Human Actions Tracking and Recognition Based on Body Parts Detection via Artificial Neural Network

Amir Nadeem
Dept. of Computer Science
Air University
Islamabad, Pakistan
171269@students.au.edu.pk

Ahmad Jalal
Dept. of Computer Science
Air University
Islamabad, Pakistan
ahmadjalal@mail.au.edu.pk

Kibum Kim
Dept. of Computer Science
Hanyang University
South Korea
kibm@hanyang.ac.kr

Abstract*— Human body action recognition has drawn a good deal of interest in the community of computer vision, owing to its wide range of applications. Recently, the video / image sequence base action recognition techniques are believed to be ideal for its efficiency and lower cost compared to other techniques such as the ambient sensor and the wearable sensor. However, given to a large amount of variation in human pose and image quality, reliable detection of human action is still a very challenging job for scientists. In this document, we used linear discriminant analysis for the generation of features from the body parts detected. The primary goal of this study is to combine linear discriminant analysis with an artificial neural network for precise human action detection and recognition. Our proposed mechanism detects complicated human actions in two state-of-the-art datasets, i.e. KTH-dataset and Weizmann Human Action. We obtained multidimensional features from twelve body parts, which are estimated from body models. These multidimensional characteristics are used as inputs for the artificial neural network. To access the efficiency of our suggested method, we compared the outcomes with other state-of-the-art classifiers. Experimental results show that our proposed technique is reliable and applicable in health exercise systems, smart surveillance, e-learning, abnormal behavioral detection, protection for child abuse, care of the elderly people, virtual reality, intelligent image retrieval and human computer interaction.

Keywords— Artificial neural network, body parts detection, human action recognition, linear discriminant analysis.

I. INTRODUCTION

Human action recognition is the process of transforming human actions to digital actions. This is an important and difficult problem in the subject of computer vision, which is wildly used in human detection, image retrieval, human behavior analysis [1], as well as human motion capture [2],[41]. Human action can be detected by estimating parts of human body and dividing the body around these body parts. One such division may include head, both shoulders, hands, torso, hips, feet and knees. Once this segmentation is achieved, the human action can be detected accurately using segmented body parts as input for linear discriminant analysis [3] to generate multidimensional features. Two main techniques are used widely to detect human action [31]-[39]. In first technique machine learning [4] is directly applied on raw image sequence to detect action. But this technique is too complex because there is so much noise and training time may be infinity long. In 2nd technique, human action is detected by using features [5],[40], which are extracted from human body parts.

II. RELATED WORK

In the human action recognition field, many researchers have proposed different methods to contribute in the improvement of society. To segment foreground, R. Achant et al. classified all regions of an image as salient region and non-salient regions [6]. A. Jalal et al. estimated and detected human body parts in physical sports movements [7]. In other work, two body parts model are used to detected body parts. Action detection using markov model on translation and scale-invariant features proposed by S. Kamal et al. [8]. A. Jalal et al. used Multi-features descriptors for human action tracking and detection in different environments [9]. A new ANN technique which accumulates a key-points tracker and an object detector is proposed by V. T. Hoang et al. [10]. G. Liang et al. proposed a limb based graphical model for detection of human action [11]. Due to several limitations in existing works, we proposed method in which foreground is segmented to detect body parts. These body parts are used to generate multidimensional features which are input to ANN for action recognition.

The rest of the paper is organized as following: Section III presents an overview of the proposed technique which comprises of foreground segmentation, body parts detection, features extraction by using LDA, and ANN for human action detection. Section IV is used to describe experimental results and compared with other state-of-the-art human action detection systems. Finally, in Section V paper is concluded.

III. PROPOSED SYSTEM METHODOLOGY

This section elaborates the proposed methodology of human action recognition system. Weizmann Human action [12] dataset and KTH-dataset [13], have been chosen to evaluate the performance of proposed method. The human action recognition system is carried out in four modules where each module has multiple phases as; (1) Foreground segmentation, (2) Body parts detection, (3) Features extraction and (4) Artificial Neural Network. During testing on our both datasets, learning parameters from the training module are used for distinguishing each action as a separate pose. Fig. 1 illustrates the overview of proposed system methodology.

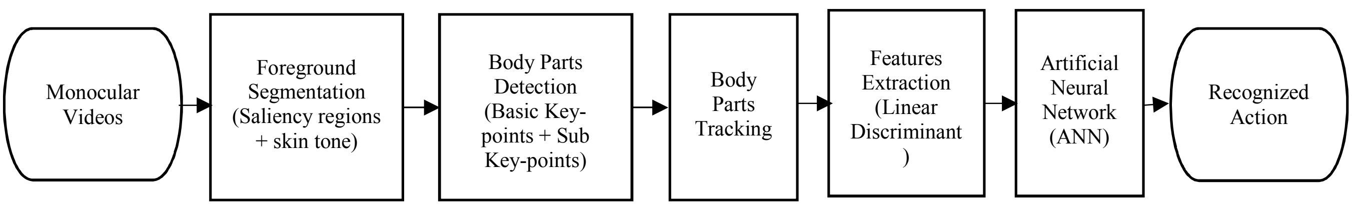

Fig. 1. Flow diagram of the human action recognition system

978-1-7281-4235-7/20/$31.00 ©2020 IEEE

A. Foreground Segmentation:

Foreground segmentation via skin tone is achieved using heuristic thresholding [14] which is an approach of image transform operation. It identifies skin regions with the help of suitable YCC model. Conversion from RGB color model to YCC model is illustrated with the help of Equations (1-3);

$$Y = 16 + \frac{1}{255}(65.48R + 128.55G + 24.97B) \quad (1)$$

$$C_r = 128 - \frac{1}{255}(37.80R - 74.20G + 112.44B) \quad (2)$$

$$C_b = 128 + \frac{1}{255}(112.44R - 94.15G - 18.28B) \quad (3)$$

An algorithm has been created that takes advantage of human skin color's spatial features [15]. A skin color map is obtained and it is used to identify pixels that appear to be skin on the input image's chrominance parts. It is discovered that the range of both, Cb and Cr for the most skin-color reference map are:

$$77 \leq C_b \leq 127 \; and \; 133 \leq C_r \leq 173 \quad (4)$$

Our aim is to discover human skin from distinct races. The above-mentioned thresholds work only with the skin of a Caucasian people because only individuals with white skin can be found with the first threshold. For this reason, a new threshold is proposed to segment people within the image as shown by equating (5):

$$80 \leq C_b \leq 120 \; and \; 133 \leq C_r \leq 173 \quad (5)$$

To enhance the foreground detection result, salient areas are identified in image that are visually more noticeable as compare to other areas in the image [16]. This method is effective on a wide range of images including video frames, paintings, and pictures containing noisy information etc. Saliency is calculated at different levels as the local contrast of an image region compared to its neighboring areas [17]. This contrast based saliency value $S_{i,j}$ for any pixel at location (i, j) in the given image is calculated as the distance l between the average characteristic vectors of the inner region $R1$ and the outer region $R2$ as following:

$$s_{i,j} = l\left[\left(\frac{1}{n_1}\sum_{p=1}^{n_1} c_p\right), \left(\frac{1}{n_2}\sum_{q=1}^{n_2} fc_q\right)\right] \quad (6)$$

where n_1 and n_2 are the number of pixels in region $R1$ and $R2$. The distance l is a Euclidean distance; if c is the characteristic vector having correspondence and uncorrespondence measurements (See Fig. 2).

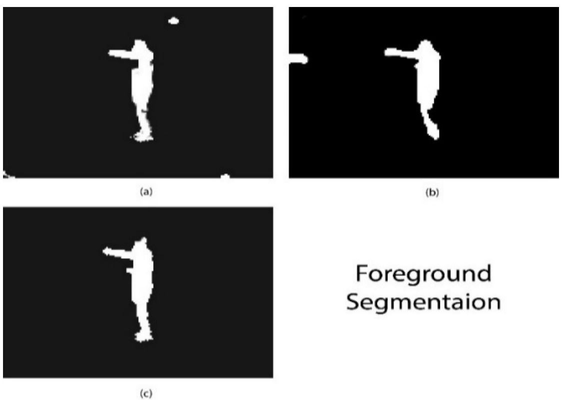

Fig. 2. . Foreground Segmentation. (a) foreground detected via skin tone method (b) foreground detection via saliency method and (c) combined result

Forgruound detected via skin tone method is combined with forground detected via saliency method, to get a better result.

B. Body Parts Detection:

In body basic key-points, five body parts are detected which includes head, hands and feet. Human is detected from silhouette as given in the Algorithm 1. Distance from the upper side of the human silhouette to the feet is calculated to estimate overall height of the body.

Algorithm 1 Extraction of human from silhouette

Input: Y: Monocular Image Pixels, $B_{0:}$ initial floor normal l_0: initial floor distance

Output: Detected Human

/* Background removal
$Y_f = \{ Y_n \mid Y_n B_0 - l_0 < \delta_f \}$;
i = i + 1;
Repeat
l_i -1 = $Y_f \mathbf{B}_i - 1$;
$B_i = (Y_f \mid Y_f)^{-1} Y^t_f \, l_i$-1;
$Y_f = \{ Y_n \mid Y_n A_i - l_i - 1 < \delta f \}$;
i= i+1;
until Y_f is altered.
$Y = Y/Y_f$;
/*Connected part specifying
For j=1 to |Y| **do**
If Yj is vector that is non zero.
Neighbor=combined pixels with the same pixel value;
If neighbor pixel is empty, **then**
Linked [nextpixel]=label j= Next pixel;
Nextpixel+=1;
Else
Pixels j =min (neighbours labels);
Linked [pixel] union (linked [pixel], L;
End
End
End
/* Human Detection
For j=1 to |Y| do
If Yj is vector with all non-zero values, then
$Pixel_j$ = find ($pixel_j$);
End
End

Human head width is approximately 1/8.5 of the height. Using this lead, head width ratio to the overall height is determined by dividing number of pixels from head to the feet, with number of pixels counted horizontally in upper area. Limbs are detected with the help of forward kinematics technique [18]. Joint space for the kinematics technique is calculated as:

$$Y_i^p = Y_{i-1}^p + \left(R_{i-1}^{p-1} \ldots.. R_0^{p-1}\right).\left(Y_i^{p-1} - Y_{i-1}^{p-1}\right) \quad (7)$$

In the equation, Y_i^p is ith joint position of given limb Y in the specific frame p. The portion $\left(R_{i-1}^{p-1} \ldots.. R_0^{p-1}\right)$ is the combination of the rotation matrix of the joints space. Lowest points on the lower limbs are designated as feet. In both

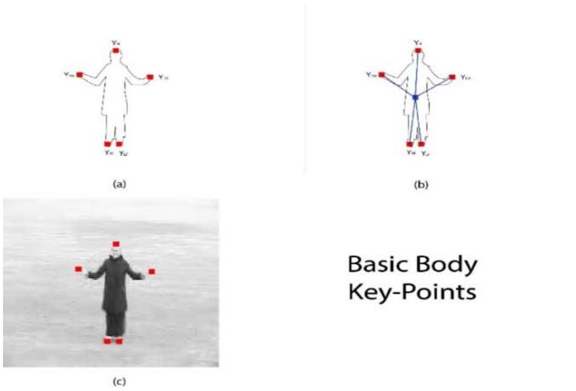

Fig. 3. Basic key-points. (a) Body basic key-points representation (b) Key-points connected (c) key points on original image

directions, these limbs are searched to locate the feet up to a certain height as shown in Fig. 3.

If point is discovered on the left side of the second foot, then right foot is called as first foot and the left foot as second foot and vice versa. If there is only one foot than there is occlusion. The detection of the upper limbs is used to direct the detection of the lower arms to detect both hands.

Seven more key-points can be detected on the body after five fundamental body key-points. Torso point is determined with the help of two-foot and upper-head center. Hips points are estimated with the help of torso points (See Fig. 4).

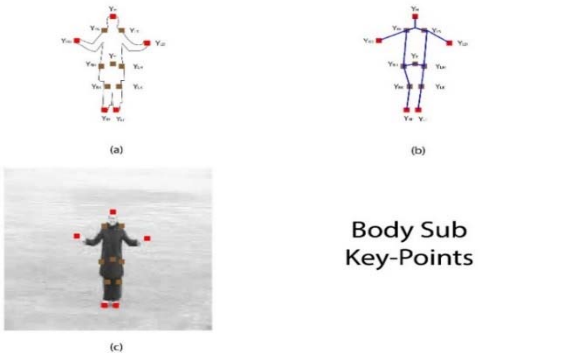

Fig. 4. Body parts models (a) All twelve key-point with body models. (b) Connected key-points, and (c) Representation on real image.

The aspect ratio data of both torso height and width is fixed to their surroundings to estimate hips points. While, each of these points has a specific distribution that is averaged to generate torso points along with other points. These points are used during initialization parameters to calculate the anticipated length and joint angle [19].

The two shoulders points are identified by converging an appropriate function. The function domain is an isosceles triangle whose base vertices are the two shoulder points and its main vertex is the neck point. When the base of the triangle is maximum, the respective function is then minimized and vice versa. Using equation (7), position of the limbs is determined already. We also took into consideration the position of the upper limbs [20] to find the precise place of the shoulder point. At the end, by considering the center of the hip and the foot, we recognized knee. Finally, knee-point estimation is formulated as:

$$Y_K^p = \frac{Y_F^p - Y_{Hip}^p}{2} \tag{8}$$

The point Y_K^p is the locality of the knee, Y_F^p is the locality of the respected foot and Y_{Hip}^p is the position of the hips.

C. Tracking Body Parts:

Body key-points tracking is very helpful when there is occlusion or complex images [21]. Tracking is only possible when there is sequence of frames or video, and in this case there is sequence of images. Body parts are tracked by calculating the difference between the position of key-point in current frame and its position in previous frame as:

$$Y^p = Y^{p-1} + \Delta Y^{p-1} \tag{9}$$

Here, Y^p is position of the key-point on human silhouette at any given frame p and it is calculated by finding the distance between sequences of frames of the given dataset.

D. Feature Generation:

Accurate detection of features is very impotent stage for action recognition [47]-[52]. LDA determines low-dimensional characteristics from high-dimensional feature space that have the most delicate discriminating capacity [22], [42,46]. These characteristics can combine the same categories and separate distinct sample kinds as much as possible. This means choosing the features which make the greatest percentage difference between inter-class scatter and intra-class scatter. It is defined as.

$$W_s = \sum_{j=1}^{t} p(j)E\left\{(m_j - x)(m_j - x)^T \middle| x \in class\ j\right\} \tag{10}$$

$$B_s = \sum_{j=1}^{t} p(j)(m_j - m)(m_j - m)^T \tag{11}$$

In the above equations, m_j denotes the mean data of class j and m denotes the mean of the complete data. In equation (10), W_s denotes within class scatter and in equation (11) B_s denotes between class scatter. Since the optimum direction of projection within each class should be with minimized deviation and at the same moment projection between all classes should be with maximized deviation. Fig.5 shows the generation of features for the KTH-Dataset.

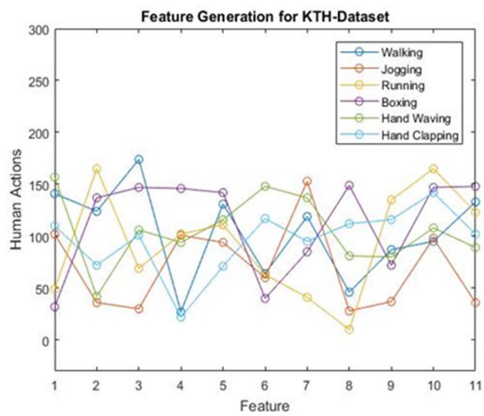

Fig. 5. Multidimensional features extraction for 6 classes of KTH Multiview Football dataset

New 22-dimensional features are generated in the proposed method [23]. These feature are divided into three groups: (1) Torso features which are six dimensional, (2) First degree features which are eight-dimensional and (3) Second

Fig. 3. Basic key-points. (a) Body basic key-points representation (b) Key-points connected (c) key points on original image degree features which are also eight-dimensional. Fig. 6 shows that how linear discriminant analysis discriminates actions on the basis of features.

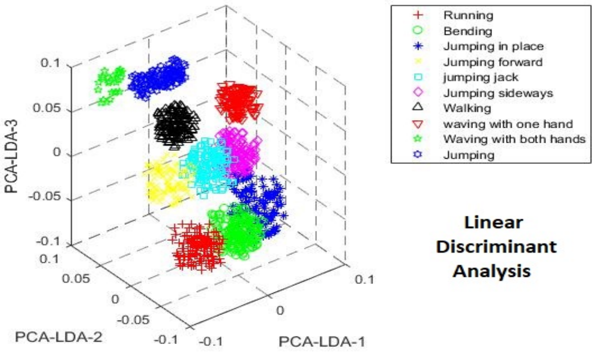

Fig. 6. Discriminating features with linear discriminant analysis.

E. B. *Artificial Neural Network:*

Finally, these multidimensional features are fed as input to artificial neural network for human action detection [24]. Artificial neural network comprises of various layers of neurons as input layer, hidden layers and the layer of output.

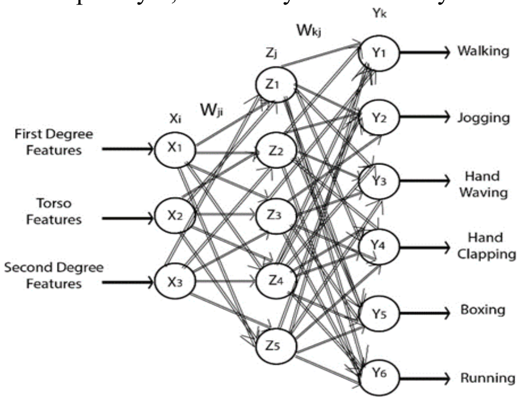

Fig. 7 Artificial Neural Network for KTH Multiview Football dataset

There are also weighted connections that links neural network neighboring layers and a transfer function in each neuron to calculate net input. If $w_{i,j}$ is the weights to connections p_j is the partiality, and x_i represents input multidimensional features, then transforming function F_j can be written as;

$$F_j = \sum_i w_{i,j} \times x_i + p_j \qquad (12)$$

A three-layer artificial neural network is implemented and supplied at the input layer with outward, neighboring, distinctive and enclosed characteristics. During the . Outputs probability distribution for all interactions can be seen in the equation (13) as;

$$\sigma(F)_j = \frac{e^{F_j}}{\sum_{n=1}^{N} e^{F_j}} \,, j = 1, \dots \dots, N \qquad (13)$$

Here, $\sigma(F)_j$ is the soft max function and F_j shows the composition of input features multiplied with their corresponding weights. Resulting layer estimates the poses for KTH-Dataset as walking, jogging, hand waving, hand clapping, boxing and running. Fig. 7 depicts the designed artificial neural network for human interaction recognition over six interaction classes.

IV. Experimental Results

In the proposed method, two datasets are used. Datasets are Wizmann Human Action and KTH-dataset.

A. Body Parts Recognition::

Weizmann Human Action dataset comprises of a ninety low quality videos of 180 × 144 resolution. Some examples of Weizmann dataset is shown in Fig. 8.

Fig. 5. Some samples from the Weizmann Dataset. (a) Jogging (b) Walking (c) Waving Both Hands

This is an easily available dataset, in which 9 subjects are performing ten actions. The dataset involves activities including running, bending, jumping in place, jumping forward, jumping jack, jumping sideways, walking, waving with one hand, waving with two hands and jumping on one leg. In TABLE I, we calculated body parts detection accuracy results with respect to Weizmann dataset by defining a local threshold value around the labeled body parts. If the detected body part lies within that threshold it is accurately detected. On Weizmann dataset our proposed method shows significant body parts detection accuracy of 89.41%.

Table I. Representing accuracy results for Weizmann dataset

Body Parts	Distance from ground truth	Detection Accuracy (%)
Upper head	10.0	89
Right Hand	9.6	91
Left Hand	9.8	94
Right Shoulder	5.3	94
Left Shoulder	13.2	92
Right Hip	10.7	81
Left Hip	10.8	91
Right Knee	13.6	82
Left Knee	11.7	82
Torso Point	10.8	88
Right Foot	9.8	94
Left Foot	10.4	95
Mean Detection Accuracy rate =89.41 %		

While, in TABLE II, accuracy of our method for body parts detection is compared with three state of the art methods. Results shows that our method performs better on both datasets than existing methods as given in TABLE I

Table II. Comparison of results for Body Parts Detection

Methods	Weizmann dataset	KTH-Dataset
S. Hong, et al. [25]	76.6	80.1
C. Dorin et al. [26]	81.9	79.2
S.Gomathi et al. [27]	82.01	83.0
Proposed Method	89.41	86.67

B. Action Recognition:

There are 6 types of videos in the KTH-dataset. Twenty-five subjects have performed these actions in different scenes having frame rate of 25fps. Spatial resolution for each video sequence is 160×120 pixels. In Fig. 9 there are some samples from the KTH-dataset.

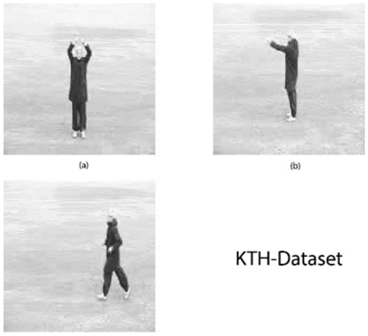

Fig. 6. Some samples from the KTH-Dataset. (a) Hand Waving (b) Boxing (c) Jogging

In this proposed approach, for both datasets, one third of image sequences are kept for validation purpose and one third for testing purpose. Rest are given to proposed model for training. Results are illustrated by making the confusion matrices with actual class label. Table III represents the confusion matrix for KTH-dataset, where results are 100% accurate for running with ball, dribbling and standing because these actions are unique and our multidimensional features are able to classify them easily. Additionally, there is some confusion in taking pass, shooting, running and passing. Mean recognition accuracy for this dataset is 87.57%. To understand table, I, you should understand these contractions;

Table III Representing Confusion matrix for KTH-Dataset

	SH	RB	R	TP	DB	ST	P
SH	0.83	0.0	0.0	0.07	0.0	0.0	0.10
RB	0.0	1.0	0.0	0.0	0.0	0.0	0.0
R	0.0	0.11	0.89	0.0	0.0	0.0	0.0
TP	0.12	0.0	0.0	0.70	0.0	0.0	0.18
DB	0.0	0.0	0.0	0.0	1.0	0.0	0.0
ST	0.0	0.0	0.0	0.0	0.0	1.0	0.0
P	0.0	0.0	0.0	0.0	0.0	0.0	0.71

shooting=SH, running with ball=RB, running=R, taking pass=TP, dribbling=DB, standing=ST, passing=P

Results are also compared with state of the arts method. In table IV there is comparison of accuracy of our proposed method on 3 state of the art methods.

Table IV. Comparison of proposed technique with state of the art methods

Approaches	KTH-Dataset	Weizmann Human Action Dataset
M. S. Ryoo et al. [28]	71.2%	67.5%
Y. Yang et al. [29]	72.5%	75.6%
M. Mahmood et al. [30]	83.5%	72.5%
Proposed Method	87.57%	86%

V. CONCLUSION

The proposed paper presents a new methodology for human action detection and recognition using a combination of linear discriminant analysis and artificial neural network. Human body parts are detected from the human silhouette and these body parts are used to generate multidimensional features with the help of a linear discriminant analysis. These features are inputs to an artificial neural network which recognizes human actions. The Weizmann Human Action dataset and KTH-Dataset validate the precision of our technique. By comparing the proposed technique with state-of - the-art techniques, it is shown that the proposed technique is better suited to complicated human actions. In our future work, we aim at optimizing body parts and using a quadratic discriminant analysis for the generation of features to see the distinction in outcomes.

CONCLUSION

The proposed paper presents a new methodology for human action This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by Ministry of Education (No. 2018R1D1A1A02085645).

REFERENCES

[1] A. Jalal, N. Sharif, J. T. Kim, and T.S. Kim, “Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart homes,” Indoor and Built Environment, Vol. 22, pp. 271-279, 2013.

[2] A. Jalal, S. Kamal and D.-S. Kim, "Detecting Complex 3D Human Motions with Body Model Low-Rank Representation for Real-Time Smart Activity Monitoring System," KSII Transactions on Internet and Information Systems, vol. 12, no. 3, pp. 1189-1204, 2018.

[3] T. K. Kim, and J. Kittler, “Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume. 27 , Issue. 3, March 2005.

[4] A. Jalal, S. Kamal, and D. Kim, “A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments,” Sensors, vol. 14(7), pp. 11735-11759, 2014.

[5] A. Jalal, Y. Kim, and D. Kim, “Ridge body parts features for human pose estimation and recognition from RGB-D video data,” in Proceedings of the IEEE International Conference on computing, communication and networking technologies, pp. 1-6, 2014.

[6] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, “Salient Region Detection and Segmentation,” International Conference on Computer Vision Systems (ICVS ’08), Vol. 5008, Springer Lecture Notes in Computer Science, pp. 66-75, January 2008.

[7] A. Jalal, A. Nadeem, and S. Bobasu, “Human Body Parts Estimation and Detection for Physical Sports Movements,” 2nd International Conference on Communication, Computing and Digital systems (C-CODE), March 2019.

[8] M. Z. Uddin, W Khaksar, and J. Torresen, “Activity Recognition Using Deep Recurrent Neural Network on Translation and Scale-Invariant Features,” 25th IEEE International Conference on Image Processing (ICIP), October 2018.

[9] A. Jalal, M. Mahmood, and A. S. Hasan, “Multi-features descriptors for Human Activity Tracking and Recognition in Indoor-Outdoor Environments,” 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), January 2019.

[10] V. T. Hoang, and K. H. Jo, “Multi-Person Pose Estimation with Human Detection: A Parallel Approach,” 44th Annual Conference of the IEEE Industrial Electronics Society, October 2018.

[11] G. Liang, X. Lan, J. Wang, Ji. Wang, and N. Zheng, “A Limb-Based Graphical Model for Human Pose Estimation,” IEEE Transactions on Systems, Man, and Cybernetics: Systems, Volume. 48 , Issue. 7 , July 2018.

[12] Treves, and Haim, “Weizmann Dataset,” Mendeley Data, v1, July 2017.

[13] C. Schuldt, I. Laptev, and B. Caputo, “Recognizing human actions: a local SVM approach,” Proceedings of the 17th International Conference on Pattern Recognition, August, 2004.

[14] Maheswari, and R. Korah, “Enhanced skin tone detection using heuristic thresholding,” Biomedical Research (2017) Volume. 28, Issue. 9, February 2017.

[15] M. A. K. Quaid and Ahmad Jalal, “Wearable Sensors based Human Behavioral Pattern Recognition using Statistical Features and Reweighted Genetic Algorithm, Multimedia Tools and Applications, 2019.

[16] A. Ahmed, A. Jalal, and A. A. Rafique, “Salient Segmentation based Object Detection and Recognition using Hybrid Genetic Transform”, IEEE ICAEM conference, 2019.

[17] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, “Saliency filters: Contrast based filtering for salient region detection,” IEEE Conference on Computer Vision and Pattern Recognition, June 2012.

[18] D. Zhang, and Z. Gaoa, “Forward kinematics, performance analysis, and multi-objective optimization of a bio-inspired parallel manipulator,” Robotics and Computer-Integrated Manufacturing, Volume. 28, Issue. 4, PP 484-492, August 2012.

[19] I. Akhter, and M. J. Black, “Pose-Conditioned Joint Angle Limits for 3D Human Pose Reconstruction,” The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1446-1455, June 2015.

[20] G. Liang, X. Lan, J. Wang and N. Zheng, “Human pose estimation based on human limbs,” 23rd International Conference on Pattern Recognition (ICPR), December 2016.

[21] J. Tian, L. Li, and W. Liu, “Monocular Human Motion Tracking with Non-Connected Body Part Dependency,” International Conference on Digital Image Computing: Techniques and Applications (DICTA), November 2015.

[22] Z. Li, Y. Fu,J. Yuan, T, S. Huang, and Y. Wu, “Query Driven Localized Linear Discriminant Models for Head Pose Estimation,” IEEE International Conference on Multimedia and Expo, July 2007.

[23] Y. Kim, and D. Kim, “Real-time dance evaluation by markerless human pose estimation,” Multimedia Tools and Applications, Volume. 77, Issue. 23, pp 31199–31220, December 2018.

[24] Y. Liu, Y. Xu, and S. B. Li, “2-D Human Pose Estimation from Images Based on Deep Learning: A Review,” 2nd IEEE Advanced Information Management,Communicates,Electronic and Automation Control Conference (IMCEC), May 2018.

[25] S. Hong, and M. Kim, “A Framework for Human Body Parts Detection in RGB-D Image,” Journal of Korea Multimedia Society, December 2016.

[26] C. Dorin, and B.Hurwitz, “Automatic Body Part Measurement of Dressed Humans Using Single RGB-D Camera,” Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, May 2016.

[27] S. Gomathi, and T. Santhanam, “Application of Rectangular Feature for Detection of Parts of Human Body,” Advances in Computational Sciences and Technology, Volume 11, Number 1, pp. 43-55, 2018.

[28] M. S. Ryoo, and J. K. Aggarwal, “Spatio-temporal relationship match: video structure comparison for recognition of complex human activities,” Proc. of ICCV, pp. 1593-1600, May 2010.

[29] Y. Yang, and D. Ramanan, “Articulated Human Detection with Flexible Mixtures of Parts,” IEEE Transactions on Pattern Analysis and Machine Intelligence , December 2013.

[30] M. Mahmood, A. Jalal, and M. A. Sidduqi, “Robust Spatio-Temporal Features for Human Interaction Recognition Via Artificial Neural Network,” International Conference on Frontiers of Information Technology (FIT) , December 2018.

[31] A. Jalal and Y. Kim, “Dense Depth Maps-based Human Pose Tracking and Recognition in Dynamic Scenes Using Ridge Data,” in Proceedings of the IEEE International Conference on Advanced Video and Signal-based Surveillance, pp. 119-124, 2014.

[32] A. Jalal, S. Kamal and D. Kim, “Individual Detection-Tracking-Recognition using depth activity images,” in Proceedings 12th IEEE International Conference on Ubiquitous Robots and Ambient Intelligence, pp. 450-455, 2015.

[33] S. Kamal and A. Jalal, “A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors, Arabian Journal for Science and Engineering, vol. 41(3), pp. 1043-1051, 2016.

[34] A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal and D. Kim, “Robust human activity recognition from depth video using spatiotemporal multi-fused features, Pattern recognition, vol. 61, pp. 295-308, 2017.

[35] S. Kamal, A. Jalal and D. Kim, “Depth Images-based Human Detection, Tracking and Activity Recognition Using Spatiotemporal Features and Modified HMM, Journal of Electrical Engineering and Technology, pp. 1921-1926, 2016.

[36] A. Jalal, S. Kamal and D. Kim, “Human depth sensors-based activity recognition using spatiotemporal features and hidden markov model for smart environments, Journal of computer networks and communications, vol. 2016, pp. 1-11, 2016.

[37] A. Jalal, S. Kamal and D. Kim, “Facial Expression recognition using 1D transform features and Hidden Markov Model, Journal of Electrical Engineering & Technology, vol. 12(4), pp. 1657-1662, 2017.

[38] A. Jalal, S. Kamal, and D. Kim, “A depth video-based human detection and activity recognition using multi-features and embedded hidden Markov models for health care monitoring systems, International Journal of Interactive multimedia and Artificial Intelligence, vol. 4(4), pp. 54-62, 2017.

[39] M. Mahmood, A. Jalal, and H. A. Evans, “Facial Expression Recognition in Image Sequences Using 1D Transform and Gabor Wavelet Transform, IEEE conference on International Conference on Applied and Engineering Mathematics, 2018.

[40] A. Jalal, Majid A. K. Quaid, and A. S. Hasan, “Wearable Sensor-Based Human Behavior Understanding and Recognition in Daily Life for Smart Environments, IEEE conference on International Conference on Frontiers of information technology, 2018.

[41] A. Jalal, M. A. K. Quaid and M. A. Sidduqi "A Triaxial acceleration-based human motion detection for ambient smart home system," IEEE International Conference on Applied Sciences and Technology, 2019.

[42] A. Jalal, M. Mahmood and A. S. Hasan "Multi-features descriptors for human activity tracking and recognition in Indoor-outdoor environments," IEEE International Conference on Applied Sciences and Technology, 2019.

[43] A. Jalal and M. Mahmood, “Students’ Behavior Mining in E-learning Environment Using Cognitive Processes with Information Technologies,” Education and Information Technologies, Springer, 2019.

[44] A. Jalal, and M. A. K. Quaid, “IM-SportingBehaviors dataset.” [Online]. Available: http://portals.au.edu.pk/imc/Pages/Datasets.aspx. [Accessed: 11- Jun-2019].

[45] A. A. Rafique, A. Jalal, and A. Ahmed, “Scene Understanding and Recognition: Statistical Segmented Model using Geometrical Features and Gaussian Naïve Bayes”, IEEE conference on International Conference on Applied and Engineering Mathematics, 2019.

[46] M. Batool, A. Jalal, and Kibum Kim, “Sensors Technologies for Human Activity Analysis Based on SVM Optimized by PSO Algorithm”, IEEE ICAEM conference, 2019.

[47] A. Jalal, M. A. K. Quaid and K. Kim, “A Wrist Worn Acceleration Based Human Motion Analysis and Classification for Ambient Smart Home System,” Journal of Electrical Engineering & Technology, 2019.

[48] K. Kim, A. Jalal and M. Mahmood, “Vision-based Human Activity recognition system using depth silhouettes: A Smart home system for monitoring the residents, Journal of Electrical Engineering and Technology, 2019.

[49] A. Ahmed, A. Jalal, and K. Kim, “Region and decision tree-based segmentations for Multi- objects detection and classification in Outdoor Scenes”, IEEE Conference on Frontiers of Information Technology, 2019.

[50] A. A. Rafique, A. Jalal, and K. Kim, “Statistical multi-objects segmentation for indoor/outdoor scene detection and classification via depth images”, IEEE conference on applied sciences and technology, 2020.

[51] A. Ahmed, A. Jalal, and K. Kim, “RGB-D images for object segmentation, localization and recognition in indoor scenes using feature descriptor and Hough voting”, IEEE conference on applied sciences and technology, 2020.

[52] M. Mahmood, Ahmad Jalal, and K. Kim, “WHITE STAG Model: Wise Human Interaction Tracking and Estimation (WHITE) using Spatio-temporal and Angular-geometric (STAG) Descriptors, Multimedia Tools and Applications, 2020.