5. Discussion

This section presents a detailed analysis of experimental outcomes through the proposed method on the basis of accuracy measures such as precision, sensitivity, specificity, and accuracy. The proposed algorithm consists of five main stages. These five main stages include the preprocessing which is performed first in which datasets are normalized to get better results. The accurate results will give more accuracy. In the second step, feature extraction is implemented using HOG, Gabor, and chromatic feature extractor. In the third step, feature selection is implemented separately based on PCA to get the best features. In the fourth step, features are fused, while in the final stage, results are taken through the classification learner. In preprocessing, background subtraction is done to detect the human from the image and the noise is removed using morphological operations. After removing the noise, a bounding box is drawn to around the person to ignore the unnecessary parts using cropping. In the next step, three kinds of features comprising shape, texture, and color are extracted from segmented images. Five classifiers containing linear-SVM, cubic-SVM, complex tree,

Table 2: Classification results of experiment 1 with all possible values.

Weizmann					KTH			
Method	SEN (%)	SPE (%)	PRE (%)	ACU (%)	SEN (%)	SPE (%)	PRE (%)	ACU (%)
Linear-SVM	98.8	99.7	98.52	98.8	99.85	99.95	99.03	99.8
Cubic-SVM	98.84	99.81	98.98	99.3	99.8	99.94	99.09	99.7
Complex tree	85.92	97.33	86.06	89.0	98.26	99.67	97.77	98.4
Fine-KNN	98.99	99.78	99.23	99.0	99.77	99.91	99.75	99.6
Subspace-KNN	90.3	98.30	91.7	93.3	99.86	99.96	99.74	99.8

Table 3: Classification results of experiment 2 along with the sensitivity and other measures.

Weizmann					KTH			
Method	SEN (%)	SPE (%)	PRE (%)	ACU (%)	SEN (%)	SPE (%)	PRE (%)	ACU (%)
Linear-SVM	98.83	99.84	99.0	99.3	99.89	99.95	99.89	99.8
Cubic-SVM	99.34	99.89	99.5	99.5	98.81	99.94	98.60	99.7
Complex tree	85.40	97.15	85.0	88.3	98.32	99.67	97.37	98.5
Fine-KNN	87.26	97.21	95.4	91.1	99.42	99.84	99.38	99.2
Subspace-KNN	90.25	98.3	91.9	93.9	99.97	99.99	99.95	99.9

Table 4: Confusion matrix of KTH dataset of experiment 2 on subspace-KNN.

Classification classes	Total images	Clapping	Jogging	Hand waving	Running	Walking	Boxing
Clapping	312	312					
Jogging	191		191				
Hand waving	581	1		580			
Running	109				109		
Walking	27					27	
Boxing	408						408

Table 5: Confusion matrix of Weizmann dataset of experiment 2 on cubic-SVM.

Classification classes	Total images	Hand waving	Running	Jumping	Walking	Bending
Hand waving	624	624				
Running	206		201	2	3	
Jumping	421	1		420		1
Walking	271		2	2	267	
Bending	375					375

Table 6: Classification results of experiment 3 using the linear-SVM method and others.

Weizmann					KTH			
Method	SEN (%)	SPE (%)	PRE (%)	ACU (%)	SEN (%)	SPE (%)	PRE (%)	ACU (%)
Linear-SVM	97.90	99.69	98.39	98.7	100	99.72	98.70	98.7
Cubic-SVM	97.86	99.67	98.544	98.7	91.51	99.21	96.39	96.5
Complex tree	85.43	97.212	85.53	88.6	98.34	99.69	97.93	98.5
Fine-KNN	63.7	91.63	90.79	71.9	95.30	99.32	98.02	97.0
Subspace-KNN	89.99	98.18	92.11	93.0	99.97	99.99	99.95	99.9

fine-KNN, and subspace-KNN are used to test the proposed algorithm. Subspace-KNN and cubic-SVM have achieved higher accuracy than the other classification learners on the "Weizmann" and "KTH" datasets, respectively.

In the experimental results, six classes of KTH datasets and five classes of Weizmann datasets are used to get results. 99.90% and 99.5% accuracies for KTH datasets using subspace-KNN are achieved using subspace-KNN and

Table 7: Classification results of experiment 4.

Method	Weizmann SEN (%)	Weizmann SPE (%)	Weizmann PRE (%)	Weizmann ACU (%)	KTH SEN (%)	KTH SPE (%)	KTH PRE (%)	KTH ACU (%)
Linear-SVM	93.38	98.96	96.09	95.9	59.25	96.7	90.90	85.3
Cubic-SVM	91.56	98.5	95.63	94.5	57.56	96.45	74.49	84.3
Complex tree	80.56	97.18	85.57	88.5	97.74	99.63	97.51	98.2
Fine-KNN	46.29	87.54	88.81	58.2	70.14	96.30	90.90	84.0
Subspace-KNN	89.21	98.37	91.44	92.5	99.97	99.99	99.45	99.9

Table 8: Classification results of experiment 5.

Method	Weizmann SEN (%)	Weizmann SPE (%)	Weizmann PRE (%)	Weizmann ACU (%)	KTH SEN (%)	KTH SPE (%)	KTH PRE (%)	KTH ACU (%)
Linear-SVM	80.0	96.68	90.0	87.5	50.83	95.49	57.16	80.2
Cubic-SVM	80.0	96.20	90.31	86.3	50.42	95.37	55.03	79.5
Complex tree	85.32	97.8	85.31	88.4	98.49	99.69	97.93	98.6
Fine-KNN	33.42	84.1	87.63	46.7	39.47	93.6	84.92	57.6
Subspace-KNN	88.72	97.96	90.81	92.0	99.97	99.99	99.95	99.9

Table 9: AUC results.

Method	Classes	Experiment 1 Weizmann	Experiment 1 KTH	Experiment 2 Weizmann	Experiment 2 KTH	Experiment 3 Weizmann	Experiment 3 KTH	Experiment 4 Weizmann	Experiment 4 KTH	Experiment 5 Weizmann	Experiment 5 KTH
Linear-SVM	C1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	C2		1.00		1.00		1.00		0.98		0.90
	C3	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.98	1.00
	C4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98
	C5	1.00	1.00	1.00	1.00	1.00	0.98	0.99	1.00	0.98	1.00
	C6		1.00		1.00		1.00	1.00	1.00	1.00	1.00
Cubic-SVM	C1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	C2	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.98	0.90
	C3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
	C4	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.98
	C5	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00
	C6		1.00		1.00		1.00		1.00		0.99
Complex tree	C1	0.98	1.00	0.99	1.00	0.98	1.00	0.98	0.99	0.98	0.99
	C2		0.98		0.96		0.98		0.99		0.99
	C3	0.87	1.00	0.89	1.00	0.85	1.00	0.88	0.99	0.87	0.99
	C4	0.94	0.99	0.93	0.98	0.94	0.99	0.94	0.99	0.94	0.96
	C5	0.90	0.98	0.88	0.96	0.90	0.98	0.88	1.00	0.87	1.00
	C6	0.97	1.00	0.95	1.00	0.97	1.00	0.96	1.00	0.96	1.00
Fine-KNN	C1	1.00	0.98	0.93	0.97	0.79	0.98	0.69	0.95	0.60	0.75
	C2		1.00		1.00		0.99		0.88		0.54
	C3	0.99	0.98	0.87	0.99	0.63	0.98	0.51	0.91	0.50	0.76
	C4	0.99	0.81	0.91	0.69	0.71	0.64	0.65	0.61	0.62	0.50
	C5	0.99	1.00	0.92	1.00	0.76	1.00	0.61	0.76	0.51	0.54
	C6	1.00	0.98	1.00	0.96	0.99	0.94	0.89	0.91	0.71	0.81
Subspace-KNN	C1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	C2		1.00		1.00		1.00		1.00		1.00
	C3	0.98	1.00	0.98	1.00	0.98	1.00	0.98	1.00	0.98	1.00
	C4	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00
	C5	0.99	1.00	0.99	1.00	0.99	1.00	0.99	1.00	0.99	1.00
	C6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 10: Comparison of action recognition results.

Reference	Year	Recognition (%)
Weizmann dataset		
[52]	2013	95.45
[53]	2014	95.56
[54]	2015	95.10
[55]	2016	88.10
[7]	2017	95.80
[56]	2019	99.0
[57]	2020	96.0
Proposed	2018	99.5
KTH dataset		
[58]	2014	95.0
[59]	2015	95.21
[60]	2015	96.50
[61]	2016	97.10
[62]	2017	94.92
[7]	2017	99.30
[63]	2020	94.83
[64]	2019	91.67
Proposed	2018	99.90

cubic-SVM, respectively. Comparison of the previously implemented algorithms with the proposed algorithm is shown in Table 10. The table is explained in a better way. On the basis of the above discussion, it is clear that the combination of three feature extractors used in the proposed algorithm gives better accuracy as compared to the already implemented algorithms.

6. Conclusion

In this research, a technique is proposed for the "detection" and "classification" of several activities from videos and multimedia frames. The proposed algorithm consists of five pipeline processes which are preprocessing, feature extraction, feature selection, serial feature fusion, and classification. From all results shown above and in Discussion, it is obvious that by using the proposed technique, the detection of human activities is tackled. Five different experiments are executed to judge the authenticity of this algorithm. The results of all five experiments are discussed in detail in Results and Analysis of Experiment. The KTH and Weizmann datasets are selected to check the reliability of this algorithm. This method executed better on the KTH and Weizmann datasets. Moreover, it is determined that shape features are very important for the classification of chosen classes such as bending, jumping, running, walking, and hand waving. The texture and color features are very essential for the detection and classification of different usual activities performed by a human being. To enhance the system performance, feature selection and feature fusion seem to be quite significant as accuracy and sensitivity. In contrast to the existing techniques, the proposed technique has achieved higher accuracy which is 99.9% on KTH datasets and 99.5% on Weizmann datasets.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

There are no conflicts of interest associated with publishing this paper.

Acknowledgments

This research was supported by the Researchers Supporting Project number (RSP-2021/244), King Saud University, Riyadh, Saudi Arabia.

References

[1] D. Thombre, J. Nirmal, and D. Lekha, "Human detection and tracking using image segmentation and Kalman filter," in *in Intelligent Agent & Multi-Agent Systems, 2009. IAMA 2009. International Conference on*, pp. 1–5, Chennai, India, 2009.

[2] R. Xu, Y. Guan, and Y. Huang, "Multiple human detection and tracking based on head detection for real-time video surveillance," *Multimedia Tools and Applications*, vol. 74, pp. 729–742, 2015.

[3] J.-W. Hsieh, S.-H. Yu, Y.-S. Chen, and W.-F. Hu, "Automatic traffic surveillance system for vehicle tracking and classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, pp. 175–187, 2006.

[4] G. C. Dedes and K. C. Mouskos, *GPS/IMU/video/radar absolute/relative positioning communication/computation sensor platform for automotive safety applications Google Patents*, 2014.

[5] Y. Tian, P. Luo, X. Wang, and X. Tang, "Pedestrian detection aided by deep learning semantic tasks," in *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5079–5087, Boston, MA, USA, 2015.

[6] Y. Xu, D. Xu, S. Lin, T. X. Han, X. Cao, and X. Li, "Detection of sudden pedestrian crossings for driving assistance systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 3, pp. 729–739, 2012.

[7] M. Sharif, M. A. Khan, T. Akram, M. Y. Javed, T. Saba, and A. Rehman, "A framework of human detection and action recognition based on uniform segmentation and combination of Euclidean distance and joint entropy-based features selection," *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, 2017.

[8] R. Krerngkamjornkit, *Novel robust computer vision algorithms for micro autonomous systems*, 2014.

[9] M. A. Goodrich and A. C. Schultz, "Human–robot interaction: a survey," *Foundations and Trends® in Human–Computer Interaction*, vol. 1, pp. 203–275, 2007.

[10] Q. Ye, Z. Han, J. Jiao, and J. Liu, "Human detection in images via piecewise linear support vector machines," *IEEE Transactions on Image Processing*, vol. 22, pp. 778–789, 2013.

[11] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang, "Illumination normalization for face recognition and uneven

background correction using total variation based image models," in *in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, pp. 532–539, San Diego, CA, USA, 2005.

[12] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: an overview," *Computer Science Review*, vol. 11-12, pp. 31–66, 2014.

[13] Z. Lin and L. S. Davis, "Shape-based human detection and segmentation via hierarchical part-template matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 4, pp. 604–618, 2010.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, pp. 886–893, San Diego, CA, USA, 2005.

[15] S. Zhang, C. Bauckhage, and A. B. Cremers, "Informed Haar-like features improve pedestrian detection," in *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 947–954, Columbus, OH, USA, 2014.

[16] W. Gao, H. Ai, and S. Lao, "Adaptive contour features in oriented granular space for human detection and segmentation," in *in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 1786–1793, Miami, FL, USA, 2009.

[17] A. Satpathy, X. Jiang, and H.-L. Eng, *Human detection by quadratic classification on subspace of extended histogram of gradients*, 2014.

[18] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," *International Journal of Computer Vision*, vol. 77, pp. 259–289, 2008.

[19] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," in *in Computer vision, 2009 IEEE 12th international conference on*, pp. 24–31, Kyoto, Japan, 2009.

[20] P. Viola, M. J. Jones, and D. Snow, *Detecting pedestrians using patterns of motion and appearance*, p. 734, 2003.

[21] R. Poppe, "A survey on vision-based human action recognition," *Image and Vision Computing*, vol. 28, pp. 976–990, 2010.

[22] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequential deep learning for human action recognition," in *in International Workshop on Human Behavior Understanding*, pp. 29–39, Zurich, Switzerland, 2011.

[23] N. Yala, B. Fergani, and A. Fleury, "Towards improving feature extraction and classification for activity recognition on streaming data," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, pp. 177–189, 2017.

[24] U. M. Nunes, D. R. Faria, and P. Peixoto, "A human activity recognition framework using max-min features and key poses with differential evolution random forests classifier," *Pattern Recognition Letters*, vol. 99, pp. 21–31, 2017.

[25] V. Kantorov and I. Laptev, "Efficient feature extraction, encoding and classification for action recognition," in *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2593–2600, Columbus, OH, USA, 2014.

[26] J. Liu, S. Ali, and M. Shah, "Recognizing human actions using multiple features," in *in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8, Anchorage, AK, USA, 2008.

[27] S. Azary and A. Savakis, "3D action classification using sparse spatio-temporal feature representations," in *in International Symposium on Visual Computing*, pp. 166–175, Crete, Greece, 2012.

[28] O. Oreifej and Z. Liu, "Hon4d: histogram of oriented 4d normals for activity recognition from depth sequences," in *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 716–723, Portland, OR, USA, 2013.

[29] C. Conde, D. Moctezuma, I. M. De Diego, and E. Cabello, "HoGG: Gabor and HoG-based human detection for surveillance in non-controlled environments," *Neurocomputing*, vol. 100, pp. 19–30, 2013.

[30] K. Bhuvaneswari and H. A. Rauf, "Edgelet based human detection and tracking by combined segmentation and soft decision," in *in Control, Automation, Communication and Energy Conservation, 2009. INCACEC 2009. 2009 International Conference on*, pp. 1–6, Perundurai, India, 2009.

[31] Q. Wang, D. Gong, M. Li, C. Zhao, and Y. Lei, "Sparse feature auto-combination deep network for video action recognition," in *in 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 712–716, Guilin, China, 2017.

[32] B. Zhang, Y. Yang, C. Chen, L. Yang, J. Han, and L. Shao, "Action recognition using 3D histograms of texture and a multi-class boosting classifier," *IEEE Transactions on Image Processing*, vol. 26, pp. 4648–4660, 2017.

[33] Z. Lan, S.-I. Yu, D. Yao, M. Lin, B. Raj, and A. Hauptmann, "The best of both worlds: combining data-independent and data-driven approaches for action recognition," in *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 123–132, Las Vegas, NV, USA, 2016.

[34] A. B. Sargano, X. Wang, P. Angelov, and Z. Habib, "Human action recognition using transfer learning with deep representations," in *in Neural Networks (IJCNN), 2017 International Joint Conference on*, pp. 463–469, Anchorage, AK, USA, 2017.

[35] A. Chaaraoui, J. Padilla-Lopez, and F. Flórez-Revuelta, "Fusion of skeletal and silhouette-based features for human action recognition with RGB-D devices," in *in Proceedings of the IEEE international conference on computer vision workshops*, pp. 91–97, Sydney, NSW, Australia, 2013.

[36] D. Tran and A. Sorokin, "Human activity recognition with metric learning," in *in European conference on computer vision*, pp. 548–561, Berlin, Heidelberg, 2008.

[37] F. Lv and R. Nevatia, "Single view human action recognition using key pose matching and viterbi path searching," in *in Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1–8, Minneapolis, MN, USA, 2007.

[38] J. Yin, Q. Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082–1090, 2008.

[39] W. Niu, J. Long, D. Han, and Y.-F. Wang, "Human activity detection and recognition for video surveillance," in *In 2004 IEEE international conference on multimedia and expo (ICME)(IEEE Cat. No. 04TH8763)*, vol. 1, pp. 719–722, IEEE, 2004.

[40] J.-f. Li, K.-Q. Wang, and D. Zhang, "A new equation of saturation in RGB-to-HSI conversion for more rapidity of computing," in *in Proceedings. International Conference on Machine Learning and Cybernetics*, pp. 1493–1497, Beijing, China, 2002.

[41] P. Salembier and J. Ruiz, "On filters by reconstruction for size and motion simplification," in *in Proc. of Int. Symposium in*

Mathematical Morphology, pp. 425–434, Orlando, FL, USA, 2002.

[42] Y. H. Liu, M. Muftah, T. Das, L. Bai, K. Robson, and D. Auer, "Classification of MR tumor images based on Gabor wavelet analysis," *Journal of Medical and Biological Engineering*, vol. 32, no. 1, pp. 22–28, 2012.

[43] B. Shabbir, M. Sharif, W. Nisar, M. Yasmin, and S. L. Fernandes, "Automatic cotton wool spots extraction in retinal images using texture segmentation and Gabor wavelet," *Journal of Integrated Design and Process Science*, vol. 20, pp. 65–76, 2016.

[44] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, pp. 786–804, 1979.

[45] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," in *in Matters of intelligence*, pp. 115–141, Springer, 1987.

[46] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, pp. 1517–1525, 2004.

[47] D. G. Stork, R. O. Duda, P. E. Hart, and D. Stork, *Pattern Classification*, A Wiley-Interscience Publication, 2001.

[48] Z. Chen, C. Jiang, S. Xiang, J. Ding, W. Min, and X. Li, "Smartphone sensor-based human activity recognition using feature fusion and maximum full a posteriori," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 3992–4001, 2020.

[49] T. K. Ho, "Nearest neighbors in random subspaces," in *in Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pp. 640–648, Berlin, Heidelberg, 1998.

[50] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, pp. 18–28, 1998.

[51] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *in Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, pp. 441–444, Hong Kong, China, 2006.

[52] G. Goudelis, K. Karpouzis, and S. J. P. R. Kollias, "Exploring trace transform for robust human action recognition," *Pattern Recognition*, vol. 46, no. 12, pp. 3238–3248, 2013.

[53] J. A. Nasiri, N. M. Charkari, and K. J. S. P. Mozafari, "Energy-based model of least squares twin support vector machines for human action recognition," *Signal Processing*, vol. 104, pp. 248–257, 2014.

[54] J. Jiang, X. He, M. Gao, X. Wang, and X. Wu, "Human action recognition via compressive-sensing-based dimensionality reduction," *Optik*, vol. 126, no. 9-10, pp. 882–887, 2015.

[55] Z. Gao, H. Zhang, A. A. Liu, G. Xu, and Y. Xue, "Human action recognition on depth dataset," *Neural Computing and Applications*, vol. 27, no. 7, pp. 2047–2054, 2016.

[56] S. Aly, "An effective human action recognition system based on Zernike moment features," in *in 2019 International Conference on Innovative Trends in Computer Engineering (ITCE)*, pp. 52–57, Aswan, Egypt, 2019.

[57] D. K. Vishwakarma, "A two-fold transformation model for human action recognition using decisive pose," *Cognitive Systems Research*, vol. 61, pp. 1–13, 2020.

[58] L. Shao, X. Zhen, D. Tao, and X. Li, "Spatio-temporal Laplacian pyramid coding for action recognition," *IEEE Transactions on Cybernetics*, vol. 44, no. 6, pp. 817–827, 2014.

[59] M. Sreeraj, "Multi-posture human detection based on hybrid HOG-BO feature," in *in Advances in Computing and Communications (ICACC), 2015 Fifth International Conference on*, pp. 37–40, Kochi, India, 2015.

[60] J. Yang, Z. Ma, and M. Xie, "Action recognition based on multi-scale oriented neighborhood features," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 8, no. 1, pp. 241–254, 2015.

[61] S. Cheng, J. Yang, Z. Ma, and M. Xie, "Action recognition based on spatio-temporal log-Euclidean covariance matrix," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 9, no. 2, pp. 95–106, 2016.

[62] H. Liu, Z. Ju, X. Ji, C. S. Chan, and M. Khoury, "Study of human action recognition based on improved spatio-temporal features," in *in Human Motion Sensing and Recognition*, pp. 233–250, Springer, 2017.

[63] M. Abdellaoui and A. Douik, "Human action recognition in video sequences using deep belief networks," *Traitement du Signal*, vol. 37, pp. 37–44, 2020.

[64] S. Salahuddin, I. Ahmed, M. Rashid, and N. Minallah, "Automatic recognition of human actions," in *in 2019 13th International Conference on Open Source Systems and Technologies (ICOSST)*, pp. 1–6, Lahore, Pakistan, 2019.