International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Human Action Recognition using Rule based Fuzzy Motion Feature Templates

Chandra Mani Sharma, Alaknanda Ashok, Alok Kumar Singh Kushwaha

Abstract: ***This paper proposes a technique for human activity recognition in a video stream. To achieve high accuracy in activity recognition results, the method in its initial step deploys temporal template matching to recognize activities. As temporal templates are susceptible to get affected by speed, style and performance pattern of activity, so it becomes difficult to accurately differentiate among closely similar activities (e.g walking, running and jogging). The confusion in recognizing activities is reconciled by subsequent rule based activity distinction. The proposed method recognizes the human activities in video on various bench-marked data sets including KTH Dataset and Weizmann Dataset. Experimental results demonstrate the novelty of method with a wide spectrum of varied conditions. The average accuracy of the method is 97.20% under standard conditions.***

Keywords **: *Human activity recognition, computer vision, rule based approach, video content management and analysis, motion feature templates.***

I. INTRODUCTION

The human activity recognition is important because of its various useful applications in *video surveillance, content based video indexing, retrieval and filtering, performance evaluation and improvement in sports, augmented reality, intention recognition and analysis, and traffic video analysis* [1-5]. The multidisciplinary nature of the task of human activity recognition has drawn the interest of large research community in recent time. Still, recognition of human actions by vision is a challenging task. Recently, several vision based activity recognition techniques have been proposed in literature but most of them work in constrained set-ups tackling with only specific challenges of the task [1]-[3]. Various bench-marked data sets exist for human action recognition in video. In the work of Chaquet et al [4], various contemporary video datasets for human activities have been discussed. When it comes to devise methods for human action recognition, the task is full of challenges such as dealing with dynamic nature of video content, occlusion, background, variation in speed and pattern of performing an action by various actors, view-point etc.

Revised Manuscript Received on November 30, 2019.
* Correspondence Author
Chandra Mani Sharma*, School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India. Email: cmsharma.its@gmail.com
Alaknanda Ashok, Women Institute of Technology, Email: alakn@rediff.com
Alok Kumar Singh Kushwaha, Department of CSE, IKGPTU, Jalandhar, India. Email: dr.alok@ptu.ac.in

II. RELATED WORK

The idea of activity recognition by vision, particularly by motion, is about 4 decade old. Jonssons et al [6] in 1973 demonstrated how human actions can be interpreted by the motion of light bulbs attached to body. An activity is a harmonic repetition of certain steps performed repeatedly after certain time interval. Image template matching based methods are a popular choice for activity recognition due to their simplicity. As argued in [7], image model based action recognition methods are computationally efficient. The detection and labeling of body parts is not necessary for recognition of the activities. Motion History Image (MHI) and Motion Energy Image (MEI) based methods also fall under image template based methods. One such method is proposed by Bobick et al in [8]. This method works fine to recognize simple aerobic exercise activities. The limitation of the method is that it can work only with the static background and uses frame differencing for background segmentation. Also only motion of the object is taken into account to represent and recognize the different activities. The improvement over Bobick's approach was proposed by us in [9], where not only motion but also the shape of the actor is taken into account for creating spatio-temporal activities and background segmentation is performed with the help of a statistical model. The approach reports improvements over Bobick's method in terms of accuracy. The limitation of the method is that the activities of a single actor can only be recognized from almost frontal view.

Some optical flow based approaches have also been proposed in literature. The upside of these approaches is that the methods are simple and you need not fit any model to recognize the actions. The downside includes the slow processing speed of the methods and the assumption that the optical flow between consecutive is induced by actors only and not by other protuberances such as changing lighting conditions etc. Kumar et al. [10] describe a rule-based activity recognition system. The system can lexically interpret the behavior of target objects in traffic network scenario into two categories. Apriori knowledge of the context of the behaviors of system is required to analyze the system. This is hard coded in the system. The limitation here is that system cannot predict the random dynamic behaviors. In order to commensurate the difference between the video plane coordinates and real world coordinates, camera calibration is required.

Suppose, two different objects are moving in video with the same speed then object appearing near to camera will seem to be moving faster than the farther object. Camera calibration helps in scaling this mismatch. Rahman et al. present a nice discussion on motion history and its applications in [11]. Motion based pattern recognition opens up various new dimensions for research such as gait recognition, activity recognition etc. The MHI representation of motion has many useful applications in this regard. Object motion has been considered for long as an important cue for representing and recognizing the object actions. Template matching based methods first convert an image sequence into a static activity template. Multiple activities will result into distinguishable different templates. At run time the activities are matched with the stored pattern templates.

Pros of template matching based approaches:

- Easy to implement
- Efficient and fast in speed

Cons of the template matching based methods:

- More prone to noise
- More susceptible to the variations of the time intervals of the movements.
- The simple MHI based method is not suitable for the scenes having dynamic background.
- Cannot handle occlusion

Most of the human activity recognition methods do not take into consideration the changes in performing an activity. However, as the way and style to do an activity is generally not fixed. It changes and evolves over time. We propose an activity recognition method that uses a fusion approach and takes advantage of multiple cues.

III. THE PROPOSED METHOD

Fig. 1 shows the schematic diagram of the proposed method. It deploys a fusion approach for activity recognition.

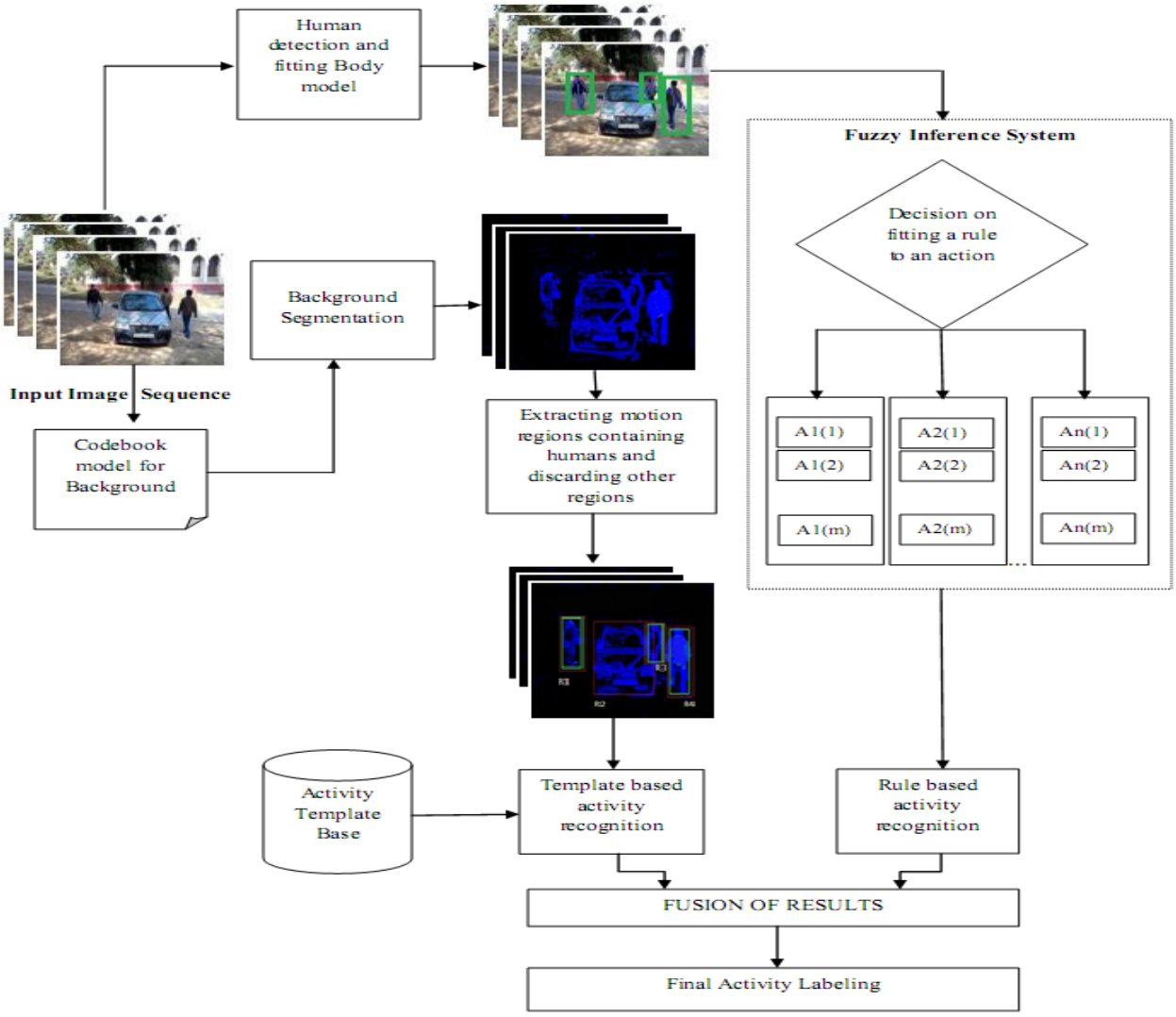


Fig. 1: Schematic Diagram of the Proposed Method

The spatio-temporal activities for various activities are constructed and stored in activity base. Fig 2 shows some such activity templates stored in activity base. We construct codebook [12] of scene to model the variations in the background. The reason behind choosing this approach is that the simple average frame differencing method does not give good background segmentation results when the dynamics in scene change.

Retrieval Number: A4855119119/2019©BEIESP
DOI: 10.35940/ijitee.A4855.119119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Codebook method is efficient in computation in comparison with multi-model GMM or neural network based techniques. As background segmentation is an intermediate step fast computational speed is a desirable parameter. Also if there are multiple objects (of different types) in video e.g. humans, animals, vehicles etc., how to isolate the objects of interest out of these in order to monitor their activities. The methods described in [8-9] suffer from the following limitations-

- These techniques take into account only the motion caused by objects and hence fail to distinguish the humans among other objects. So, techniques are suitable where video contains only humans as moving objects.
- If objects other than human are present in video, then motion history generated by some non-human object is mistaken to be some nearly similar activity performed by a human object (e.g. a moving dog can be misinterpreted by as a bend-moving human object). This limitation may lead to inaccurate recognition of activities.

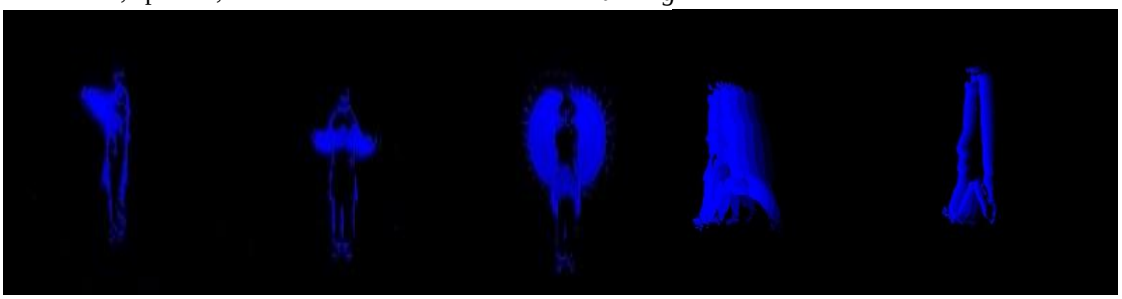
On the contrary, the proposed method has the ability to identify human objects among several other types of objects. Also activity recognition is performed only on the regions of interest (RoI) containing human object and other regions are discarded. Discarding the not-of-interest regions saves on unnecessary computational overhead. The proposed approach minimizes the confusion in activity recognition using a fusion approach. In a raw input video frame humans are detected by face/head in frontal, profile, and back views. A cross validation is made among human motion regions and the regions of human presence. A typical video frame as shown in *Fig.1* contains four moving objects. Four motion regions R_1, R_2, R_3 & R_4 are created by these objects. After applying human detector in these regions it is estimated that region R_2 does not contain a human object, there R_2 is discarded for further processing as here we are concerned to recognize only human activities.

It is to note that the activities of other objects can also be easily recognized using the proposed method as the motion pattern of a running car is altogether different and distinguishable from that of a human being. On regions R1, R3 and R4, we apply template matching which tags these regions with the following actions:

R1<- "Walking", R2<- "Walking" and R3<- "Walking"

We take into account some fuzzy activity rules to improve the acctivity recognition accuracy.

If the Object is walking then there is small change in angles θ_1,, θ_2, θ_3 ,and θ_4.The angles θ_1 and θ_2 fluctuates between 180-150 degrees. If the angle fluctuates between 180-90 degree the activity is running. Let ε_j , ε_w , and ε_r be the mean optical flow of human objects for "*jumping*", "*walking*" and "*running*" activities. Further strengthening the concept, we take fact into consideration related to object motion that $\boldsymbol{\varepsilon_j<\varepsilon_w<\varepsilon_r}$**.** It means that net motion of a human actor is in increasing order while performing "*jumping*", "*walking*" and "*running*" activities.



(a) (b) (c) (d) (e)

Fig 2. Activity Template Base (ATB) containing activities; (a) boxing, (b) clap, (c) hand- wave, (d) run, and (e) walk

(a) (b)

Fig 3: (a) A raw input video frame (b) Background segmentation and object extraction from the fame

Table 1: The cross matching in the regions

Region of Interest	Motion	Detection of Human
R1	Y	Y
R2	Y	N
R3	Y	Y
R4	Y	Y

We take into account some fuzzy activity rules to improve the acctivity recognition accuracy.If the Object is walking then there is small change in angles θ_1, θ_2, θ_3 ,and θ_4.The angles θ_1 and θ_2 fluctuates between 180-150 degrees. If the angle fluctuates between 180-90 degree the activity is running. Let ε_j , ε_w , and ε_r be the mean optical flow of human objects for "*jumping*", "*walking*" and "*running*" activities. Further strengthening the concept, we take fact into consideration related to object motion that $\boldsymbol{\varepsilon_j<\varepsilon_w<\varepsilon_r}$**.** It means that net motion of a human actor is in increasing order while performing "*jumping*", "*walking*" and "*running*" activities.

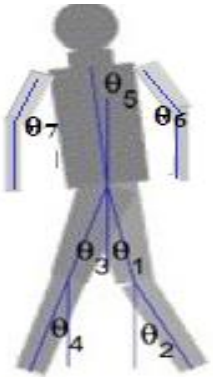


Fig. 4: Important angles in making decision of an activity

If during the performance of an activity i-

Rule 1: If the minimum values of key angles θ_1, θ_2, θ_3 ,and θ_4 fluctuates between 180-150 degree and $\varepsilon_i \geq \varepsilon_j$, then activity label for the ROI under consideration is *"walking"*.

Rule 2: If the minimum values of key angles θ_1, θ_2, θ_3 ,and θ_4 lies between 150-90 degree

If $\varepsilon_i \geq \varepsilon_j$,

Then activity label is "running".

Else

Activity label is "jumping".

Rule 3: If the minimum values of key angles θ_1 and θ_2, lies between 90-30 degree and θ_3 and θ_4 does not change then activity label is "clapping".

Rule 4: If the value of key angles $\theta_1$and θ_2 does not change and θ_3,and θ_4 lies between 180-150 degree then activity label is "hand-waving".

Rule 5: If angles θ_1, θ_2, θ_3 ,and θ_4 do not change and ε_i is 0, then activity label for the ROI under consideration is "standing".

Rule 6: Other tested activities such as "crawling" can be accurately measured using template matching only as they have clear distinguishable templates and do not require further rectification.

Description of the Proposed Method:

i. Background segmentation using background codebook modeling

ii. If there are N frames in a video sequence $V_{i(1\leq i\leq N)}$ having corresponding segmented frame sequence $S_{i(1\leq i\leq N)}$ obtained in step I. n_i connected components regions are extracted from a given segmented frame s_i. Draw the bounding box around each of these component regions.

iii. In corresponding v_i frame of input video detect for frontal, profile or back view of human face or head using an ensemble of binary classifiers for frontal, profile or back views.

if detected view is *frontal*
then fit stick_model_1
if detected view is *profile*
then fit stick_model_2
if detected view is *back*
then fit stick_model_3

iv. Calculate tangents, optical flow and Mahalanobis distance to match activity templates. The popular moment invariants (which is a standard statistical concept and is characterized by seven equations) are used as the means to describe an activity. Mahalanobis distance is calculated between the moment descriptions of the input and each of the known actions. Subsequently, the input action is recognized. The distance matrix obtained after this step is critically analyzed. Here, separation distances for different actions are taken into consideration. For each given activity in activity set $A(a_1, a_2, a_3 \ldots a_k)$, activity templates are constructed with the help of the *motion history images* (MHIs). In video, we estimate the direction of motion by gradient orientation. Gradients of the MHI are calculated by convolving it with separable Sobel filters in both directions (x and y). Smaller Mahalanobis distance signifies a better match for activity recognition. It is also used as a tool for conflict resolution.

v. Apply fuzzy rules for activity prediction.

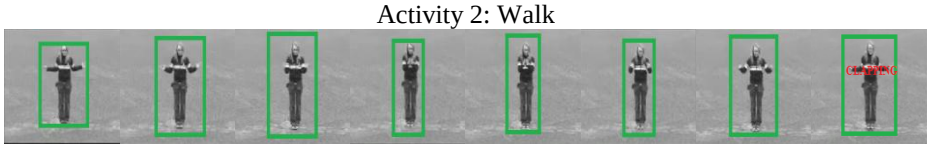
vi. Fuse the results of step iv and v for activity recognition.

The modern research is focused on minimizing the search area in object detection and classification task. In the proposed method eliminates brute force searching for face/head detection. It does so by reducing Search Area to a few small sized (as compared to size of entire video frame) pre-determined regions of interest (ROIs). Further, in all activities considered here, human face/ head falls only in half of the upper area of bounding rectangle. This means that an actor perform different activities by standing, sitting or bending. This consideration further reduces the potential search area.

IV. EXPERIMENTAL RESULTS

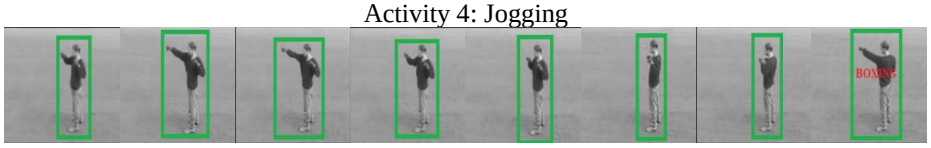
Activity 1: Both Hand-waving

Activity 2: Walk

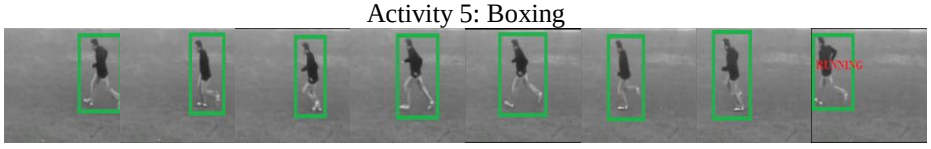


Activity 3: Handclapping

Activity 4: Jogging



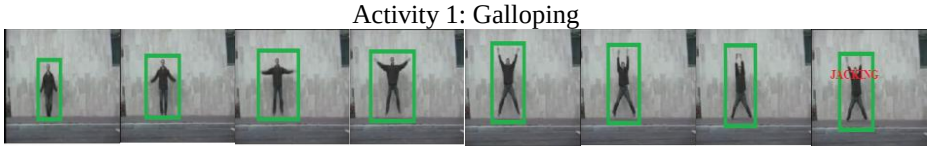
Activity 5: Boxing



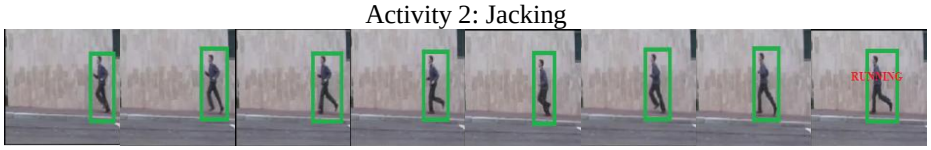
Activity 6: Running

Fig 5: Activity Recognition on KTH Dataset.

Activity 1: Galloping

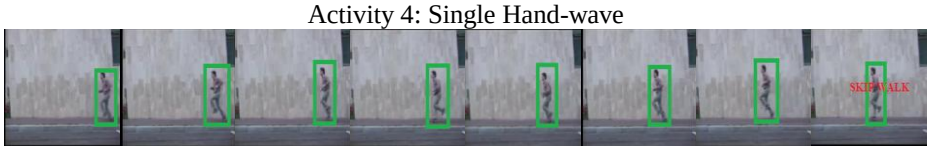


Activity 2: Jacking



Activity 3: Running

Activity 4: Single Hand-wave



Activity 5: Skip Run

Fig 6: Activity Recognition on Weizmann Dataset

Method has been tested on a variety of activity sets including two standard activity data sets KTH and Weizmann activity datasets. Figure 5 and 6 represent the output results respectively on the samples from KTH and Weizmann data sets. For every activity, left most frame is the starting frame and right most frame is the ending frame. The average accuracy of the method is 97.20%.

The method can easily recognize a wide range of activities including Running, Walking, Jogging, Jacking, Single Hand Waving, Double Hand Waving, Clapping, Boxing, Side-walking, Skip Walking, Sitting, Bending, Bend-moving etc.

The proposed method has following advantages as listed below:

- It can accommodate a wide spectrum of human actions and is customizable to handle new actions required for a particular application.
- The robust code book technique employed in the method makes it suitable for outdoor environment and background can be performed.
- The application of fuzzy rules considerably improves the performance of the method in terms of false negatives and false positives. The method has an average accuracy rate of 97.20% on different standard datasets.
- The method is simple and straightforward.

Moreover, the technique is suitable for the following applications:

- Visual surveillance
- Training and performance evaluation of players in sports
- Content based video indexing and retrieval
- Automatic monitoring of work floor etc.

V. CONCLUSION

The paper describes a technique for activity recognition in videos. The task of activity recognition is valuable and important due to its many applications in visual surveillance, access control, content based video management, sports etc. The method uses a fusion approach and improves the accuracy of the typical MHI based approach. It has been tested on two standard datasets namely KTH and Weizmann and recognizes the activities like running, walking, jogging, clapping, skip-running, galloping, boxing etc. The average activity recognition accuracy is 97.20%.

REFERENCES:

1. Vishwakarma, S., & Agrawal, A. (2013). A survey on activity recognition and behavior understanding in video surveillance. Int. J. The Visual Computer, 29(10), pp 983-1009.
2. Robertson, N., & Reid, I. (2006). A general method for human activity recognition in video. Computer Vision and Image Understanding, 104(2-3), 232-248.
3. Kushwaha, A. K. S., Sharma, C. M., Khare, M., Srivastava, R. K., & Khare, A. (2012). Automatic multiple human detection and tracking for visual surveillance system. In 2012 IEEE International Conference on Informatics, Electronics & Vision (ICIEV) , pp. 326-331
4. Chaquet, J. M., Carmona, E. J., & Fernández-Caballero, A. (2013). A survey of video datasets for human action and activity recognition. Computer Vision and Image Understanding, 117(6), pp 633-659.
5. Kushwaha, A. K. S., Sharma, C. M., Khare, M., Prakash, O., & Khare, A. (2014). Adaptive real-time motion segmentation technique based on statistical background model. The Imaging Science Journal, 62(5), pp 285-302.
6. Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Perception & psychophysics, 14(2), pp 201-211.
7. Sharma, C. M., Kushwaha, A. K. S., Nigam, S., & Khare, A. (2011, September). On human activity recognition in video sequences. In 2011 2nd International Conference on Computer and Communication Technology (ICCCT-2011), pp. 152-158
8. Bobick, A. F., & Davis, J. W. (2001). The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis & Machine Intelligence, (3), pp 257-267.
9. Sharma, C. M., Singh Kushwaha, A. K., Nigam, S., & Khare, A. (2011, July). Automatic human activity recognition in video using background modeling and spatio-temporal template matching based technique. In Proceedings of the ACM International Conference on Advances in Computing and Artificial Intelligence (pp. 97-101)
10. Kumar, P., Ranganath, S., Weimin, H., & Sengupta, K. (2005). Framework for real-time behavior interpretation from traffic video. IEEE Transactions on Intelligent Transportation Systems, 6(1), 43-53.
11. Ahad, M. A. R., Tan, J. K., Kim, H., & Ishikawa, S. (2012). Motion history image: its variants and applications. Machine Vision and Applications, 23(2), 255-281.
12. Kim, K., Chalidabhongse, T. H., Harwood, D., & Davis, L. (2005). Real-time foreground–background segmentation using codebook model. Real-time imaging, 11(3), 172-185.

AUTHORS' PROFILE

Chandra Mani Sharma He is currently working with School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India. He holds masters degree in technology and currently pursuing PhD in the area of computer vision. He is an active researcher and has more than 30 publications with reputed publishers like ACM, IEEE, Springer, IGI-G, Taylor & Francis etc.

Alaknanda Ashok Dr Alaknanda Ashok is working as Professor and Director of Women Institute of Technology, Dehradun, India. She is a seasoned academician, administrator and researcher. She earned her PhD from Indian Institute of Technology, Roorkee. She is actively involved in a number of research, development and consultancy activities. She is the recipient of many awards and recognitions for her notable contribution.

Alok Kumar Singh Kushwaha Dr. Alok Kumar Singh Kushwaha is working as an Assistant Professor in the Department of CSE, IKGPTU, Jalandhar. He earned his PhD from Indian Institute of Technology, Varanasi. He has authored and co-authored several national and international publications and also working as a reviewer for reputed professional journals. Dr. Kushwaha is having an active association with different societies and academies around the world.