ISSN 2229-5518

Human Activity Recognition using Motion History Algorithm

Muhammad Hassan, Tasweer Ahmad, Muhammad Ahsan Javaid

Abstract-- In this research, I have worked to recognize various human actions and activities using Motion History Algorithm. Firstly, I studied different techniques already implemented by various researchers in this field like motion image energy etc and finally develop an approach. In this approach, videos are converted into image frames and the images are pre-processed plus background subtraction is computed for these frames. Then, the motion history of images is computed, and features are extracted. After that, Discrete Cosine Transform (DCT) of the Region of Interest is computed which is then thresholded for classification. The proposed algorithm is able to distinguish between various human actions sittings, standing, hand waving etc.

Keywords: Discrete Cosine Transform, Motion History, Motion Image Energy, Region of Interest.

I. Introduction:

1.1. Overview:

Since the demands of the applications are increasing on daily basis therefore automatically recognition of human activities is in our sights. In surveillance environment, the detection of not normal activities will be very helpful in production of alerts for potential crime. In entertainment environment, this recognition will update the human computer interaction (HCI). In healthcare system, recognition of activities will provide help the rehabilitation of patients.

Generally, recognition of human activities is divided into two levels of representations as shown in figure 1:

1. Low-level core technology.
2. High-level applications.

In core technology, there are 3 main stages for processing any activity. These stages are object segmentation, feature extraction and classification algorithms. First of all human object in a video sequence will be segmented out. The main characteristics are carefully extracted and the properly represented by labels or features such as:

- Shape.
- Silhouette.
- Colors.
- Poses.
- Body motions.

So in a nut shell, once the features are extracted, the algorithms are applied for the detection, recognition and classification of human activities. High level of applications only works on the results which are obtained when the algorithm are successfully applied on surveillance environments, entertainment environments or healthcare systems.

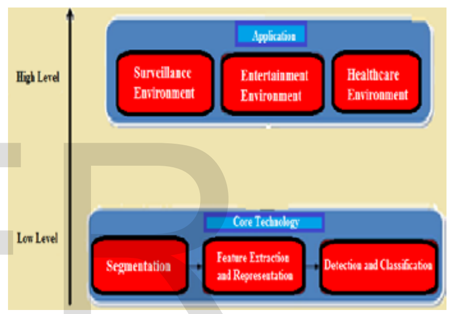

Fig.1. Overview

1.2. Segmentation:

In core technology, the target object is obtained by the help of object segmentation which is applied on each and every frame in the video sequence. Since the cameras are either mobile or static so object segmentation is divided in two categories as shown in Figure 2:

- Static camera segmentation.
- Moving camera segmentation.

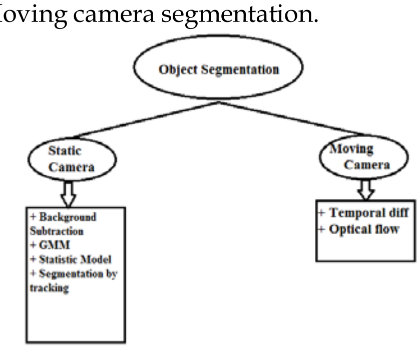

Fig.2. Types of Object Segmentation

In static camera segmentation, the position and the angle of the camera are fixed therefore viewpoint of the object and background remains the same. The method used most commonly used for the static segmentation of an

- *Muhammad Hassan has completed masters degree program in electrical control engineering from GC University, Lahore, E-mail: engr.muhamadhassan@gmail.com*
- *Tasweer Ahmad is currently pursuing PhD degree program in Computer engineering in LUMS University, Lahore, E-mail: tasveerahmad@gmail.com*
- *M Ahsan Javaid has completed masters degree program in electrical control engineering from University Of South Asia, Lahore, E-mail: ahsanjavaid888@gmail.com*

object is background subtraction [1-3]. This is because of simplicity and efficiency of background subtraction. First step is to obtain an image of background image with no foreground object in it. The next step is to subtract each frame in the video sequence from the background image to gather the entire set of objects needed to be analyzed. Although a slight change in illumination affects the background. Many complex algorithms are proposed to make a perfect background model from a couple of background images. The background model will be made by the following algorithms:

- Single Gaussian for each pixel.
- Mixture of Gaussians for each pixel [4, 5].
- Statistical parameters with intensity change for each pixel.
- Statistical parameters with chromaticity change for each pixel [6].

Static camera segmentation by tracking [7, 8] is not like the algorithms that are mentioned above. In the above mention algorithms background model is prepared in advance. While tracking algorithm is kind of a dynamic time warping in which the object will be obtained by tracking regions. It can easily obtain when there is locally smooth motion.

In case of moving camera segmentation camera will be mobile either with the photographer or with the robot. The moving camera segmentation is tough than static because it has to face three continuously varying factors:

- Motion of the target object.
- Motion of the camera.
- Variation in background.

Since the background is dynamic therefore creating background model is not possible. Since the proper background model will not be created so we use the temporal difference algorithm [9, 10]. In this algorithm we focus on the difference between two consecutive image frames in a video sequence. The motion of the camera can be reduced by the optical flow algorithm [11, 12]. Optical flow calculates the pixel-level motion present in two consecutive frames. The features are tracked and also the transformations of the coordinates of the two consecutive frames are obtained. So the segmentation using moving camera can be obtained by the transformed coordinate algorithm.

1.3. Feature Extraction and Representation:

In next stage of core technology characteristics obtained from the segmented objects are represented as feature points. The characteristics of the segmented objects are:

- Shape.
- Silhouette.
- Colors.
- Motions.

These feature points are divided into four groups as shown in figure 3:

1. Space-time information.
2. Frequency transforms.
3. Local descriptors.
4. Body modeling.

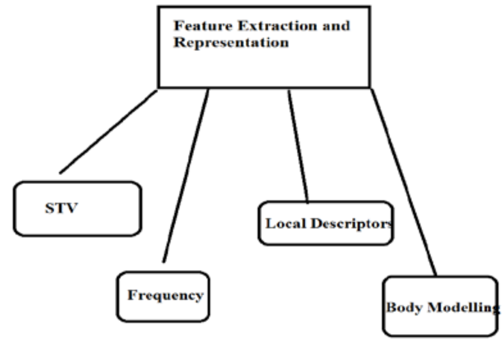

Fig.3. Feature Extraction and Representation

1.3.1. Space Time Volume (STV):

First of all the information of space time is discussed. Space time volume (STV) [13] is made as the image features and this can be done by linking together consecutive silhouettes with axis of time. 3D XYT volume is obtained which is comprised of x-y spatial coordinates and time. This volume will capture the continuity of aperiodic activities. STV is limited on viewpoint changes so they detect global features.

1.3.2. Frequency Domain Transformation:

The frequency domain information is also very helpful in recognizing the features. Discrete Fourier transforms (DFT) [14] is most commonly used to get the information regarding geometry of the object. DFT features belong to global features i.e., they consider the whole image instead of a small area and it is also detects the global features.

1.3.3. Descriptors:

Local descriptors [15] detect the local features instead of global features therefore they work on an image patch instead of whole image. The main types of local descriptors are:

- Scale invariant feature transform (SIFT) [16, 17].
- Histogram of oriented gradient (HOG) [18].

They are more invariant to the following:

1. Background clutters.
2. Appearances.
3. Occlusions.
4. Rotation.
5. Scale.

Although the whole body activities are not fully recognized by the above features.

1.3.4. Body Modeling:

ISSN 2229-5518

To recognize full body features locally we need body modeling algorithm in which 2D or 3D pose estimation is required. Once the pose estimated and the coordinates of body are transformed into discriminative feature representations like:

- Polar coordinate representation [19].
- Boolean features [20].
- Geometric relational features (GRF) [21].

1.4. Activity Classification:

Next stage in the core technology is detection and classification in which algorithms are used to recognize features of human activities. They are divided in three types:

- Dynamic time warping (DTW).
- Generative models.
- Discriminative models.

1.4.1. Dynamic Time Warping (DTW):

Dynamic time warping (DTW) [22] is a simple classification algorithm which is used to calculate the similarity between two temporal sequences whose time and speed can be variable. DTW is not good for a huge amount of classes that contains bulk of variations.

1.4.2. Generative models:

Generative models are dynamic classifiers and use probability based approaches. Some of the famous types are:

- Hidden Markov Models (HMM) [23].
- Dynamic Bayesian Networks (DBN) [24].

1.4.3. Discriminative models:

Discriminative models are static classifiers and most common types are:

- Support Vector Machine (SVM) [25].
- Relevant Vector Machine (RVM) [26].
- Artificial Neural Network (ANN) [27].

II. Literature Review

Vision based human activity recognition has been the most common dialogue in the field of research. Many publications are made on this topic like Aggarwal and Shangho[28] works on recognition of simple actions. Valera and Velastin[29] constructed an autonomous visual surveillance systems. Moeslund et al.[30] focuses on

- Human motion capture.
- Human model initialization.
- Tracking.
- Pose estimation.
- Action recognition.

Krüger et al.[31] works on various levels of complexities and Turaga et al.[32] recognizes activities on the lower level processing modules. Enzweiler and Gavrila[33] put all their resources in of walking video sequences. Candamo et al.[34] produces the research on recognition of human behaviors in changing scenerios. Aggarwal and Ryoo[35] focus on recognition of human group activities. Jiang et al.[36] emphasis on event recognition. Enzweiler and Gavrila[37] focus on the pedestrian recognition.

Zhenyu He, Lianwen Jin[37] develops a research on human activity recognition system with the help of tri-axis accelerometer. The system consists of:

- Discrete Cosine Transform (DCT).
- Principal Component Analysis (PCA).
- Support Vector Machine (SVM).

The features are extracted by the help of accelerometer and then data is processed by DCT. Now the extracted features are reduced by PCA in DCT domain. Then the invariant information is used for recognition of activity and Support Vector Machines are used to classify various human activities.

Ivan Laptev and Tony Lindeberg[38] investigates the local space time descriptors for classification of activities. Space time interest points are calculated and local space time events are represented. Many image descriptors are evaluated as spatio temporal neighborhoods. Video datasets of human activities are used for classification. Local position dependent histograms give advantage over different techniques.

Ivan Laptev and Tony Lindeberg[39] also works on image features and interest points to represent patterens. Spatio temporal events are detected using Harris and F¨orstner interest point operators. Descriptors are used to classify events.

James Davis and Gary Bradski[40] presents the research in Motion History Image (MHI) that add layers of consecutive image silhouettes into one template form. MHI extracts motion characteristics in real time scenarios.

III. Motion History Algorithm

Action Energy Image (AEI) is the key component in the motion history algorithm. AEI contains both characteristics of any action or activity such as:

- Structural characteristics.
- Motion characteristics.

The AEI representation of activities has the following advantages:

- ✓ Represents major shapes of silhouettes.
- ✓ Represents human motion sequence in a single image.
- ✓ Preserve average temporal information.
- ✓ Reduces the noise effects of background subtracted.
- ✓ No need for time alignment.
- ✓ Compact 2D representation of 3D spatio temporal information.
- ✓ Save storage space.

- ✓ Computationally efficient because of Fast Fourier Transform algorithms.
- ✓ Computation is extremely simple.

Once the AEI is computed then activities are easily classified by the help of classification algorithms. The classification algorithms are the machine learning techniques used in this research work. The most frequent technique used for classification of various activities is Support Vector Machine (SVM) and Discrete Cosine Transform (DCT). The following steps are the algorithm used to identify and classify various activities using machine learning techniques along with AEI of the image.

3.1. Pre processing Stage:

The pre processing includes RGB to Gray, background subtraction and foreground extraction. First of all convert the image frame from RGB to Gray level image as shown in figure 10.

Fig.10. RGB to Gray

Then, background subtraction technique is applied to compute the area of interest as shown in figure 11. Once the region of interest is segmented then area of interest will be crop from each and every frame of the video sequence.

Fig.11. Background Subtraction

3.2. Main Algorithm:

Therefore each and every frame in the video sequence is considered for the cropping and stored. The AEI can only be evaluated successfully if only area of interest is considered as shown in figure 12.

Fig.12. Crop region of the area of interest

Once all the cropped regions are obtained then add all of them to obtain the motion history of the image as shown in figure 13.

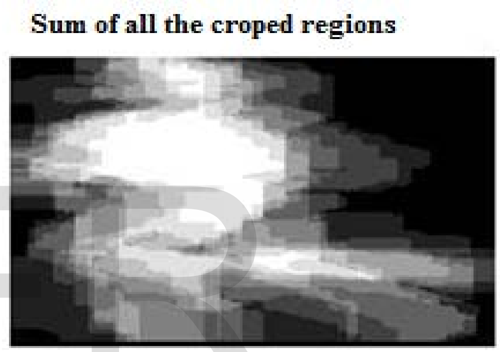

Fig.13. Sum of all the region of interests

3.3. Classification:

After the feature extraction is done we have to apply the classification algorithm in this research it is Discrete Cosine Transform (DCT). A discrete cosine transform (DCT) gives sum of cosine oscillation of frequencies instead of sequence of data points. DCTs act as a key component for the following fields:

- Science and engineering.
- Lousy compression of audio and images.
- Spectral methods for partial differential equations.

The cosine transform is more useful then sine transform in the following applications:

- ✓ Compression.
- ✓ Differential equations with boundary conditions.

DCT is related to Fourier transform and almost similar to discrete Fourier transform (DFT). But the only difference is the usage of real numbers only. DCTs are almost twice in length with DFT. DCT operates on real data with even symmetry. DCT have 8 standard variants in which 4 are commonly used. The general DCT equation for 2D image is:

ISSN 2229-5518

$$F(u,v) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \left(\frac{2}{M}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1}\sum_{j=0}^{M-1} \Lambda(i).\Lambda(j).cos\left[\frac{\pi.u}{2.N}(2i+1)\right] cos\left[\frac{\pi.v}{2.M}(2j+1)\right].f(i,j)$$

While the inverse DCT transform of this 2D image is F-1(u,v):

$$\Lambda(\xi) = \begin{cases} \frac{1}{\sqrt{2}} & \text{for}\,\xi = 0 \\ 1 & \text{otherwise} \end{cases}$$

The basic operation of the DCT is as follows:

- Input N * M image.
- Intensity of the pixel (i, j).
- DCT coefficient in (k1, k2) DCT matrix.
- Signal energy lies at low frequencies.
- Upper left corner of the DCT shows the energy.
- Compression is achieved.
- Each DCT input pixel is in gray scale level.

After applying the DCT as a Classifier to the region of interest, the output form will be as shown in figure 14.

Fig.14. DCT and Multiplication applied

And once we applied DCT, we apply the pseudo inverse technique to get the square matrix of the resultant, so that we can evaluate the determinant. And on the basis of the determinants, threshold is set so that various activities are recognized. The classification is shown in figure 15.

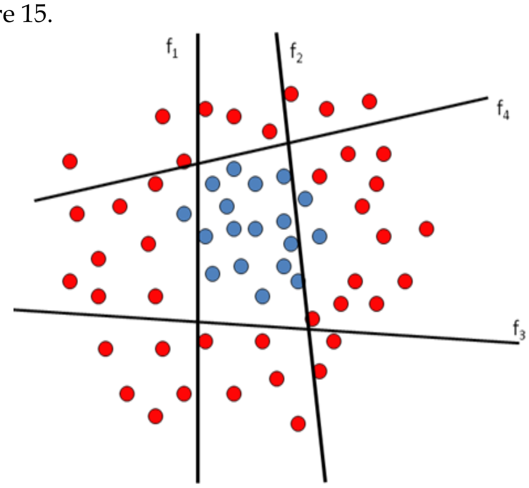

Fig.15 Classification

The block diagram of the motion history algorithm is shown in figure 16 and each stage is also represented with a separate block.

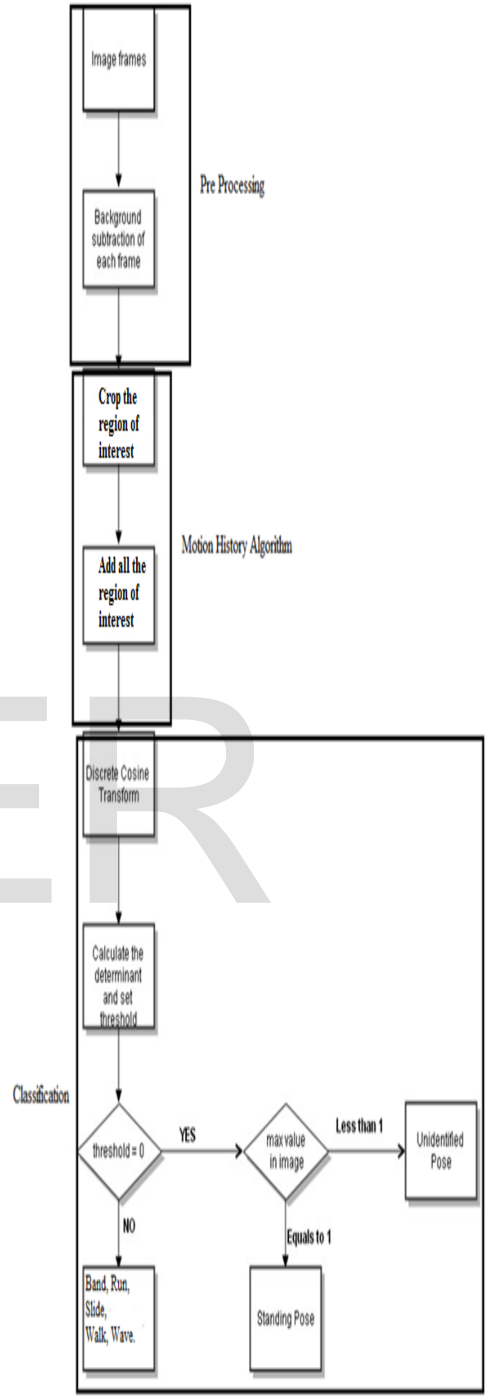

Fig.16. Block Diagram of Algorithm

.

IV. Results & Conclusion

In this research, 6 human activities are identified on various videos on single activity which are discussed below. There output results are shown along with their threshold.

4.1. Bending:

ISSN 2229-5518

In case of bending, the threshold is less than 1e-104. The output is shown in figure 17 in which the activity along with the motion history is shown.

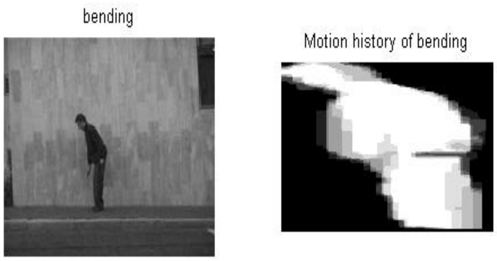

Fig.17. Bending Output

4.2. Running:

In case of running, the threshold is greater than 1e17 && threshold is less than 1e60. The output is shown in figure 18 in which the activity along with the motion history is shown.

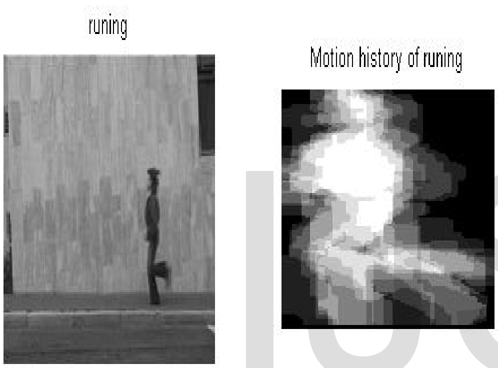

Fig.18. Running Output

4.3. Sliding:

In case of Sliding, the threshold is less than 1e-10 && threshold is greater than 1e-55. The output is shown in figure 19 in which the activity along with the motion history is shown.

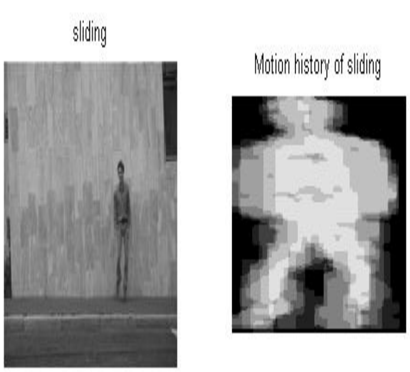

Fig.19. Sliding Output

4.4. Standing:

In case of standing, the threshold is equals to zero and then the maximum value of the motion history image is also 1. The output is shown in figure 20 in which the activity along with the motion history is shown.

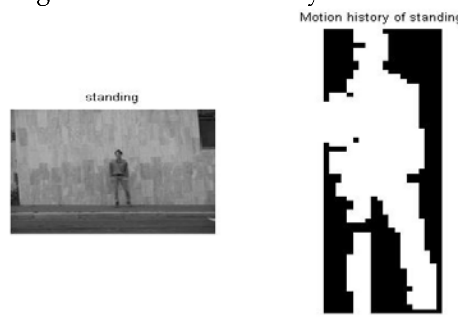

Fig.20. Standing Output

4.5. Walking:

In case of bending, the threshold is greater than 1e17 && threshold is less than 1e60. The output is shown in figure 21 in which the activity along with the motion history is shown.

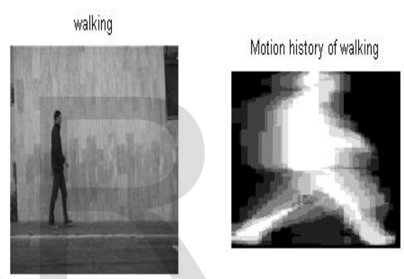

Fig.21. Walking Output

4.6. Waving both hands:

In case of waving both hands, the threshold is greater than 1e-10 && threshold is less than 1e17. The output is shown in figure 22 in which the activity along with the motion history is shown.

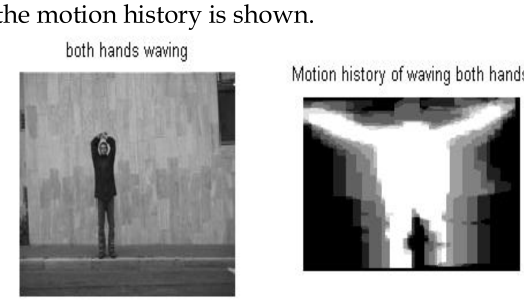

Fig.22. Waving Both Hands Output

4.7. Conclusion:

The final results of the 6 activities which are recognized and classified as well are shown in the confusion matrix in table 1. The diagonal of the confusion matrix represents the success rate of each and every activity that has been recognized and classified.

ACTIVITY	BEND	RUN	SLIDE	STAND	WALK	WAVE 2	UNIDENTFIED
BEND	0.70	0.00	0.10	0.00	0.00	0.00	0.20
RUN	0.10	0.70	0.10	0.00	0.00	0.10	0.00
SLIDE	0.10	0.00	0.70	0.00	0.10	0.00	0.10
STAND	0.00	0.00	0.00	0.89	0.00	0.00	0.11
WALK	0.00	0.20	0.10	0.00	0.70	0.00	0.00
WAVE 2	0.00	0.00	0.00	0.00	0.00	0.70	0.30

Table.1. Confusion Matrix

The overall success rate in using motion history algorithm along with classification technique is 73%. Since the success rate in using Silhouette extraction technique using star skeleton algorithm is less than expected because of the usage of the correlation. And silhouette extraction using Ellipse technique has the drawback that is works only on the still activities like standing, sitting and lying etc. The ellipse technique will not recognize the moving activities like walking, running etc.

So in a nutshell the human activity recognition using Machine Learning techniques can be achieved with the success rate of 70+% by the help of the motion history algorithm.

V. References

1. Wren, C.R.; Azarbayejani, A.; Darrell, T.; Pentland, A.P. Pfinder: Real-time tracking of the human body. IEEE Trans. Pattern Anal. Mach. Intell. 1997, 19, 780–785.

2. Cucchiara, R.; Grana, C.; Piccardi, M.; Prati, A. Detecting moving objects, ghosts, and shadows in video streams. IEEE Trans. Pattern Anal. Mach. Intell. 2003, 25, 1337–1342.

3. Seki, M.; Fujiwara, H.; Sumi, K. A Robust Background Subtraction Method for Changing Background. In Proceedings of Fifth IEEE Workshop on Applications of Computer Vision, Palm Springs, CA, USA, 4–6 December 2000; pp. 207–213.

4. Permuter, H.; Francos, J.; Jermyn, I. A study of Gaussian mixture models of color and texture features for image classification and segmentation. Pattern Recogn. 2006, 39, 695–706.

5. Yoon, S.; Won, C.S.; Pyun, K.; Gray, R.M. Image Classification Using GMM with Context Information and with a Solution of Singular Covariance Problem. In IEEE Proceedings of Data Compression Conference (DCC), Snowbird, UT, USA, 25-27 March 2003; p. 457.

6. Horprasert, T.; Harwood, D.; Davis, L.S. A statistical approach for real-time robust background subtraction & shadow detection. IEEE ICCV 1999, 99, 1–19.

7. Brendel, W.; Todorovic, S. Video Object Segmentation by Tracking Regions. In proceedings of IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 833–840.

8. Yu, T.; Zhang, C.; Cohen, M.; Rui, Y.; Wu, Y. Monocular Video Foreground/Background Segmentation by Tracking Spatial-color Gaussian Mixture Models. In Proceedings of IEEE Workshop on Motion and Video Computing (WMVC), Austin, TX, USA, 23–24 February 2007; p. 5.

9. Murray, D.; Basu, A. Motion tracking with an active camera. IEEE Trans. Pattern Anal. Mach. Intell. 1994, 16, 449–459.

10. Kim, K.K.; Cho, S.H.; Kim, H.J.; Lee, J.Y. Detecting and Tracking Moving Object Using an Active Camera. In Proceedings of IEEE 7th International Conference on Advanced Communication Technology (ICACT), Phoenix Park, Dublin, Ireland, 21–23 February 2005; Volume 2, pp. 817–820.

11. Daniilidis, K.; Krauss, C.; Hansen, M.; Sommer, G. Real-time tracking of moving objects with an active camera. Real-Time Imaging 1998, 4, 3–20.

12. Huang, C.; Chen, Y.; Fu, L. Real-time Object Detection and Tracking on a Moving Camera Platform. In Proceedings of IEEE ICCAS-SICE, Fukuoka, Japan, 18–21 August 2009; pp. 717–722.

13. Shechtman, E.; Irani, M. Space-time Behavior Based Correlation. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 405–412.

14. Kumari, S.; Mitra, S.K. Human Action Recognition Using DFT. In Proceedings of the third IEEE National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Hubli, India, 15–17 December 2011; pp. 239–242.

15. Sedai, S.; Bennamoun, M.; Huynh, D. Context-based Appearance Descriptor for 3D Human Pose Estimation from Monocular Images. In Proceedings of IEEE Digital Image Computing: Techniques and Applications (DICTA), Melbourne, VIC, Australia, 1–3 December 2009; pp. 484–491.

16. Lowe, D.G. Object Recognition from Local Scale-invariant Features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–25 September 1999; Volume 2, pp. 1150–1157.

17. Lowe, D.G. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 2004, 60, 91–110.

18. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of IEEE

Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 886–893.

19. Müller, M.; Röder, T.; Clausen, M. Efficient content-based retrieval of motion capture data. ACM Trans. Graph. (TOG) 2005, 24, 677–685.

20. Hoang, L.U.T.; Ke, S.; Hwang, J.; Yoo, J.; Choi, K. Human Action Recognition based on 3D Body Modeling from Monocular Videos. In Proceedings of Frontiers of Computer Vision Workshop, Tokyo, Japan, 2–4 February 2012; pp. 6–13.

21. Hoang, L.U.T.; Tuan, P.V.; Hwang, J. An Effective 3D Geometric Relational Feature Descriptor for Human Action Recognition. In Proceedings of IEEE RIVF International Conference on Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), Ho Chi Minh City, Vietnam, 27 February–1 March 2012; pp. 1–6.

22. Veeraraghavan, A.; Roy-Chowdhury, A.K.; Chellappa, R. Matching shape sequences in video with applications in human movement analysis. IEEE Trans. Pattern Anal. Mach. Intell. 2005, 27, 1896–1909.

23. Hoang, L.U.T.; Ke, S.; Hwang, J; Tuan, P.V.; Chau, T.N. Quasi-periodic Action Recognition from Monocular Videos via 3D Human Models and Cyclic HMMs. In Proceedings of IEEE International Conference on Advanced Technologies for Communications (ATC), Hanoi, Vietnam, 10–12 October 2012; pp. 110–113.

24. Murphy, K.P. Dynamic Bayesian networks: Representation, inference and learning. PhD diss., University of California, Berkeley, CA, USA, 2002.

25. Vapnik, V.; Golowich, S.E.; Smola, A. Support vector method for function approximation, regression estimation, and signal processing. Adv. Neural Inf. Process. Syst. 1997, 9, 281–287.

26. Tipping, M.E. Sparse Bayesian learning & the relevance vector machine. J. Mach. Learn. Res. 2001, 211-244.

27. Fiaz, M.K.; Ijaz, B. Vision based Human Activity Tracking using Artificial Neural Networks. In Proceedings of IEEE International Conference on Intelligent and Advanced Systems (ICIAS), Kuala Lumpur, Malaysia, 15–17 June 2010; pp. 1–5.

28. Aggarwal, J. K.; Park, S. Human Motion: Modeling and Recognition of Actions and Interactions. In Proceedings of IEEE 2nd International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), Thessaloniki, Greece, 6–9 September 2004; pp. 640–647.

29. Valera, M.; Velastin, S.A. Intelligent distributed surveillance systems: A review. IEE Proc. Vis. Image Signal Process. 2005, 152, 192–204.

30. Moeslund, T.B.; Hilton, A; Krüger, V. A survey of advances in vision-based human motion capture and analysis. Comput. Vis. Image Underst. 2006, 104, 90–126.

31. Krüger, V.; Kragic, D.; Ude, A.; Geib, C. The meaning of action: A review on action recognition and mapping. Adv. Robot. 2007, 21, 1473–1501.

32. Turaga, P.; Chellappa, R.; Subrahmanian, V.S.; Udrea, O. Machine recognition of human activities: A survey. IEEE Trans. Circuits Syst. Video Technol. 2008, 18, 1473–1488.

33. Enzweiler, M.; Gavrila, D.M. Monocular pedestrian detection: Survey and experiments. IEEE Trans. Pattern Anal. Mach. Intell. 2009, 31, 2179–2195.

34. Candamo, J.; Shreve, M.; Goldgof, D.B.; Sapper, D.B.; Kasturi, R. Understanding transit scenes: A survey on human behavior-recognition algorithms. IEEE Trans. Intell. Transp. Syst. 2010, 11, 206–224.

35. Aggarwal, J.K.; Ryoo, M.S. Human activity analysis: A review. ACM Comput. Surv. (CSUR) 2011, 43, 16.

36. Jiang, Y.; Bhattacharya, S.; Chang, S.; Shah, M. High-level event recognition in unconstrained videos. In International Journal of Multimedia Information Retrieval, 2013, 2, 73–101

37. Enzweiler, M.; Gavrila, D.M. Monocular pedestrian detection: Survey and experiments. IEEE Trans. Pattern Anal. Mach. Intell. 2009, 31, 2179–2195.

38. Ivan Laptev and Tony Lindeberg: "Local Descriptors for Spatio-Temporal Recognition", Spatial Coherence for Visual Motion Analysis Lecture Notes in Computer Science Volume 3667, 2006, pp 91-103.

39. Ivan Laptev and Tony Lindeberg: "Space-time Interest Points", International Journal of Computer Vision archive Volume 64 Issue 2-3, September 2005, pp 107 – 123.

40. James Davis and Gary Bradski: "Motion Segmentation and Pose Recognition with Motion History Gradients", Machine Vision and Applications - Special issue: IEEE WACV archive Volume 13 Issue 3, July 2002, Pages 174 – 184.