Human Recognization Activity and Maximum Motion Representation in Surveillance Video

Mr. Mariyan Richard A [1], Dr. Prasad N Hamsavath [2]

1Assistant Professor, Department of Masters of Computer Applications,

Nitte Meenakshi Institute of Technology

2Professor & H.O.D., Department of Masters of Computer Applications

1mariyanrich01@gmail.com, 2naikphd@gmail.com,

doi: 10.48047/ecb/2023.12.s2.333

Abstract

Recognizing human actions in the real-world environment finds vital applications, Hence, machine vision studies in this field become crucial. This research aims to extract human activity from video sequences. The number of activities are required for human action recognition, including the gathering of visual data, the identification and presentation of robust features, and the training of classifiers with strong discriminative abilities. The action recognition method employed in this research is based on maximal motion identification. The Region of Interest (ROI) difference picture is utilized to extract motion information, and the Block Based Motion Intensity Code (BBMIC) is extracted as a feature. The Weizmann action dataset, which includes 10 actions (including "walk," "run," "jump," "side," "bend," "wave1," "wave2," "pjump," "jack," and "skip"), was used in the trials. A variety of the tree-based classifiers, including Random-Tree, the Random-Forest, and the Decision-Tree (J48), were also used. According to the experimental findings, this method uses the Weizmann dataset to recognize the activities with an overall quality rate of 94.20%. This is more effective than other tree-based algorithms. Performance of the suggested strategy is on par with that of well-established, established procedures.

Keywords: Video Surveillance; Activity Recognition; Frame Difference; The Random Forest; The Random Tree; The Decision Tree.

1. Introduction

Many applications that require actual user activities to be recognized, particularly those that need proactive assistance. If robots could automatically understand the activities that people could undertake in daily life, several occupations would undergo a revolution. This has resulted in an enormous amount of activity in the field of computer vision's study of human action, which aims to automatically segment, capture, and recognize human action in real time as well as possibly forecast continuing human behavior's. Identifying user behavior's entails drawing conclusions about the user's recent behaviour from data in the form of videos, which are regarded as a rich source of knowledge. The Automatic detection of human activities brings up a number of intriguing application areas, including robotics employing abstract concepts from films, healthcare activities, autonomous visual observation, human-computer interaction, gesture recognition, and content-based retrieval and indexing. An action can be thought of as a series of simple activities, such as leaping, walking, or kicking a ball, however it can be challenging to recognize actions in videos without the aid of a person because actors can carry out the same action in multiple ways and also at various speeds.

Human actions are influenced by a variety of factors, including the environment, culture, individual differences, and emotions. Different people will do the same thing in different ways, and even the same person will do it differently at different times. Automatic action recognition is an extremely challenging undertaking because of the variety of human body size, look, shape, and sophistication of human actions. First off, since most algorithms can only encompass a limited number of action types due to a lack of training data, labelling a broad range of actions is time-consuming. Second, multiple environmental settings, lighting conditions, camera settings, and either fixed or moving camera positions are used to gather the data. Because of this, most algorithms are constrained to specific usage cases. Recent technological advancements make it easier to resolve these problems.

1.1 Related Work

The feature descriptor, representation, and classification models in video sequences are the subject of recent reviews in the field of human action analysis in [1–3]. A space-time shape based on 2D human silhouette is proposed by Gorelick et al. [4] and contains the spatial features about the attitude and shape descriptors that were employed to recognise the human action. The activities were characterized using human silhouettes, and the actions were represented using temporal templates named Motion History Image (MHI) and Motion Energy Image (MEI) [5]. Hu moments or an SVM classifier were used to connect the actions. Each frame of an image sequence receives a visual word from Y. Wang et al[7] .'s analysis of the motion of the subject in the frame. A bag of word classifier is used to recognize complex events at various temporal scales by combining temporal structure and local features [8]. The motion history frames are used to build the object shape information and spatial-temporal template for various human activities like walking, running, bending, sleeping, and jumping in the human activity recognition in video technique based on background modeling [9]. An approach for both object detection and action recognition in video surveillance scenarios is presented by Churn-Hao Wang [10].

The Histogram of Oriented Gradients (HOG) method is used to the find objects, and the Hidden Markov Model (HMM) is used to record the features' temporal structure. Decisions are made based on an understanding of the object's motion trajectory and the connection between events and the object's movement. A method based on human silhouettes' extended motion templates was proposed in [11]. To differentiate human action, which represents both local and global information, holistic structural features were extracted from motion templates. Wang L and al. [12] presents an edge based learning discriminative element technique for activity acknowledgment. The experiments in [13] are carried out the most widely used human-action recognition datasets, such as the KTH and Weizmann-datasets, and they are based on motion information excerpted from the difference image based on Region of Interest (ROI) using 18-dimensional features known as Block Intensity Vector (BIV). A Motion Intensity Code (MIC) is presented in [14] to recognize human actions based on motion data. Better performance was achieved when the KTH and Weizmann datasets were used to learn and classify action using the PCA and SVM.

1.2 Outline of the work

This study examines activity recognition, which aims at recognizing human actions from video clips. With the person performing actions like "walk," "run," "jump," "side," "bend," "wave1," "wave2," "pjump," "jack," and "skip," Weizmann's [15] action dataset is used to assess the proposed approach. The sequential pictures are subtracted to get the difference image. Choosing the Area of Interest allows for the collection of motion information (ROI). The ROI that's been extracted is broken up into 9 x 6 chunks. The block that has the greatest motion in the divided ROI is used to extract the feature known as Block Based Motion Intensity Code (BBMIC). For activity recognition, the obtained features is fed into tree-based classifiers like Random Forest, Random Tree, and Decision Tree (J48).

The remainder of the paper is structured as follows. Section 2 depicts and discusses the feature extraction process. Section 3 describes the proposed approach's workflow. The Experimental results are presented in Section 4. Section 5 brings the paper to a close.

2. Feature Extraction

The feature, which is representative of the significant information required for extensive research, is a typical characteristic taken from an image/photo or video sequence. The feature used in this study is described in the ensuing subsections.

2.1 Frame Difference

Motion is an essential active information used to Identify the human activity. A pixel-by-pixel comparison of the current frame at time t + 1 with the previous frame t yields the segmented image.

(a) **(b)** **(c)**

Fig. 1. (a), (b) Two Consecutive frames. (c) Difference image of (a) and (b) from Weizmann dataset.

$$D_k(i,j) = |I_k(i,j) - I_{k+1}(i,j)| \quad (1)$$

$$1 \leq i \leq w, 1 \leq j \leq h$$

The extracted motion data is referred to as the Region of Interest (ROI). Figures 1(a) and 1(b) show two successive frames from the Weizmann dataset. Figure 1 depicts the resulting difference image (c). Dk(i, j) is the difference image, Ik(i, j) is the intensity of the pixel I (j) in the kth frame, and w and h are indeed the image's width and height. The difference image or movement information Tk is calculated using

$$T_k(i,j) = \begin{cases} 1, & \text{if } D_k(i,j) > t \\ 0, & \text{otherwise;} \end{cases} \tag{2}$$

where *t* is the threshold.

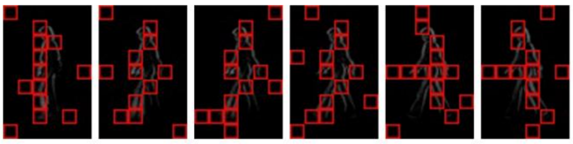
2.2 Block Based Motion Intensity Code (BBMIC)

Motion is an important cue extracted from video to recognise the performed action. In this regard, the video sequences are used to extract Block Based Motion Intensity Code (BBMIC). This section describes the procedure for extracting the feature.

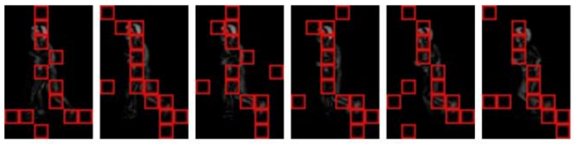
(a) (b) (c) (d)

Fig. 2. (a) Data Motion information. (b) ROI derived from (a). (c) 9x6 block description of (b). (d) Max motion block.

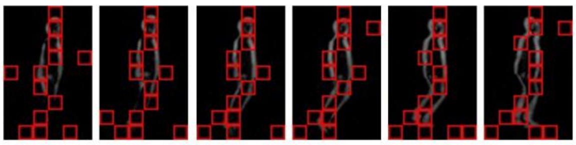
Initially motion identified region is considered as ROI of size 90 X 60, as seen in Fig. 2(a). The extracted ROI is shown in Fig. 2(b). ROI divided into 9 x 6 blocks, thus giving a total of 54 blocks as shown in Fig. 2(c). As shown in Fig. 2, only the block with the largest motion along the row and column is taken into consideration for further analysis in order to compress the number of calculation (d). Among the 54 blocks, only top 15 blocks that show maximum motion only are considered for further analysis. The average intensity of each block in 9 X 6 regions calculated and thus an 15 - dimensional feature vectors are extracted from the selected blocks and it is fed to the tree based classifiers for further processing. Fig. 3 shows the motion information extracted for different actions from Weizmann dataset.



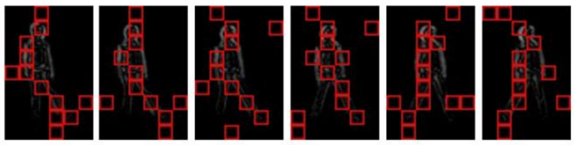
(a) walk



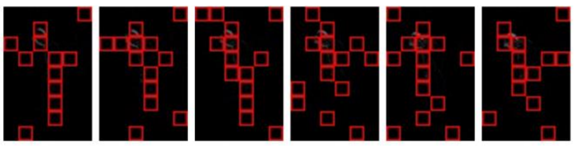
(b) run



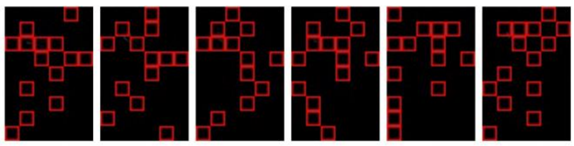
(c) jumping



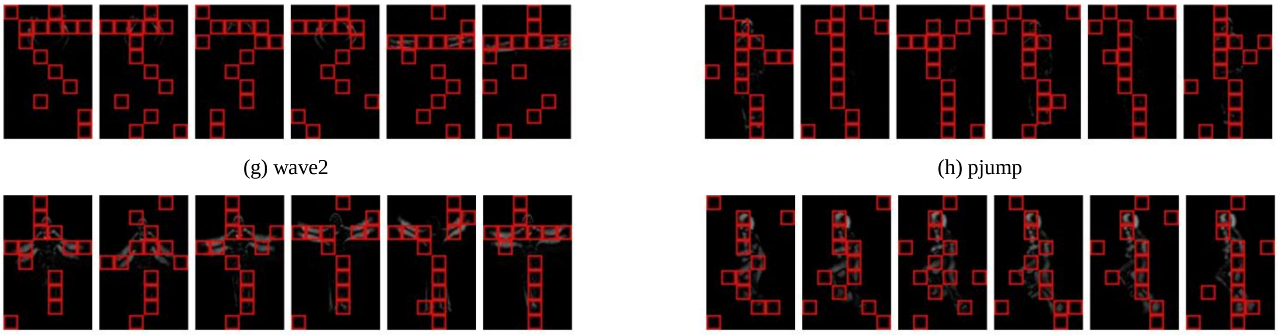
(d) side



(e) bend



(f) wave1



(g) wave2 (h) pjump

(i) jack (j) skip

Fig. 3. Extracted maximum motion block from ROI.

3. BBMIC Action Recognition Approach's workflow proposed

Weizmann-action datasets are used in experiments. The video is rendered at a rate of 25 frames per second. Smoothing is accomplished through Gaussian convolution with a kernel size of 5 X 5 and variance = 0.5. To erase noise for fine feature extraction and classification, all video sequences must be pre-processed. As discussed in Section 2, ROI is derived from the video sequence, and BBMIC features are excerpted. For action recognition, the features are fed into a tree-based classifier. The workflow of the proposed approach is shown in Fig. 4. In this work, different tree based classifier like Random Forest [16], Random Tree, and Decision Tree (J48) [17] are employed in order to assess the efficacy of these classifiers on Weizmann dataset.

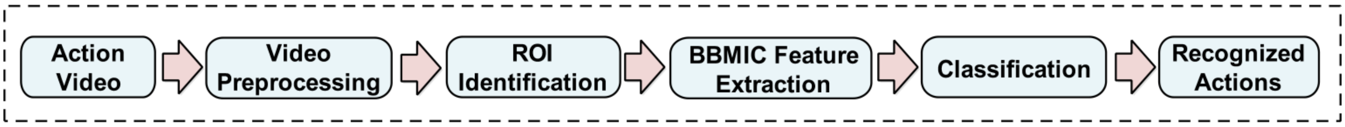


Fig. 4. Workflow of the proposed BBMIC approach.

4. Experimental Results

The proposed method is tested against by the Weizmann-action dataset. The following experiments can be carried out in C++ programing comcepts with OpenCV 2.4 on a computer with an Intel Xeon Processor 2.40 GHz and 4 GB RAM running Ubuntu 12.04. The BBMIC features are fed into the tree-based classifiers such as The Random-Forest, The Random-Tree, and The Decision-Tree (J48) using the source code Deep Learning Tool WEKA [18] to build a model for each activity, and so these models are used to gauge how well each tree-based classifier succeeds.

4.1 Weizmann Dataset

On the Weizmann action dataset, the proposed method is assessed. It has 81 video clips from nine different people depicting 10 different actions: “walk”, “run”, “jump-forward-on-two-legs” (or “jump”), “gallopsideways” (or “side”), “bend”, “wave one-hand” (or “wave1”) , “wave-two-hands” (or “wave2”), “jump-in-place-on-two-legs” (or “pjump”), “jumping-jack” (or shortly “jack”) and “skip”. The video sequences have 180 x 144 pixel resolution at 25 fps. The sample frames of action sequence are shown in Fig. 5. 10 actions like “walk”, “run”, “jump”, “side”, “bend”, “wave1”, “wave2”, “pjump”, “jack” and “skip” are used in this experiment. For conducting experiments, video clips showing all these 10 actions performed by all nine people were considered.

Fig. 5. A Sample frames from the Weizmann-dataset.

4.2 Quantitative Evaluation

A 10-fold cross validation approach is used to measure performance. The tree-based classifier is fed the features. The performance of the suggested technique is assessed using the formulas Precision (P) = TP/TP+FP, Recall (R) = TP/TP+FN, and F1 = 2.(P.R)/(P+R). Where TN and FN are the true negative predictions and false negative expectations for the specific class, and

TP & FP are true positive and false positive predictions for the specific class, respectively. The build-time of all the tree-based classifiers is also measured. Table I displays accuracy findings from the Weizmann-dataset using The Random-Forests, The Random-Tree, and Decision-Tree (J48).

4.3 Experimental Findings for Classifiers Based on Trees

WEKA, machine learning tool and is an open source was used for the experiment. The classifiers' performance was evaluated using a 10-fold cross validation model. to forecast the performance of each tree-based classifier. Classification accuracy, Precision, Recall, and F-measure scores are used to evaluate the performance of the classifiers.. Table I shows the comparison of tree based classifiers in terms of Precision, Recall, Fmeasure and Time complexity. Compared to other tree-based classifiers, Random Forest requires more time to learn for the provided dataset, but it also performs well, with an accuracy rate of 94.20%.

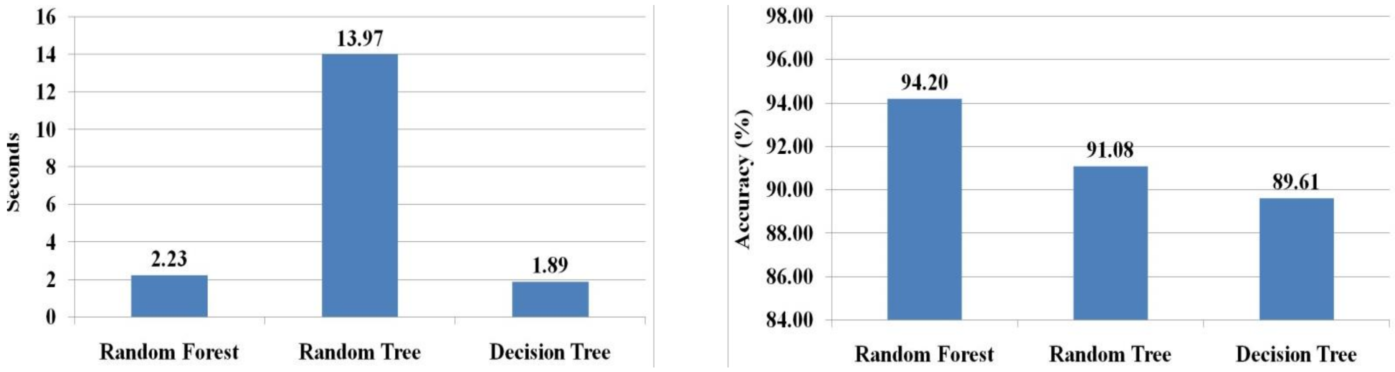
Table I. The tree-based categorization algorithms' performance metrics

Algorithm	Precision (%)	Recall (%)	F-Score (%)	Time (Sec.)
Random Forest	94.70	94.72	94.71	2.23
Random Tree	91.10	91.11	91.10	0.23
J48	89.60	89.60	89.60	1.89

Table. II shows the average classification results of the Random Forest classifier. Ten actions from the Weizmann action dataset are taken into consideration, including "walk," "run," "jump," "side," "bend," "wave1," "wave2," "pjump," "jack," and "skip." The majority of the actions are correctly identified, with the appropriate reaction defining the primary diagonal.As seen in Table II, since the “side” and “walk” action are mispredicted, some of the side sequences are misclassified as “walk” and some actions like “skip” and “run” are similar. Thus, it needs further attention. The overall performance accuracy of the proposed BBMIC and time consumed to build the model with different tree based classifier on Weizmann dataset is shown in Fig. 5.

Table II. Confusion matrix (%) of the Weizmann dataset using Random forest classifier (**94.20%**)

Action	walk	run	jump	side	bend	wave1	wave2	pjump	jack	skip
walk	**97.71**	0.00	0.42	1.04	0.00	0.00	0.00	0.00	0.21	0.62
run	0.96	**93.95**	0.32	1.27	0.00	0.00	0.00	0.00	0.00	3.50
jump	0.47	0.94	**94.37**	0.47	0.00	0.00	0.00	0.00	0.00	3.76
side	8.24	2.51	1.08	**84.59**	0.00	0.00	0.00	0.00	0.36	3.23
bend	0.00	0.00	0.00	0.00	**98.80**	0.60	0.30	0.30	0.00	0.00
wave1	0.00	0.00	0.00	0.00	1.78	**97.72**	0.25	0.00	0.25	0.00
wave2	0.00	0.00	0.00	0.00	0.69	1.15	**98.16**	0.00	0.00	0.00
pjump	0.58	0.00	0.00	0.58	0.58	0.00	0.00	**95.38**	2.89	0.00
jack	0.00	0.00	0.41	0.00	0.61	0.00	0.00	2.04	**96.95**	0.00
skip	3.40	6.80	2.55	2.83	0.00	0.00	0.00	0.00	0.00	**84.42**



(a) Build time

(b) Performance Measure

Fig. 5 Illustrates the build time and overall accuracy of all tree based classifiers

4.4 Comparative Study

To quantify the effectiveness of the suggested method, the findings from the suggested protocol are quantitatively compared with results from state-of-the-art methods. This comparison is shown in Table III. The comparison reveals that the suggested technique performs well on the Weizmann action dataset.

Table III. Comparison of performance on the Weizmann dataset

Method	Classifier	Accuracy (%)
Proposed approach	**Random Forest**	**94.20**
Jia et al. [19]	*k*-NN	90.9
Liu et al. [20]	*k*-NN (*k*=5)	89.3
Thurau [21]	*k*-NN (*k* = 90)	86.7
Klaser et al. [22]	SVM	84.3
Niebles et al. [23]	SVM	81.5

5. Conclusion

This paper presented a method for human activity recognition from video sequences. Using Block Based Motion Intensity Code (BBMIC) as features. The categorization based on motion information uses the ROI that was taken from the segmented images. A variety of motions, including "walk," "run," "jump," "side," "bend," "wave1," "wave2," "pjump," "jack," and "skip," were taken into consideration in experiments on the Weizmann dataset. The maximum motion information is used to extract BBMIC features using the ROI that was recovered from the difference picture. Using the help of tree-based algorithms like Random Forest, Random Tree, and Decision Tree, this approach evaluates the efficacy of the BBMIC feature in a video sequence. The suggested technique produced recognition values for the Weizmann dataset of 94.20%, 91.08%, and 89.61%, respectively. The experiments revealed that the system could not accurately distinguish between side and skip, which is of future interest.

Acknowledgements

The authors greatfully acknowledge **University Grants Commission of India** [F. No. 41-636/2012 (SR)], for funding this work.

References

1. Poppe, Ronald. "Vision-based human motion analysis: An overview." *Computer vision and image understanding* 108, no. 1 (2007): 4-18.
2. Poppe, Ronald. "A survey on vision-based human action recognition." *Image and vision computing* 28, no. 6 (2010): 976-990.
3. Turaga, Pavan, Rama Chellappa, Venkatramana S. Subrahmanian, and Octavian Udrea. "Machine recognition of human activities: A survey." *Circuits and Systems for Video Technology, IEEE Transactions on* 18, no. 11 (2008): 1473-1488.
4. Blank, Moshe, Lena Gorelick, Eli Shechtman, Michal Irani, and Ronen Basri. "Actions as space-time shapes." In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2, pp. 1395-1402. IEEE, 2005.
5. Bobick, Aaron F., and James W. Davis. "The recognition of human movement using temporal templates." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23, no. 3 (2001): 257-267.
6. Meng, Hongying, Nick Pears, and Chris Bailey. "A human action recognition system for embedded computer vision application." In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1-6. IEEE, 2007.
7. Wang, Yang, and Greg Mori. "Human action recognition by semilatent topic models." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31, no. 10 (2009): 1762-1774.
8. Niebles, Juan Carlos, Chih-Wei Chen, and Li Fei-Fei. "Modeling temporal structure of decomposable motion segments for activity classification." In *Computer Vision–ECCV 2010*, pp. 392-405. Springer Berlin Heidelberg, 2010.
9. Sharma, Chandra Mani, Alok Kr Singh Kushwaha, Swati Nigam, and Ashish Khare. "Automatic human activity recognition in video using background modeling and spatio-temporal template matching based technique." In *Proceedings of the International Conference on Advances in Computing and Artificial Intelligence*, pp. 97-101. ACM, 2011.
10. Wang, Chun-hao, Yongjin Wang, and Ling Guan. "Event detection and recognition using histogram of oriented gradients and hidden markov models." In*Image Analysis and Recognition*, pp. 436-445. Springer Berlin Heidelberg, 2011
11. Wu, Di, and Ling Shao. "Silhouette Analysis-Based Action Recognition Via Exploiting Human Poses." *Circuits and Systems for Video Technology, IEEE Transactions on* 23, no. 2 (2013): 236-243.
12. Wang, Liang, Yizhou Wang, Tingting Jiang, Debin Zhao, and Wen Gao. "Learning discriminative features for fast frame-based action recognition."*Pattern Recognition* 46, no. 7 (2013): 1832-1840.

13. Arunnehru, J., and M. Kalaiselvi Geetha. "Automatic Activity Recognition for Video Surveillance." *International Journal of Computer Applications* 75 (2013).

14. Arunnehru, J., and M. Kalaiselvi Geetha. "Motion Intensity Code for Action Recognition in Video Using PCA and SVM." In *Mining Intelligence and Knowledge Exploration*, pp. 70-81. Springer International Publishing, 2013.

15. Blank, M., Gorelick, L., Shechtman, E., Irani, M., Basri, R. "Actions as space time shapes". in Proc. IEEE International Conferences on Computer Vision, pp. 1395–1402, 2005.

16. L. Breiman, Random Forest, Machine Learning, 45(1) 5–32 (2001)

17. J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, (1993).

18. I. H. Witten and E. Frank. "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations", Morgan Kaufmann, 1999.

19. Jia, Kui, and Dit-Yan Yeung. "Human action recognition using local spatio-temporal discriminant embedding." In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1-8. IEEE, 2008.

20. Liu, Jingen, Saad Ali, and Mubarak Shah. "Recognizing human actions using multiple features." In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1-8. IEEE, 2008.

21. C. Thurau, ``Behaviour histograms for action recognition and human detection," in International Workshop on Human Motion with ICCV, 2007.

22. Klaser, Alexander, and Marcin Marszalek. "A spatio-temporal descriptor based on 3d-gradients." in Proc. BMVC, pp. 1--8, (2008).

23. Niebles, Juan Carlos, and Li Fei-Fei. "A hierarchical model of shape and appearance for human action classification." In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1-8. IEEE, 2007.