Journal of Ambient Intelligence and Humanized Computing (2023) 14:4707–4725
https://doi.org/10.1007/s12652-022-04374-y

ORIGINAL RESEARCH

Multiview human activity recognition using uniform rotation invariant local binary patterns

Swati Nigam[1] · Rajiv Singh[1] · Manoj Kumar Singh[2] · Vivek Kumar Singh[2]

Received: 15 January 2022 / Accepted: 30 July 2022 / Published online: 24 September 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Significant efforts have been made to monitor human activity, although it remains a challenging area for computer vision research. This paper has introduced a framework to identify the most common types of video surveillance activities. The proposed framework consists of three consecutive modules: (i) human detection by background subtraction, (ii) retrieval of uniform and rotation invariant local binary pattern (LBP) feature, and (iii) identification of human activities with a support vector machine (SVM) multiclass classifier. This framework provides a consistent view of the human actions that look at multiple subjects from different views. In addition to this, uniform patterns provide better performance in discriminating human activities. A multiclass SVM is used for classification of human activities. SVM classifier is set and trained to achieve the better efficiency by selecting the appropriate feature before it is integrated. Weizmann's Multiview dataset, CASIA dataset and IXMAS dataset confirm the high efficiency and better robustness of the proposed framework.

Keywords Human activity recognition · Multiview activity recognition · Foreground detection · Rotation invariance · Uniform LBP · Multiclass SVM

1 Introduction

Notable success achieved in the monitoring of human actions allows a variety of advanced multimedia applications (Singh et al. 2020a). Owing to its great importance, human activity recognition is exploited for applications like intelligent video surveillance, abnormal behavior recognition, sports, transportation, web and healthcare. It is found in literatures that computational methods are well capable of recognizing normal and abnormal activities from image and video sequences such as walking, fighting and robbing (Nigam et al. 2019; Sahoo et al. 2020; Rajagopal et al. 2020; Pillai et al. 2021; Yousef et al. 2022; Srivastava et al. 2021). Visual surveillance equipment are used for monitoring human activities and retrieving relevant information. These are further integrated to build advanced systems (Lv et al. 2021a, 2022). Depending on number of persons involved in it, human activities can be of four types—(i) actions involving one person, (ii) interactions involving two people, (iii) interactions involving a person and an object, and (iv) group interactions involving a number of persons.

Even though action recognition is a challenging issue, much work is done on it over the past decade (Binh et al. 2013; Poppe 2010; Weinland et al. 2011; Aggarwal and Ryoo 2011; Ji and Liu 2009). One of the major issues with actions performed in a real 3D environment is that cameras capture only 2D projects of real-time actions. Therefore, visual analysis of the activities performed in image plane is merely a projection of actual real-world actions. This projection is based on viewpoints and does not contain complete information about the action. As a solution to this problem, concept of exploring information obtained from multiple cameras mounted on different viewpoints is used (Ji and Liu 2009). Therefore, the mechanisms are developed by presenting a view independent analysis for multiple views (Ji

✉ Rajiv Singh
jkrajivsingh@gmail.com

Swati Nigam
swatinigam.au@gmail.com

Manoj Kumar Singh
manoj.dstcims@bhu.ac.in

Vivek Kumar Singh
vivek@bhu.ac.in

[1] Department of Computer Science, Banasthali Vidyapith, Banasthali, India

[2] Department of Computer Science, Banaras Hindu University, Varanasi, India

and Liu 2009). Exploring information from multiple scenes enhances the accuracy of activity recognition by obtaining features from various 2D viewpoints to achieve visual consistency. Main objective of the whole scenario is to develop a reliable human activity recognition system.

We participate in the aforementioned solution and introduce an approach for multiview human activity recognition system for image sequences. The proposed framework consists of three steps:

(i) Finding the human objects by removing the background.
(ii) Extraction of uniform and rotation invariant characteristics of LBP.
(iii) Identification of human activities using SVM.

We use simple frame differencing approach for background removal from input data. After background removal, we extract uniform rotation invariant LBP features. Due to rotation invariance characteristic of uniform LBP, it provides an independent analysis of human activity perceptions. This feature is categorized using a radial basis function (RBF) kernel-based SVM classifier with one versus all (OVA) structure. Use of SVM is influenced by the fact that non-sequential strategies, like SVM, are highly competitive and balanced in large-scale and continuous work data (Nigam et al. 2018). Multiclass classification is achieved using hierarchical organization of several binary classifiers.

To illustrate the effectiveness of our proposed work, experiments are performed on three benchmark and publicly available multiview human activity video datasets. These datasets are Weizmann, CASIA and INRIA Xmas Motion Acquisition Sequences (IXMAS). We evaluate the proposed method with the existing and established feature descriptor based methods. The test results of the three datasets show the efficacy of the proposed framework.

Following are the major highlights of the proposed work.

(i) We introduce a rotation invariant human activity recognition framework.
(ii) Multiview human activity recognition is handled with background removal.
(iii) Uniform LBP and SVM classifier are exploited for implementation of the proposed framework.

Organization of this framework is as follows. Section 2 briefly discusses the related works. Section 3 elaborates the development of LBP features and the organization of the SVM classifier for multiple classes. Section 4 provides the implementation details of the proposed method. The results of the evaluation and discussion on the three public datasets are provided in Sect. 5. Section 6 provides conclusive remarks of the study.

2 Related works

Recognition of human activity is the process to detect human body motion patterns. Popular devices to detect human activity are sensors and cameras. Mostly, two type of activity recognition systems do exist, one is sensor-based and other is vision-based. Vision-based systems are more popular as compared to sensor-based since they provide important cues of activity recognition. Many researchers have contributed towards the reviews on human activity recognition (Saha et al. 2022). Based on these reviews, activity recognition approaches can be model-based and model-free.

Model-based activity recognition uses a pre-model to monitor human activities. These vivid 2D and 3D shape models are used to visualize people's activities. Global and local features have been combined in (Wang and Mori 2010) to implement a framework for human action recognition. This work has demonstrated that the combination of part-based model and motion features with large-scale improves the results. Instead of constructing hidden part model, the work in (Wu et al. 2014) has constructed hidden temporal models for each action class. It has focused on human action recognition in uncontrolled videos containing complex temporal structures. The work in (Lan et al. 2011) has focused on the recognition of specific actions and group activities. It has also defined a new feature called action context descriptor. This approach has demonstrated good visual results of several complex but mathematically costly tasks. Cheng et al. (2014) has developed a layered model to represent group activities at diverse granularities. New informative descriptions of the appearance of group actions have been introduced in this way. A nearest neighbour classifier and Gaussian mixture model based work has been proposed for video action recognition using motion curves in (Vrigkas et al. 2014).

However, from the analysis of the model based methods, it is observed that there is a trade-off in retrieving a detailed knowledge of the human body, and the cost of calculation and robustness. The model based methods exploit the pose and velocity vectors which may increase the computational complexity. Sometimes, major parts of body models are taken to reduce the complexity (for example hands, legs, torso etc.), still it is difficult to construct these models. Also, model based approaches need to be implemented directly and could not work in real time.

Model-free approaches overcome shortcomings of model-based approach. In model-free methods, low visibility features from area of interest are retrieved for action recognition. These methods are based on posture, global and local motions (Määttä et al. 2010). The feature based multiple view approaches obtain image data captured by

multiple cameras. A two-camera based method has been implemented for multiple humans pointing in a direction (Matikainen et al. 2011). The different views of 2D pointer configurations have been used to obtain 3D pointing vectors. Five calibrated and synchronized cameras have been used in (Souvenir and Babbs 2008). R transform and manifold learning of the silhouettes have been used for view invariant activity recognition. The circular shift invariance nature of discrete Fourier transform have been exploited in (Iosifidis et al. 2010).

Data fusion has also been exploited for multiview human activity recognition (Weinland et al. 2010). It has used 3D histogram of oriented gradients (HOG) features and applied local partitioning along with hierarchical classification on it. A similar method has been implemented using view point aggregation and multiview dynamic image fusion for cross view 3D action recognition (Wang et al. 2021). It has used 3D characterization and fisher vector for representation of 3D action.

Cross-view activity recognition is an interesting topic for researchers. This is a difficult task of human activity recognition since training and testing views are different. Numerous techniques are proposed for this purpose including learning from short video clips (Vyas et al. 2020), bilayer classification model (Li et al. 2019) and unsupervised attention transfer (Ji et al. 2021).

In recent works, deep learning and transfer learning have become useful tools (Lv et al. 2021b; Singh et al. 2020b). Few deep learning-based techniques are defined in (Jan and Khan 2021; Verma and Singh 2021; Verma et al. 2020) which are very efficient to perform recognition task.

Today, dynamic texture patterns like LBP, have become an obvious choice for the recognition of the activity of a person considered as moving texture patterns. A few examples of them are (Nigam et al. 2021; Kellokumpu et al. 2010, 2011; Vili et al. 2008). However, none of these strategies uses the rotation invariant uniform LBP. Selecting such patterns reduces length of LBP histogram and improves efficacy of a classifier (Pietikäinen et al. 2011). It is widely accepted that uniform LBP is highly effective and has been used repeatedly in several other applications in addition to texture analysis (Bianconi and Fernández 2011). Although many upgraded versions of simple LBP have been introduced, many techniques still benefit from the uniform LBP. However, it is not clear that how the uniform patterns contribute to the LBP based discrimination (Lahdenoja et al. 2013). Furthermore, uniform LBP has been used successfully to obtain rotation invariance (Fernández et al. 2011). The use of uniform binary patterns with rotation invariance is advanced version when compared to its predecessors, as they provide additional integrated representation (Ojala et al. 2002). The global rotation of LBP has been achieved in (Ahonen et al. 2009) using the discrete Fourier transformation in the uniform LBP histograms bins. Rotation invariance characteristic of LBP variants has also been discussed in (Zhao et al. 2011).

From above description of human activity recognition literature, it can be inferred that uniform LBP results in better selection of human multiview activity recognition.

3 Principles and basics

This section briefly discusses two major components of the proposed method, which are the uniform rotation invariant LBP and the multiclass SVM.

3.1 Uniform and rotation invariant LBP

- LBP

The LBP feature is built for a circular neighbourhood of radius R pixel. Intensity of P sample points is compared in the circular neighbourhood with the centre pixel in either clockwise or anticlockwise direction (see Fig. 1).

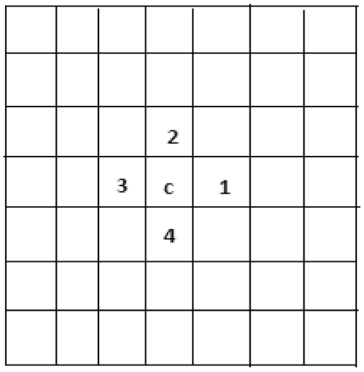
This comparison determines whether the pixel value should be zero (0) or one (1). A value 0 is given if the median pixel magnitude is greater than the neighborhood pixel and a value 1 is given if the median pixel magnitude is less than neighborhood pixel. A popular option is 8 for the number of sample points in the neighborhood and 1 for radius (i.e., $P=8$ and $R=1$). Although, other combinations may also be used. Intensity of a sample point between two pixels is determined by the bilinear interpolation. The LBP feature of an image is denoted by LBP (Pietikäinen et al. 2011). After having extracted the LBP of a pixel, intensity value of the pixel is replaced by this LBP. This procedure could not be followed for border pixels because all of the neighbor values do not exist there. Under these considerations, feature vector of an image is given by

$$LBP_{P,R}(x,y) = \sum_{P=0}^{P-1} s(g_P - g_c)2^P \tag{1}$$

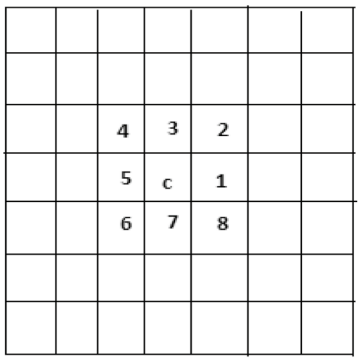
In Eq. (1), (x, y) is the center pixel location, g_c represent the center pixel intensity, g_p represent the pixel intensity of the neighborhood and $s(w)$ is defined as

$$s(w) = \left\{ \begin{array}{l} 1, w \geq 0 \\ 0, w < 0 \end{array} \right\} \tag{2}$$

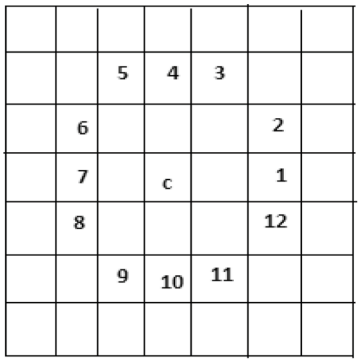
The feature vector $LBP_{P,R}$ of an image is LBP histogram of all pixels in this image. The initial dimension of this LBP histogram is 2^P since each LBP may be assigned to a different bin. If an image has M regions, then total number of histograms formed in the image are $M \cdot 2^P$ or we can say that the image has a histogram whose size equals to $M \cdot 2^P$.



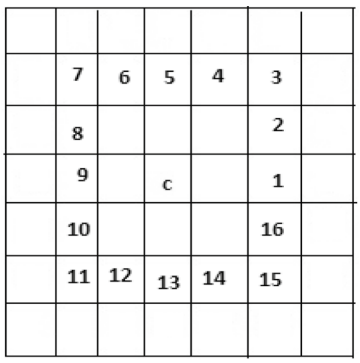
$(P = 4, R = 1)$



$(P = 8, R = 1)$



$(P = 12, R = 2)$



$(P = 16, R = 2)$

Fig. 1 Circular neighborhoods for different (P, R) (anti-clockwise)

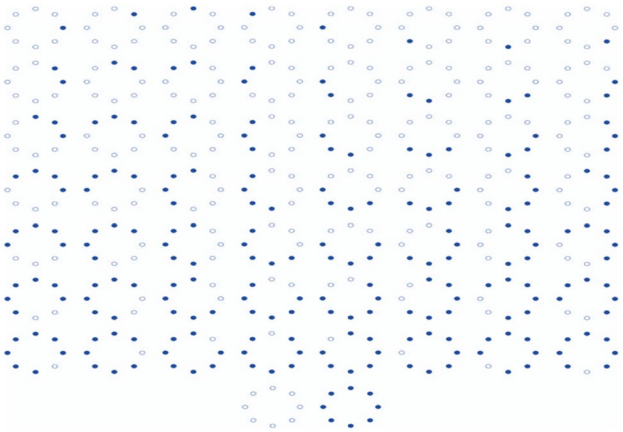


Fig. 2 Uniform LBP ($P=8$, $R=1$). Hollow circle indicates 0 and solid circle indicates 1

- Rotation invariance

Several upgraded versions of LBP, as discussed in (Pietikäinen et al. 2011), have been developed to achieve invariance against rotation and to reduce the size of LBP histogram. For the rotated image, the gray value g_p will shift along the rotation circle perimeter, hence different $LBP_{P,R}$ can be calculated. To reduce the effect of rotation, an upgraded LBP including invariance against rotation is defined as

$$LBP^{ri}_{P,R}(x, y) = min\left\{ROR(LBP_{P,R}, i)|i = 0, 1, ..., R - 1\right\} \quad (3)$$

In Eq. (3), $ROR(LBP_{P,R}, i)$ makes a circular bitwise right shift i times to the R-bit number $LBP_{P,R}$. The $LBP^{ri}_{P,R}$ feature vector can have 36 different values for $R=8$, and it can have the histogram size 36 for an image region.

- Uniform patterns

Uniform LBP is having 0, 1 or 2 circular transitions between binary value 0 and 1. Let us take few examples of uniform and non-uniform patterns. The 0–1 transitions of uniform patterns for $P=8$ and $R=1$ are shown in Fig. 2. In a circular neighborhood of P pixels, number of uniform patterns found is $P+1$. A brief description of uniform and non-uniform patterns is shown in Table 1.

Formal definition of uniform LBP with rotation invariance is given by

$$LBP^{riu2}_{P,R} = \begin{cases} \sum_{P=0}^{P-1} s(g_P - g_C), if\ \mathrm{U}(\mathrm{LBP}_{P,R}) \leq 2 \\ P + 1, \qquad otherwise \end{cases} \quad (4)$$

where *riu2* denotes rotation invariant uniform patterns, g_c represent the center pixel intensity, g_p represent the pixel intensity of the neighborhood, *s(w)* is defined in Eq. (2) and

Table 1 Uniform and non-uniform patterns

Sr. no.	Pattern	Number of transitions	Uniform	Non-uniform
1	00,000,001	1	Yes	No
2	11,111,011	2	Yes	No
3	11,110,100	3	No	Yes
4	00,001,010	4	No	Yes

$$U(LBP_{P,R}) = \left|s(g_{P-1} - g_C) - s(g_0 - g_C)\right| + \sum_{P=1}^{P-1}\left|s(g_P - g_C) - s(g_{P-1} - g_C)\right| \quad (5)$$

3.2 SVM multiclass classifier

Initially, SVM classifier was proposed for binary classification and subsequently used for multiple classification successfully.

- SVM classifier

For a set of L points, where x_i consists of D attributes and belongs to class $y_i = +1\ or\ -1$. Hence, the training set is

$$\{x_i, y_i\}, i = 1.....L, y_i \in \{1, -1\}, x \in R^D \quad (6)$$

For linearly separable data, a hyperplane can be drawn for graphs of $x_1, x_2, \ldots, x_D$ where $D > 2$. SVM provides the hyperplane which is closest from both classes.

The data point x_S is represented by

$$y_s\left(\sum_{m\in S} \alpha_m y_m x_m \cdot x_s + b\right) = 1 \quad (7)$$

where α is called Lagrange Multiplier satisfying $\alpha_i \geq 0 \forall_i$ and S represents the set of indices of support vectors. S is obtained by indices i where $\alpha_i \geq 0$. The support vectors in S are

$$b = \frac{1}{N_s}\sum_{s\in S}(y_s - \sum_{m\in s}\alpha_m y_m x_m \cdot x_s) \quad (8)$$

- One Versus All (OVA) architecture

The OVA architecture is initially used to implement multiclass classification with SVM. The idea behind it is that all classes are divided into two categories. One class is included in positive class at a time and all other classes are included in the negative one. This process is repeatedly used for each and every class. Hence, classifier is applied same number of times as the classes are. Architecture of this classifier is shown in Fig. 3.

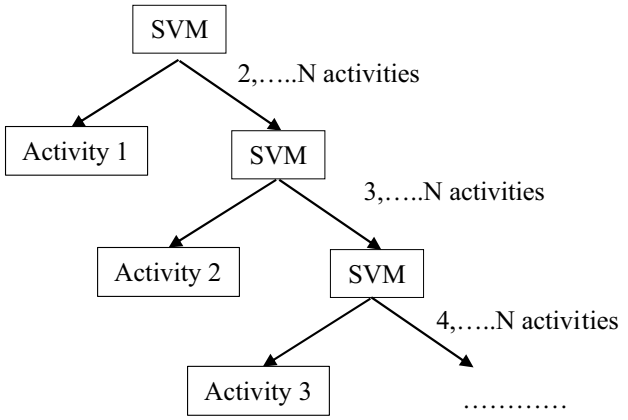


Fig. 3 SVM multiclass classifier with OVA architecture

4 The proposed framework

This section presents proposed framework for human activity recognition system. Figure 4 represents the proposed framework which has been separately discussed in this section.

4.1 Input video

It is a sequence of frames which is used for training purpose. The video can have number of frames from $1, 2, \ldots, n$. This video is represented as

$$V = \sum_{i=1}^{n} F_i \quad (9)$$

where F denotes a particular frame and i is the number of frames which lies among $1, 2, \ldots, n$.

4.2 Preprocessing

The preprocessing step is performed in order to reduce color and size variations and to have all video frames in a consistent format. Therefore, the normalized video is

$$|V| = \sum_{i=1}^{n} |F_i| \quad (10)$$

4.3 Background subtraction

Foreground object detection is an important step in activity recognition process. Background subtraction is a simple yet

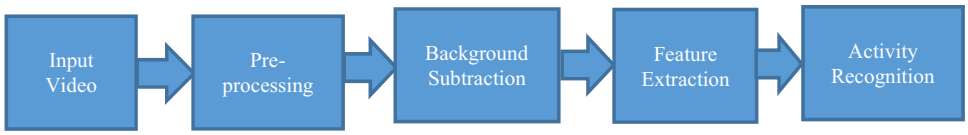


Fig. 4 Block diagram of the proposed framework

efficient technique for this purpose. Therefore, in the proposed framework, foreground object detection is performed using background subtraction approach for capturing the human activities only. Frame differencing is a popular method for this. However, threshold selection is an important step in execution of background subtraction. Few representative results of background subtraction are shown in Fig. 5. A general algorithm for frame differencing based background subtraction is provided below.

input: normalized training video $|V|$
output: normalized video after applying background subtracted $|V_{bs}|$
threshold $\leftarrow th$
background_frame $\leftarrow$ *first frame of the input video* $|F_1|$
for frame_number from 2 to n
current_frame $\leftarrow |F_i|$
difference_frame $\leftarrow$ *current_frame* $-$ *background_frame i.e.* $|F_i| - |F_1|$
if difference_frame $\geqslant th$
then pixel value has to retain as foreground object
otherwise pixel value has to be discarded
end_if
steps are repeated until frame n. Each time background_frame is reinitialized with previous step curret_frame i.e. $|F_1| = |F_i|$
end_algo

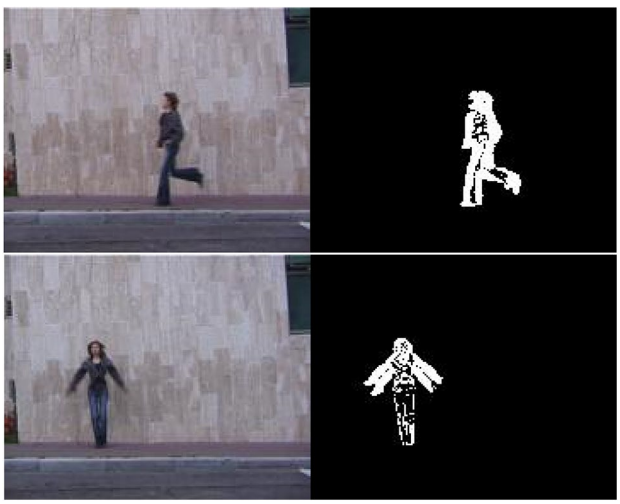


Fig. 5 Original and corresponding background subtracted images

4.4 Feature vector extraction

The next step is to extract the features from the background subtracted video $|V_{bs}|$. This motion feature is computed as following equations

$$LBP^{riu2}_{P,R} = \begin{cases} \sum_{P=0}^{P-1} s(g_P - g_C), \ if \ \mathrm{U}(\mathrm{LBP}_{P,R}) \leq 2 \\ P+1, \qquad otherwise \end{cases} \tag{11}$$

$$\text{where } s(w) = \left\{ \begin{matrix} 1, w \geq 0 \\ 0, w < 0 \end{matrix} \right\}, \tag{12}$$

$$\begin{aligned} U(LBP_{P,R}) = &\left|s(g_{P-1} - g_C) - s(g_0 - g_C)\right| \\ &+ \sum_{P=1}^{P-1} \left|s(g_P - g_C) - s(g_{P-1} - g_C)\right| \end{aligned} \tag{13}$$

g_c= center pixel of $|V_{bs}|$ and g_p= neighborhood pixel of $|V_{bs}|$.

4.5 Recognition of activities

SVM is an efficient classifier that has been extensively used in literature for recognition task. We have used SVM in the proposed framework since it produced high accuracy and performed faster predictions in comparison to other existing classifiers. After feature extraction step, activity recognition step is performed by training and testing of SVM multiclass classifier. Among different kernel functions, RBF kernel is selected owing to its better accuracy. The input video V belongs to specific activity subject to corresponding class

$$V \in \begin{cases} activity_1 \ \ iff \ \ V \in class_1 \\ activity_2 \ \ iff \ \ V \in class_2 \\ \quad \ldots \ldots \ldots \ldots \ldots \ldots \\ activity_n \ \ iff \ \ V \in class_n \end{cases} \tag{14}$$

5 Experimentation and results

This section performs experimentation on 3 benchmark datasets—the Weizmann viewpoint dataset (Gorelick et al. 2007), the CASIA dataset (Wang et al. 2007) and the IXMAS dataset (Weinland et al. 2006). First input video is supplied to the pipeline. After that preprocessing, background subtraction and feature extraction steps have been executed followed by classification. Both qualitative and

objective evaluation of the proposed framework is performed and it is tested in comparison to moment based method (Binh et al. 2013), moment invariant based method (Nigam and Khare 2016), center symmetric LBP based method (Bianconi and Fernández 2011), and GIST feature based method (Vyas et al. 2020).

5.1 Case study 1

This experiment illustrates the efficacy of the proposed framework for various rotations of an activity. For this purpose, we have selected Weizmann viewpoint dataset (Gorelick et al. 2007). It contains 10 videos showing "walking" activity captured from cameras placed at different viewpoints. These viewpoints vary between 0° and 81° angle relative to the image plane with a difference of 9°. With the gradual increase in the angle of the videos, the videos contain significant changes in scale. The videos are of frame size 180×144, image type is 'TrueColor' and no video compression is used. We have selected 40 frames from each video.

We have performed background subtraction followed by feature vector computation and classification. Visual background subtraction results are shown in Fig. 6 and recognition results are shown in Fig. 7.

The angle changes from 0° to 27°, 54°, 63°, 72°, and 81°. From Fig. 6, we get background subtracted images of the activity "walking" being performed at different rotation angles. In this case, we get foreground object segmented correctly. Not much noise is present in the images and one can easily recognize the foreground object in these videos. From Fig. 7, we get visual recognition results for activity "walking" at different rotation angles. From these results, we get correct results for recognizing the activity whether it is being performed at different rotation angles. The angle of rotation has been recognized correctly for 10 different viewing angles. Hence, one can get correct background subtracted images and visual results.

Now, we have shown quantitative results and comparatively evaluated in terms of confusion matrix and recognition accuracy. The compared methods are moment based method (Binh et al. 2013), moment invariant based method (Nigam and Khare 2016), center symmetric LBP based method (Bianconi and Fernández 2011), and GIST feature based method (Vyas et al. 2020). This section shows quantitative results for Weizmann viewpoint dataset (Gorelick et al. 2007). Confusion matrices and recognition accuracy have been shown in Tables 2, 3, 4, 5 and 6. From these tables, it is observed that the our method performs better than other methods.

Recognition accuracy of different methods for WEIZMANN dataset is shown in Fig. 8. From this table, we can observe that recognition accuracy of moment-based method is 10%, invariant moment-based method is 22.7%, CSLBP method is 32.9%, GIST based method is 33.2% and for the proposed method, it is 100%. From this table, we can conclude that the proposed framework outperforms other methods.

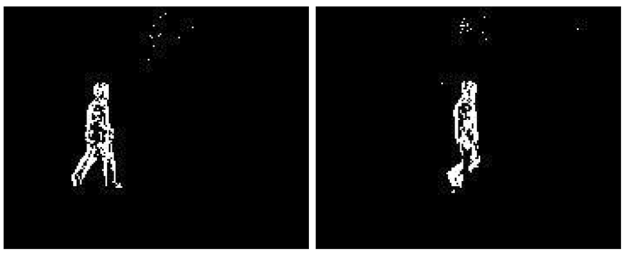


Fig. 6 Background subtraction results

5.2 Case study 2

This section demonstrates performance of the proposed framework for CASIA action recognition dataset (Wang et al. 2007). Five most common interactions have been selected which are fight, overtake, rob, follow, and meet and part. All videos have been captured from three different viewpoints: angular, horizontal and top down. The activities have been renamed as fight: angular view, fight: horizontal view, fight: topdown view, overtake: angular view, overtake: horizontal view, overtake: topdown view, rob: angular view, rob: horizontal view, rob: topdown view, follow: angular view, follow: horizontal view, follow: topdown view, meet and part: angular view, meet and part: horizontal view, and meet and part: topdown view. Videos have been compressed using HUFFYUV compression technique in AVI video format. Duration varies between 5–30 s for different activities. Visual background subtraction and recognition results are shown in Figs. 9 and 10, respectively.

Figures 9 and 10 show that we get correct background subtracted frames and visual results. The activities have been recorded using three different viewing angles. The background removal step of the proposed method provides better results which lead to correct object identification. As a result, we have better recognition of human activities and the proposed method is able to identify activities for multiple views. The proposed method also identifies suspicious activities which are fight, overtake, rob, follow, and meet and part.

Now, CASIA dataset have been analyzed and presented (Wang et al. 2007). The activities are renamed as; Fight: angular view as A1, fight: horizontal view as A2, fight: topdown view as A3, overtake: angular view as B1, overtake: horizontal view as B2, overtake: topdown view as B3, rob: angular view as C1, rob: horizontal view as C2, rob: topdown view as C3, followalways: angular view as D1, followalways: horizontal view as D2, followalways: topdown view as D3, meetapart: angular view as E1,

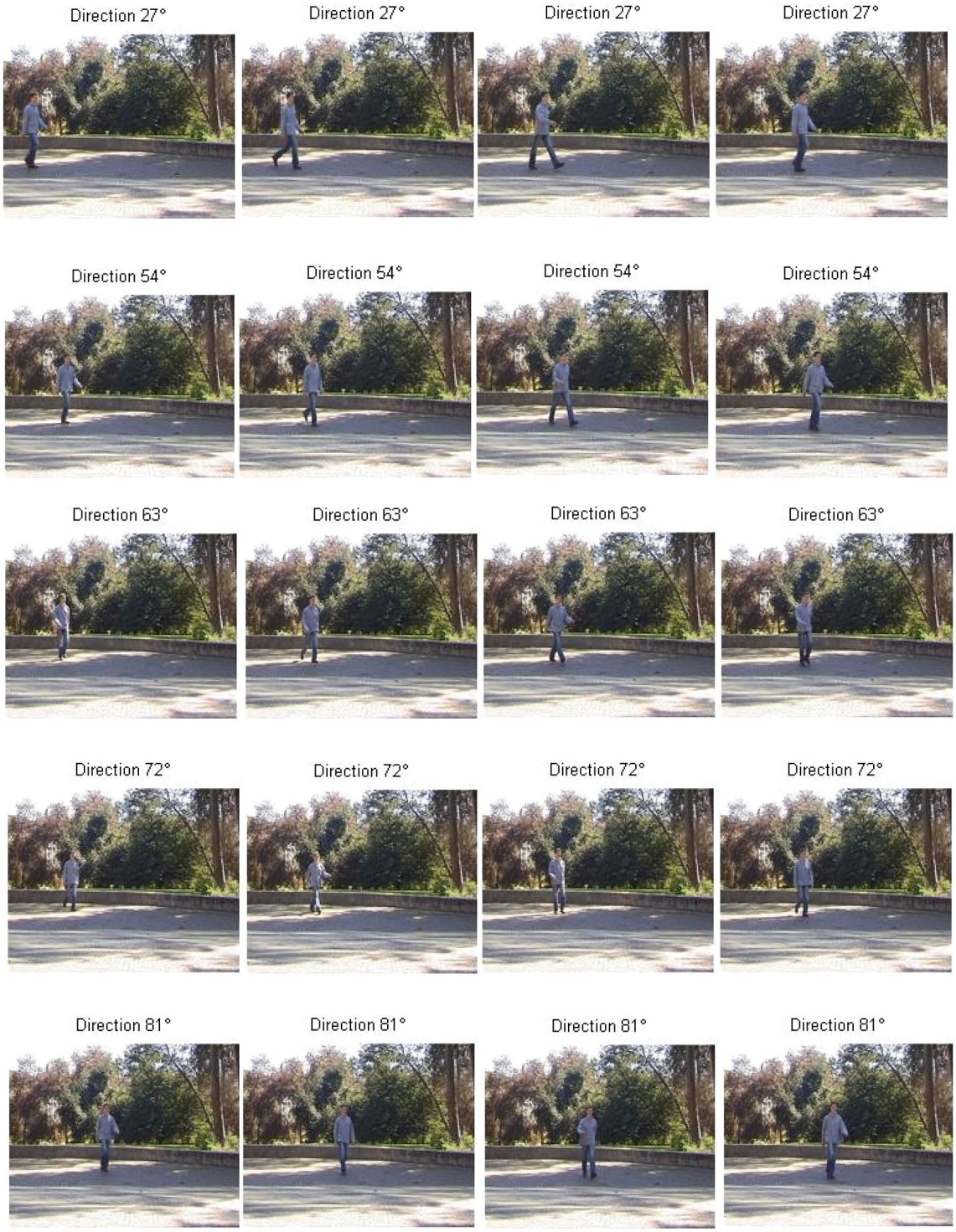


Fig. 7 Visual recognition results

Table 2 Confusion matrix for moment based method (Binh et al. 2013)

Walk	Dir 0°	Dir 9°	Dir 18°	Dir 27°	Dir 36°	Dir 45°	Dir 54°	Dir 63°	Dir 72°	Dir 81°
Dir 0°	0	0	0	0	0	0	0	0	0	1.00
Dir 9°	0	0	0	0	0	0	0	0	0	1.00
Dir 18°	0	0	0	0	0	0	0	0	0	1.00
Dir 27°	0	0	0	0	0	0	0	0	0	1.00
Dir 36°	0	0	0	0	0	0	0	0	0	1.00
Dir 45°	0	0	0	0	0	0	0	0	0	1.00
Dir 54°	0	0	0	0	0	0	0	0	0	1.00
Dir 63°	0	0	0	0	0	0	0	0	0	1.00
Dir 72°	0	0	0	0	0	0	0	0	0	1.00
Dir 81°	0	0	0	0	0	0	0	0	0	1.00

Table 3 Confusion matrix for inv-mom based method (Nigam and Khare 2016)

Walk	Dir 0°	Dir 9°	Dir 18°	Dir 27°	Dir 36°	Dir 45°	Dir 54°	Dir 63°	Dir 72°	Dir 81°
Dir 0°	0.97	0.03	0	0	0	0	0	0	0	0
Dir 9°	0.72	0.20	0.08	0	0	0	0	0	0	0
Dir 18°	0.39	0.22	0.39	0	0	0	0	0	0	0
Dir 27°	0.03	0.15	0.77	0.05	0	0	0	0	0	0
Dir 36°	0.62	0.23	0.15	0	0	0	0	0	0	0
Dir 45°	0.80	0.05	0.15	0	0	0	0	0	0	0
Dir 54°	0.26	0.10	0.64	0	0	0	0	0	0	0
Dir 63°	0	0	0.36	0.31	0	0	0	0.33	0	0
Dir 72°	0.08	0.13	0.65	0.07	0	0	0	0.07	0	0
Dir 81°	0	0	0	0.44	0	0	0.03	0.20	0	0.33

Table 4 Confusion matrix for CSLBP based method (Bianconi and Fernández 2011)

Walk	Dir 0°	Dir 9°	Dir 18°	Dir 27°	Dir 36°	Dir 45°	Dir 54°	Dir 63°	Dir 72°	Dir 81°
Dir 0°	0.97	0	0.03	0	0	0	0	0	0	0
Dir 9°	0.58	0.33	0.03	0.03	0.03	0	0	0	0	0
Dir 18°	0.56	0.15	0.26	0	0	0	0	0	0	0.03
Dir 27°	0.02	0.10	0.49	0.26	0	0.07	0.03	0.03	0	0
Dir 36°	0.13	0.23	0.10	0.15	0.26	0.05	0	0	0	0.08
Dir 45°	0.05	0.08	0.03	0.18	0.33	0.20	0.03	0.03	0	0.07
Dir 54°	0	0.03	0	0.36	0.18	0.10	0.30	0	0	0.03
Dir 63°	0.02	0	0	0.08	0.08	0.08	0.28	0.46	0	0
Dir 72°	0.03	0	0	0	0.05	0.15	0.18	0.44	0.15	0
Dir 81°	0	0	0	0.08	0.02	0.08	0.22	0.45	0.05	0.10

meetapart: horizontal view as E2 and meetapart: topdown view as E3. The results have been shown in Tables 7, 8, 9, 10 and 11. Although, CSLBP based method (Bianconi and Fernández 2011) shows a high recognition rate but proposed framework performs better than this method also.

Recognition accuracy of different methods for CASIA dataset is shown in Fig. 11. From this figure, we can observe that recognition accuracy of moment-based method is 6.7%, invariant moment-based method is 10.3%, CSLBP method is 83.7%, GIST based method is 26.7% and for the proposed method, it is 90.7%. Although, performance of the CSLBP is quite comparable to the proposed one but when we consider overall accuracy of all three cases then we can conclude that the proposed framework is better than the CSLBP based method.

Table 5 Confusion matrix for GIST based method (Vyas et al. 2020)

Walk	Dir 0°	Dir 9°	Dir 18°	Dir 27°	Dir 36°	Dir 45°	Dir 54°	Dir 63°	Dir 72°	Dir 81°
Dir 0°	1.00	0	0	0	0	0	0	0	0	0
Dir 9°	0	0	0	0	0	0	0	0	0	1.00
Dir 18°	0	0	0	0	0	0	0	0	0	1.00
Dir 27°	0	0	0	1.00	0	0	0	0	0	0
Dir 36°	0	0	0	0	0.32	0	0	0	0	0.68
Dir 45°	0	0	0	0	0	0	0	0	0	1.00
Dir 54°	0	0	0	0	0	0	0	0	0	1.00
Dir 63°	0	0	0	0	0	0	0	0	0	1.00
Dir 72°	0	0	0	0	0	0	0	0	0	1.00
Dir 81°	0	0	0	0	0	0	0	0	0	1.00

Table 6 Confusion matrix for the proposed method

Walk	Dir 0°	Dir 9°	Dir 18°	Dir 27°	Dir 36°	Dir 45°	Dir 54°	Dir 63°	Dir 72°	Dir 81°
Dir 0°	1.0	0	0	0	0	0	0	0	0	0
Dir 9°	0	1.0	0	0	0	0	0	0	0	0
Dir 18°	0	0	1.0	0	0	0	0	0	0	0
Dir 27°	0	0	0	1.0	0	0	0	0	0	0
Dir 36°	0	0	0	0	1.0	0	0	0	0	0
Dir 45°	0	0	0	0	0	1.0	0	0	0	0
Dir 54°	0	0	0	0	0	0	1.0	0	0	0
Dir 63°	0	0	0	0	0	0	0	1.0	0	0
Dir 72°	0	0	0	0	0	0	0	0	1.0	0
Dir 81°	0	0	0	0	0	0	0	0	0	1.0

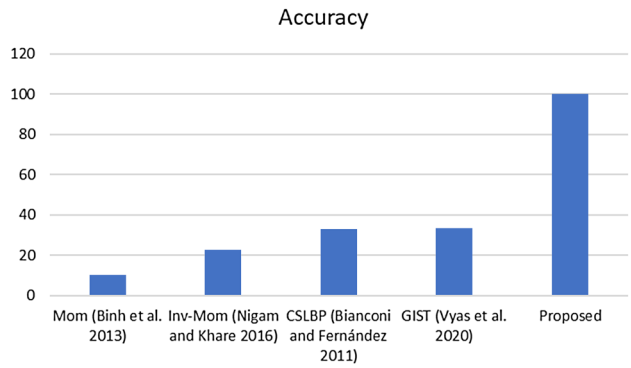


Fig. 8 Recognition accuracy of different methods

5.3 Case study 3

Now, we have selected INRIA IXMAS multiview activity dataset (Weinland et al. 2006). This includes 13 daily-life activities performed by 11 actors 3 times each. These activities are kick, punch, turn-around, check watch, cross arms, scratch head, sit down, get up, walk, wave, point, pick up, throw (overhead), throw (from bottom up) and nothing captured from five calibrated cameras. Qualitative background subtraction and recognition results are shown in Figs. 12 and

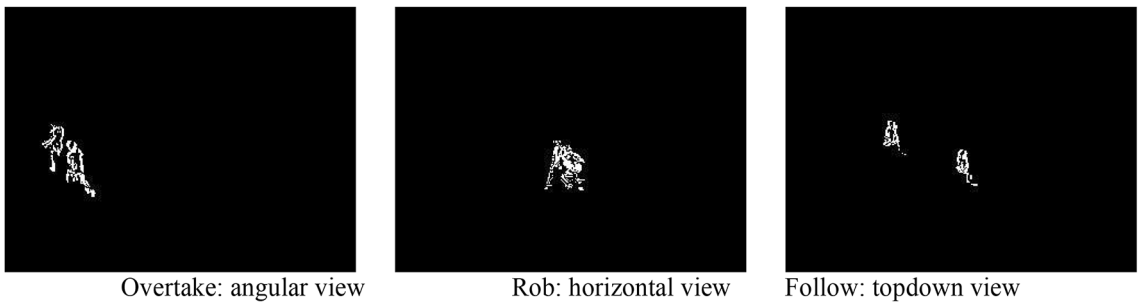


Fig. 9 Background subtraction results

Fig. 10 Visual recognition results

Table 7 Confusion matrix for moment based method (Binh et al. 2013)

Activities	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	E1	E2	E3
A1	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
A2	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
A3	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
B1	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
B2	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
B3	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
C1	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
C2	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
C3	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
D1	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
D3	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
E1	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
E2	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
E3	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0

Table 8 Confusion matrix for inv-mom based method (Nigam and Khare 2016)

Activities	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	E1	E2	E3
A1	0.96	0	0	0.01	0	0	0.01	0.02	0	0	0	0	0	0	0
A2	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B1	0.84	0	0	0.11	0	0	0.05	0	0	0	0	0	0	0	0
B2	0.86	0	0	0.08	0	0	0.06	0	0	0	0	0	0	0	0
B3	0.80	0	0	0.04	0	0.14	0.02	0	0	0	0	0	0	0	0
C1	0.77	0	0	0.01	0.01	0.01	0.16	0.04	0	0	0	0	0	0	0
C2	0.82	0	0	0	0.02	0	0.06	0.10	0	0	0	0	0	0	0
C3	0.76	0	0	0.06	0	0	0.17	0.01	0	0	0	0	0	0	0
D1	0.83	0	0	0.01	0.04	0	0.11	0.01	0	0	0	0	0	0	0
D2	0.86	0	0	0.05	0.04	0	0.05	0	0	0	0	0	0	0	0
D3	0.78	0	0	0.18	0	0.02	0.02	0	0	0	0	0	0	0	0
E1	0.50	0	0	0.01	0	0.02	0.34	0.02	0	0	0	0	0.05	0	0.06
E2	0.94	0	0	0	0	0	0.04	0.02	0	0	0	0	0	0	0
E3	0.88	0	0	0.01	0	0	0.04	0.03	0	0	0	0	0.01	0	0.03

13. From Fig. 12, we see that the foreground objects have been segmented properly. In Fig. 13, we show the obtained visual results for the proposed method.

Now, objective evaluation results are shown for IXMAS dataset (Weinland et al. 2006). Objective evaluation results have been listed in Tables 12, 13, 14, 15 and 16 which infer that the proposed framework is better than other methods.

Recognition accuracy of different methods for IXMAS dataset is shown in Fig. 14. From this figure, we can observe that recognition accuracy of moment-based method is 6.7%, invariant moment-based method is 11.2%, CSLBP method is 83.7%, GIST based method is 100% and for the proposed method, it is 90.7%. Although, performance of the GIST based method (Wang et al. 2021) is quite comparable to the proposed one but when we consider overall accuracy of all three cases then we can conclude that the proposed framework is better than the GIST based method.

In addition to these promising results, these experiments have some limitations also. All these experiments have been conducted on 70% training and 30% testing datasets.

Table 9 Confusion matrix for CSLBP based method (Bianconi and Fernández 2011)

Activities	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	E1	E2	E3
A1	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A2	0.05	0.95	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	0	0	1.00	0	0	0	0	0	0	0	0	0	0	0	0
B1	0.04	0	0	0.96	0	0	0	0	0	0	0	0	0	0	0
B2	0	0	0	0	1.00	0	0	0	0	0	0	0	0	0	0
B3	0	0	0	0	0	1.00	0	0	0	0	0	0	0	0	0
C1	0.02	0	0.10	0	0	0	0.88	0	0	0	0	0	0	0	0
C2	0	0.01	0.03	0	0.08	0.02	0	0.86	0	0	0	0	0	0	0
C3	0	0	0.03	0	0	0.22	0	0	0.75	0	0	0	0	0	0
D1	0	0	0	0.05	0	0	0	0	0	0.95	0	0	0	0	0
D2	0	0	0	0.02	0.11	0	0	0.01	0	0	0.86	0	0	0	0
D3	0	0	0	0	0	0.16	0	0	0.03	0	0	0.81	0	0	0
E1	0.02	0.01	0.17	0.01	0	0	0.03	0	0	0.09	0	0	0.67	0	0
E2	0	0.10	0.11	0	0.04	0	0.01	0.02	0	0	0.08	0	0	0.64	0
E3	0	0	0.45	0	0	0.20	0	0	0	0	0	0.12	0	0.01	0.22

Table 10 Confusion matrix for GIST based method (Vyas et al. 2020)

Activities	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	E1	E2	E3
A1	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A2	0	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	0	0	1.00	0	0	0	0	0	0	0	0	0	0	0	0
B1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
B2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
B3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
C1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
C2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
C3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
D1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
D2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
D3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
E1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
E2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
E3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00

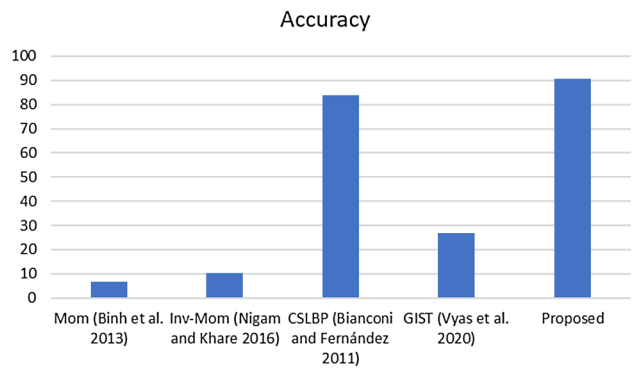


Fig. 11 Recognition accuracy of different methods

Number of samples in these sets can affect overall accuracy of the framework. Also, these experiments have been performed for training and testing sets from the same dataset. Hence, cross-dataset selection may also affect the accuracy of the method.

6 Conclusions

This study proposed an activity recognition system which is suitable for multiple views. Overall framework of the system consists of background subtraction, feature extraction and activity recognition. Background subtraction is used to

Table 11 Confusion matrix for the proposed method

Activities	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	E1	E2	E3
A1	9.0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0
A2	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0	1.0
A3	0	0	9.0	0	0	0	0	0	0	0	0	0	0	0	1.0
B1	0	0	0	9.0	0	0	0	0	0	0	0	0	0	0	1.0
B2	0	0	0	0	9.0	0	0	0	0	0	0	0	0	0	1.0
B3	0	0	0	0	0	9.0	0	0	0	0	0	0	0	0	1.0
C1	0	0	0	0	0	0	9.0	0	0	0	0	0	0	0	1.0
C2	0	0	0	0	0	0	0	9.0	0	0	0	0	0	0	1.0
C3	0	0	0	0	0	0	0	0	9.0	0	0	0	0	0	1.0
D1	0	0	0	0	0	0	0	0	0	9.0	0	0	0	0	1.0
D2	0	0	0	0	0	0	0	0	0	0	9.0	0	0	0	1.0
D3	0	0	0	0	0	0	0	0	0	0	0	9.0	0	0	1.0
E1	0	0	0	0	0	0	0	0	0	0	0	0	9.0	0	1.0
E2	0	0	0	0	0	0	0	0	0	0	0	0	0	9.0	1.0
E3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10.0

Fig. 12 Background subtraction results for Kick activity

Table 12 Confusion matrix for moment based method (Srivastava et al. 2021)

Activities	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5
A1	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A2	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A4	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A5	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B1	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B2	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B3	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B4	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B5	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C2	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C3	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C4	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C5	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

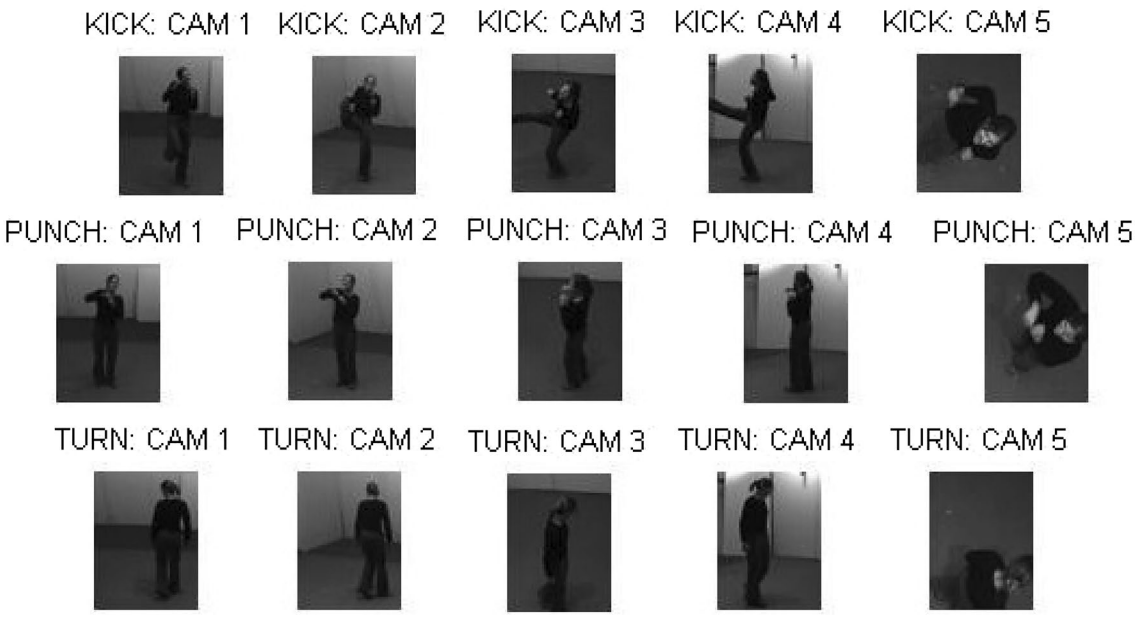


Fig. 13 Visual recognition results

Table 13 Confusion matrix for inv-mom based method (Jan and Khan 2021)

Activities	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5
A1	0.90	0.02	0	0.06	0	0.02	0	0	0	0	0	0	0	0	0
A2	0.80	0.05	0	0.09	0	0	0	0	0	0	0.06	0	0	0	0
A3	0.81	0.05	0.05	0.02	0	0.02	0	0	0	0	0.05	0	0	0	0
A4	0.54	0.07	0	0.37	0	0	0	0	0	0	0.02	0	0	0	0
A5	0.05	0.00	0	0.95	0	0	0	0	0	0	0	0	0	0	0
B1	0.53	0.02	0	0.35	0	0.06	0	0	0	0	0.04	0	0	0	0
B2	0.53	0.02	0	0.40	0	0.05	0	0	0	0	0	0	0	0	0
B3	0.86	0	0	0.07	0	0	0	0.04	0	0	0.03	0	0	0	0
B4	0.68	0.09	0	0.20	0	0	0	0	0	0	0.03	0	0	0	0
B5	0.12	0.02	0	0.86	0	0	0	0	0	0	0	0	0	0	0
C1	0.55	0.06	0	0.02	0	0.22	0	0	0	0	0.15	0	0	0	0
C2	0.80	0	0.02	0.02	0	0.06	0	0.06	0	0	0	0.04	0	0	0
C3	0.37	0.02	0.11	0.05	0	0.26	0	0	0	0	0.17	0	0.02	0	0
C4	0.91	0.05	0	0.02	0	0	0	0	0	0	0.02	0	0	0	0
C5	0.08	0.06	0	0.86	0	0	0	0	0	0	0	0	0	0	0

Table 14 Confusion matrix for CSLBP based method (Bianconi and Fernández 2011)

Activities	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5
A1	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A2	0.05	0.95	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	0.02	0.05	0.93	0	0	0	0	0	0	0	0	0	0	0	0
A4	0.02	0	0.03	0.95	0	0	0	0	0	0	0	0	0	0	0
A5	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
B1	0.07	0.08	0	0.03	0	0.82	0	0	0	0	0	0	0	0	0
B2	0.03	0.08	0.06	0	0	0.31	0.52	0	0	0	0	0	0	0	0
B3	0	0.05	0.10	0	0	0.03	0.07	0.75	0	0	0	0	0	0	0
B4	0.02	0.02	0	0.07	0	0.03	0	0	0.86	0	0	0	0	0	0
B5	0	0	0	0	0	0	0	0	0	1.0	0	0	0	0	0
C1	0.02	0	0	0.06	0	0.07	0.02	0	0.05	0	0.78	0	0	0	0
C2	0.02	0	0.08	0.03	0	0.03	0.03	0.03	0	0	0.12	0.66	0	0	0
C3	0.05	0.07	0.03	0	0	0.07	0.05	0.05	0	0	0	0	0.68	0	0
C4	0.02	0	0	0	0	0.03	0	0	0.16	0	0.03	0.10	0	0.66	0
C5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0

Table 15 Confusion matrix for GIST based method (Wang et al. 2021)

Activities	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5
A1	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A2	0	1.0	0	0	0	0	0	0	0	0	0	0	0	0	0
A3	0	0	1.0	0	0	0	0	0	0	0	0	0	0	0	0
A4	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0	0
A5	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0	0
B1	0	0	0	0	0	1.0	0	0	0	0	0	0	0	0	0
B2	0	0	0	0	0	0	1.0	0	0	0	0	0	0	0	0
B3	0	0	0	0	0	0	0	1.0	0	0	0	0	0	0	0
B4	0	0	0	0	0	0	0	0	1.0	0	0	0	0	0	0
B5	0	0	0	0	0	0	0	0	0	1.0	0	0	0	0	0
C1	0	0	0	0	0	0	0	0	0	0	1.0	0	0	0	0
C2	0	0	0	0	0	0	0	0	0	0	0	1.0	0	0	0
C3	0	0	0	0	0	0	0	0	0	0	0	0	1.0	0	0
C4	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0	0
C5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0

Table 16 Confusion matrix for the proposed method

Activities	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5
A1	9.0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0
A2	0	9.0	0	0	0	0	0	0	0	0	0	0	0	0	1.0
A3	0	0	9.0	0	0	0	0	0	0	0	0	0	0	0	1.0
A4	0	0	0	9.0	0	0	0	0	0	0	0	0	0	0	1.0
A5	0	0	0	0	9.0	0	0	0	0	0	0	0	0	0	1.0
B1	0	0	0	0	0	9.0	0	0	0	0	0	0	0	0	1.0
B2	0	0	0	0	0	0	9.0	0	0	0	0	0	0	0	1.0
B3	0	0	0	0	0	0	0	9.0	0	0	0	0	0	0	1.0
B4	0	0	0	0	0	0	0	0	9.0	0	0	0	0	0	1.0
B5	0	0	0	0	0	0	0	0	0	9.0	0	0	0	0	1.0
C1	0	0	0	0	0	0	0	0	0	0	9.0	0	0	0	1.0
C2	0	0	0	0	0	0	0	0	0	0	0	9.0	0	0	1.0
C3	0	0	0	0	0	0	0	0	0	0	0	0	9.0	0	1.0
C4	0	0	0	0	0	0	0	0	0	0	0	0	0	9.0	1.0
C5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10.0

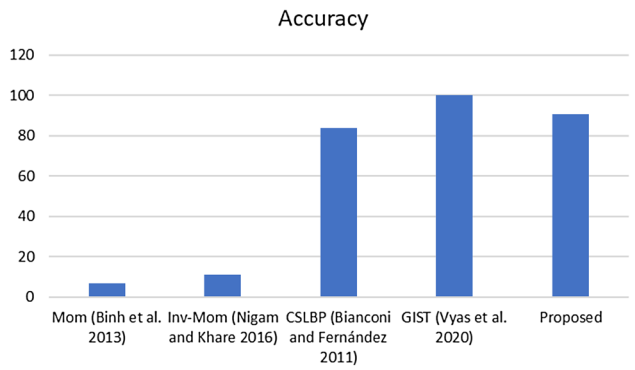


Fig. 14 Recognition accuracy of different methods

capture the object only. Then, uniform and rotation invariant LBP descriptor is computed. Multiclass SVM classifier has been applied for multiclass recognition of different activities. The framework is tested on three multiview human activity video datasets: Weizmann viewpoint dataset (Gorelick et al. 2007), CASIA action recognition dataset (Wang et al. 2007) and IXMAS dataset (Weinland et al. 2006) and compared with (Binh et al. 2013; Vyas et al. 2020; Nigam and Khare 2016; Bianconi and Fernández 2011). This comparison shows that the proposed method outperforms other feature descriptor based methods. This work can be extended for the development of context aware activity recognition which may include multiview and cross-view 2D and 3D human activities. Also, the fusion of existing feature descriptors with machine and deep learning can be done for better representation and recognition of multiview human activities.

Acknowledgements "This work was supported in part by the Ministry of Electronics and Information Technology (MeitY), Government of India under Grant No. 3(9)/2021-EG-II."

Declarations

Conflict of interest The authors declare that there is no conflict of interest associated with the manuscript and data associated will be provided on a reasonable request.

References

Aggarwal JK, Ryoo MS (2011) Human activity analysis: a review. ACM Comput Surv 43(3):1–43

Ahonen T, Matas J, He C, Pietikäinen M (2009). Rotation invariant image description with local binary pattern histogram fourier features. In: Scandinavian conference on image analysis. Springer, Berlin, Heidelberg, pp 61–70

Bianconi F, Fernández A (2011) On the occurrence probability of local binary patterns: a theoretical study. J Mathematical Imaging Vis 40(3):259–268

Binh NT, Nigam S, Khare A (2013) Towards classification based human activity recognition in video sequences. International Conference on Context-Aware Systems and Applications. Springer, Cham, pp 209–218

Cheng Z, Qin L, Huang Q, Yan S, Tian Q (2014) Recognizing human group action by layered model with multiple cues. Neurocomputing 136:124–135

Fernández A, Ghita O, González E, Bianconi F, Whelan PF (2011) Evaluation of robustness against rotation of LBP, CCR and ILBP features in granite texture classification. Mach Vis Appl 22(6):913–926

Gorelick L, Blank M, Shechtman E, Irani M, Basri R (2007) Actions as space-time shapes. IEEE Trans Pattern Anal Mach Intell 29(12):2247–2253

Iosifidis A, Nikolaidis N, Pitas I (2010) Movement recognition exploiting multi-view information. In: 2010 IEEE International Workshop on Multimedia Signal Processing, IEEE, pp 427–431

Jan A, Khan GM (2021) Real world anomalous scene detection and classification using multilayer deep neural networks. Int J Interactive Multimedia Artif Intell. https://doi.org/10.9781/ijimai.2021.10.010

Ji X, Liu H (2009) Advances in view-invariant human motion analysis: a review. IEEE Trans Syst Man Cybern Part C (Appl Rev) 40(1):13–24

Ji Y, Yang Y, Shen HT, Harada T (2021) View-invariant action recognition via Unsupervised AttentioN Transfer (UANT). Pattern Recogn 113:107807

Kellokumpu V, Zhao G, Pietikäinen M (2011) Recognition of human actions using texture descriptors. Mach Vis Appl 22(5):767–780

Kellokumpu V, Zhao G, Pietikäinen M (2010) Dynamic textures for human movement recognition. In: Proceedings of the ACM International Conference on Image and Video Retrieval, pp 470–476

Lahdenoja O, Poikonen J, Laiho M (2013) Towards understanding the formation of uniform local binary patterns. International Scholarly Research Notices

Lan T, Wang Y, Yang W, Robinovitch SN, Mori G (2011) Discriminative latent models for recognizing contextual group activities. IEEE Trans Pattern Anal Mach Intell 34(8):1549–1562

Li Y, Xu X, Xu J, Du E (2019) Bilayer model for cross-view human action recognition based on transfer learning. J Electron Imaging 28(3):033016

Lv Z, Qiao L, Singh AK, Wang Q (2021a) AI-empowered IoT security for smart cities. ACM Trans Internet Technol 21(4):1–21

Lv Z, Qiao L, Singh AK, Wang Q (2021b) Fine-grained visual computing based on deep learning. ACM Trans Multimedia Comput Commun Appl 17(1s):1–19

Lv Z, Guo J, Singh AK, Lv H (2022) Digital twins based VR simulation for accident prevention of intelligent vehicle. IEEE Trans Vehicular Tech 71(4):3414–3428

Määttä T, Härmä A, Aghajan H (2010) On efficient use of multi-view data for activity recognition. In: Proceedings of the Fourth ACM/IEEE International Conference on Distributed Smart Cameras, pp 158–165

Matikainen P, Pillai P, Mummert L, Sukthankar R, Hebert M (2011) Prop-free pointing detection in dynamic cluttered environments. In 2011 IEEE International Conference on Automatic Face and Gesture Recognition (FG), IEEE, pp 374–381

Nigam S, Khare A (2016) Integration of moment invariants and uniform local binary patterns for human activity recognition in video sequences. Multimedia Tools Appl 75(24):17303–17332

Nigam S, Singh R, Misra AK (2018) Efficient facial expression recognition using histogram of oriented gradients in wavelet domain. Multimedia Tools Appl 77(21):28725–28747

Nigam S, Singh R, Misra AK (2019) A review of computational approaches for human behavior detection. Arch Comput Methods Eng 26(4):831–863

Nigam S, Singh R, Singh MK, Singh VK (2021) Multiple views based recognition of human activities using uniform patterns. In: 2021 Sixth International Conference on Image Information Processing (ICIIP), IEEE vol. 6, pp 483–488

Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987

Pietikäinen M, Hadid A, Zhao G, Ahonen T (2011) Computer vision using local binary patterns, vol 40. Springer Science and Business Media

Pillai MS, Chaudhary G, Khari M, Crespo RG (2021) Real-time image enhancement for an automatic automobile accident detection through CCTV using deep learning. Soft Comput 25(18):11929–11940

Poppe R (2010) A survey on vision-based human action recognition. Image Vis Comput 28(6):976–990

Rajagopal A, Joshi GP, Ramachandran A, Subhalakshmi RT, Khari M, Jha S, You J (2020) A deep learning model based on multi-objective particle swarm optimization for scene classification in unmanned aerial vehicles. IEEE Access 8:135383–135393

Saha A, Rajak S, Saha J, Chowdhury C (2022) A survey of machine learning and meta-heuristics approaches for sensor-based human activity recognition systems. J Ambient Intell Hum Comput. https://doi.org/10.1007/s12652-022-03870-5

Sahoo KS, Tripathy BK, Naik K, Ramasubbareddy S, Balusamy B, Khari M, Burgos D (2020) An evolutionary SVM model for DDOS attack detection in software defined networks. IEEE Access 8:132502–132513

Singh R, Nigam S, Singh AK, Elhoseny M (2020a) Intelligent wavelet based techniques for advanced multimedia applications. Springer, pp 1–144

Singh R, Ahmed T, Kumar A, Singh AK, Pandey AK, Singh SK (2020b) Imbalanced breast cancer classification using transfer learning. IEEE/ACM Trans Comput Biol Bioinf 18(1):83–93

Souvenir R, Babbs J (2008) Learning the viewpoint manifold for action recognition. In 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp 1–7

Srivastava S, Khari M, Crespo RG, Chaudhary G, Arora P (eds) (2021) Concepts and real-time applications of deep learning. Springer International Publishing

Verma KK, Singh BM (2021) Deep multi-model fusion for human activity recognition using evolutionary algorithms. Int J Interact Multimedia Artif Intell 7(2):44

Verma KK, Singh BM, Mandoria HL, Chauhan P (2020) Two-stage human activity recognition using 2DConvNet. Int J Interactive Multimedia Artif Intell. https://doi.org/10.9781/ijimai.2020.04.002

Vili K, Guoying Z, Matti P (2008) Texture based description of movements for activity analysis. In International Conference on Computer Vision Theory and Applications (VISAPP 2008), vol 1, pp 206–213

Vrigkas M, Karavasilis V, Nikou C, Kakadiaris IA (2014) Matching mixtures of curves for human action recognition. Comput Vis Image Underst 119:27–40

Vyas S, Rawat YS, Shah M (2020) Multi-view action recognition using cross-view video prediction. In European Conference on Computer Vision. Springer, Cham, pp 427–444

Wang Y, Mori G (2010) Hidden part models for human action recognition: Probabilistic versus max margin. IEEE Trans Pattern Anal Mach Intell 33(7):1310–1323

Wang Y, Xiao Y, Lu J, Tan B, Cao Z, Zhang Z, Zhou JT (2021) Discriminative multi-view dynamic image fusion for cross-View 3-D action recognition. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2021.3070179

Wang Y, Huang K, Tan T (2007) Human activity recognition based on r transform. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp 1–8

Weinland D, Ronfard R, Boyer E (2006) Free viewpoint action recognition using motion history volumes. Comput Vis Image Underst 104(2–3):249–257

Weinland D, Ronfard R, Boyer E (2011) A survey of vision-based methods for action representation, segmentation and recognition. Comput Vis Image Underst 115(2):224–241

Weinland D, Özuysal M, Fua P (2010) Making action recognition robust to occlusions and viewpoint changes. In: European Conference on Computer Vision. Springer, Berlin, Heidelberg, pp 635–648

Wu J, Hu D, Chen F (2014) Action recognition by hidden temporal models. Vis Comput 30(12):1395–1404

Yousef R, Gupta G, Yousef N, Khari M (2022) A holistic overview of deep learning approach in medical imaging. Multimedia Syst 28(3):881–914

Zhao G, Ahonen T, Matas J, Pietikainen M (2011) Rotation-invariant image and video description with local binary pattern features. IEEE Trans Image Process 21(4):1465–1477

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.