Statistical Multi-Objects Segmentation for Indoor/Outdoor Scene Detection and Classification via Depth Images

Adnan Ahmed Rafique
Dept. of Computer Science
Air University
Islamabad, Pakistan
adnanrafique@upr.edu.pk

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

Kibum Kim
Dept. of Human-Computer Interaction
Hanyang University
Seoul, South Korea
kibum@hanyang.ac.kr

***Abstract*—With the advancement of technology, intelligence capabilities of machines are growing day by day. Researchers are committed to equip the machines with the capability of thinking humanly. Currently, the machines can sense and process information gathered from sensors. However, still there is a huge gape to improve the capability of thinking and understanding real scenes. Scene understanding is fiery area of research now a day. Therefore, we have proposed a model to understand and recognize a scene using depth data to make machines capable of interpreting the real time scenes like humans. The proposed recognition technique is a novel segmentation framework that uses statistical multi object segmentation to learn robust scene model and segregate the objects in the scene. Then, the unique features are extracted from these segregated objects to further process for recognition using linear SVM. Finally, multilayer perceptron is provided with the features and weights for the recognition of the scene. Our system demonstrated significant improvement over state-of-the-art systems. The proposed system is effective in autonomous vision systems like robotic vision, GPS based location finder, sports and security.**

Keywords –Depth images, multi-object detection, statistical segmentation, scene detection.

I. Introduction

Scene understanding has been studied comprehensively. Though scene interpretation and understanding is not only a challenge in RGB, but also in depth images. RGB with depth images provide more information to localize, detect and label the foreground objects in a scene [1-4]. These scenes are labeled and analyzed on the basis of correct segmentation and recognition of multiple objects in the scene. Although scene understanding is an advanced problem than object detection but it purely dependent on the object detection technique which leads towards the correct scene understanding, recognition and labeling [5-9]. Therefore to achieve better results in indoor/outdoor robotics navigation, virtual reality, automatic guided vehicles and security there is a requirement to improve the quality of scene recognition techniques [10-15].

Some works dealt with the scene understanding problem by using the scene segmentation employing edge detection techniques [16-18]. In [19] S. Gupta *et al*. generalized gPb – ucm technique for hierarchical segmentation using boundary detection approach on depth images. After segmentation they applied histogram of oriented gradients (HOG) combined with deformable part model (DPM) for object detection. Finally by classifying superpixels into dominant categories of objects, they perform semantic segmentation, developing generic and class-specific features to obtain the objects label. X. Ren *et al.* extended kernel descriptors to convert local kernels to patch kernels [20]. They further combine two techniques: using MRG and using segmentation tree to show contextual relationships. An edge detection followed by connected components and filtering process is proposed by M. Dimitriou *et al*. [9] to detect multiple objects in a complex scene. Moreover they used Linear Spatial Pyramid Matching for classification of objects. In [21] D. Lin *et al.* proposed extended version of Constrained Parametric Min-Cuts (CPMC) framework. They achieved the goal of scene understanding by 2D segmentation and finding the contextual relations using conditional random fields between objects and scenes. K. Lai et al. in [22] presented a scene labeling system that combines learned features from RGB-D images and point clouds. On the basis of these features, assigns a label to the objects.

In our work, we have proposed a novel idea of hybrid HOG and local geometrical features for detection and recognition and multilayer perceptron for scene understanding. The final scene labels are acquired by processing data in four steps. Initially acquisition, smoothing and edge detection is performed. A combination of HOG and geometrical features are computed as second step. Then recognition of objects using linear SVM. At the final step, for scene recognition we applied multilayer perceptron. Furthermore, the proposed method is evaluated on publically available RGB-D scenes, RGB-D 7-Scenes [23] and NYUDv1 datasets [24]. It is observed that remarkable recognition results over state-of-the-art methods are considered.

The rest of the paper is organized as follows: Section II presents the flow architecture of our system methodology which comprises of edge detection, calculating gradients, feature computing, object detection and scene classification. Section III presents the analysis and comparison of the classification rate of our work with other state-of-the-art scene recognition systems using NYUDv1, RGB-D Scenes and RGB-D 7-Scenes datasets. Finally, section IV concludes the paper.

978-1-7281-4676-8/20/$31.00©2020 IEEE

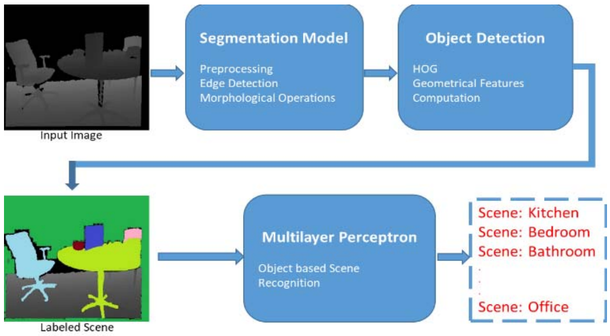

Fig. 1. Flow architecture of proposed system

II. System Methodology

The proposed system is comprised of four steps as illustrated in Fig. 1. Preprocessing to smooth the image, edge detection to segregate the objects and morphological operations to improve the segment boundaries are performed at first stage as segmentation process. Geometrical features are computed and extracted for object detection at second stage. Stage three encompassed of object recognition using linear SVM while the last and fourth stage is scene recognition of indoor scenes using multilayer perceptron.

A. Preprocessing & Enhanced Edge Detection

The depth image is acquired in the format of .png for preprocessing purpose [25-27]. Median filter [28-30] is applied as it keeps the edges preserved during noise removal as well as smoothed the image [31-34] as shown in the Fig. 2.

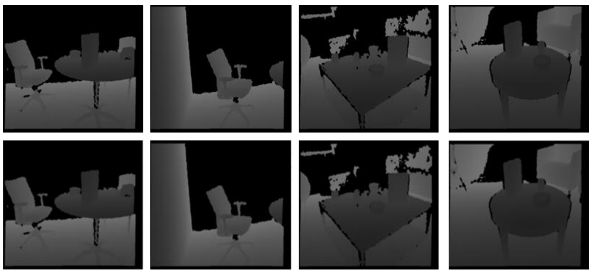

Fig. 2. Depth images before and after preprocessing, first row contains the raw depth images while second row illustrates the filtered images.

The edges are detected from the image acquired after preprocessing as median filter made it smooth and blurred. To detect the edges, blurred image is used to calculate the magnitude of the gradients [35-36]. After calculation of gradients, the edges are selected with large magnitudes. These selected edges are processed for sharp and thin edges using edge thinning process. Fig.3 demonstrates the detected edges of the depth images. Furthermore, to detect the strong edges, double thresholding is applied that uses a range of gray level values between $tv1$ *and* $tv2$. If the thresh value lies in the range, then the pixel is assigned 1(converted to white) otherwise 0 formulated as under:

$$|p| = \begin{cases} 1, & tv1 < p < tv2 \\ 0, & otherwise \end{cases} \quad (1)$$

where p is gray value of the respective pixel, $tv1$ is lower threshold value and $tv2$ is the upper threshold value. Then the connected edges are considered as the final edges which reflects the segmented objects.

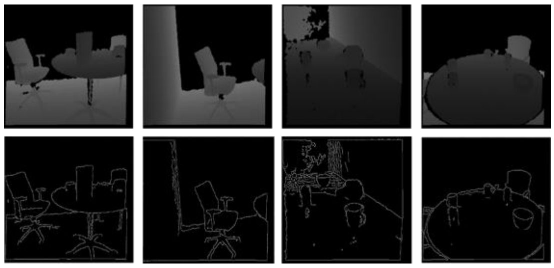

Fig. 3. Depth images after applying edge detection, upper row represents the original depth images and bottom row contains images with detected edges.

B. Features Computation

Two different type of features are computed and combined to achieve better results. The standard HOG [37-39] using depth information and geometrical features over segmented objects are used for feature computation.

HOG produces very good result against sharp edges. As we have segmented the images using edge detection and preserved the edges so HOG is used for improved accuracy purposes. Histogram of Gradients for depth images is used to obtain the cue [40-43] regarding surface orientation along with the depth discontinuities. The contrast normalization [44, 45] is then computed to clear the orientation of gradients which helps to recognize the surfaces at different depths for viewer.

Local geometrical features are calculated on the basis of four key points (extreme points) and centroid. These extreme points laid the foundation for the extraction of geometrical features including Euclidian distance, perimeter and area of triangles drawn using these extreme points and centroid.

To measure the Euclidian distance between any of two extreme points' a and b respectively, over the segmented object [46-48] in the depth image, mathematical formulation can be expressed as:

$$||dist|| = \sqrt{(a_2 - b_2)^2 + (a_1 - b_1)^2} \quad (2)$$

where $dist$ is the Euclidian distance between a, b two points having indices $_1$ *and* $_2$.

To calculate the area Å of triangle, three out of four extreme points and centroid are connected to complete a triangle for a segmented object. Area Å of the particular triangle for the segmented object may be formulated as:

$$Å = \sqrt{\mathrm{P}\,(\mathrm{P} - \overline{PQ})(\mathrm{P} - \overline{QR})(\mathrm{P} - \overline{RP})} \quad (3)$$

where Ƥ represents half of the perimeter, $\overline{PQ}$, $\overline{QR}$ and $\overline{RP}$ are the sides of triangle and Å symbolizes the area of triangle. Fig. 4 represents the average area of each class.

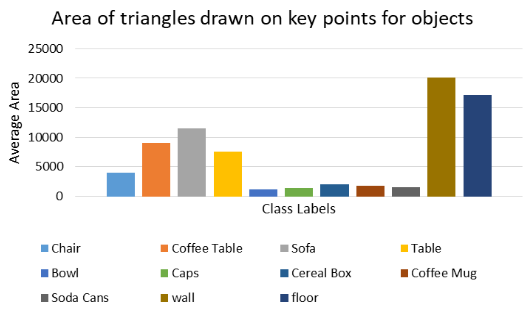

Fig. 4. Area of the triangles formed on the key points of different object categories on RGB-D Scenes dataset

C. *Object Detection & Recognition*

Computed features are provided to Linear SVM [49-52] for recognition of the detected objects. SVM can use various kernel functions for different scenarios. Here we implemented linear SVM that considers multiclass problem as multiple binary classes to classify according to the known class labels. The object is assigned with an appropriate class label on successful matching of input features whereas all the other classes are negated. Fig. 5 demonstrates the flow of linear SVM. Linear SVM can be demonstrated mathematically by a line or hyper-plane.

$$W.X + b = 0 \tag{4}$$

Where W represents a vector which is called a weight vector and $W = w_1, w_2, w_3, \dots, w_n$, b is a scalar called bias and X is feature input vector $X = x_1, x_2, x_3, \dots, x_n$

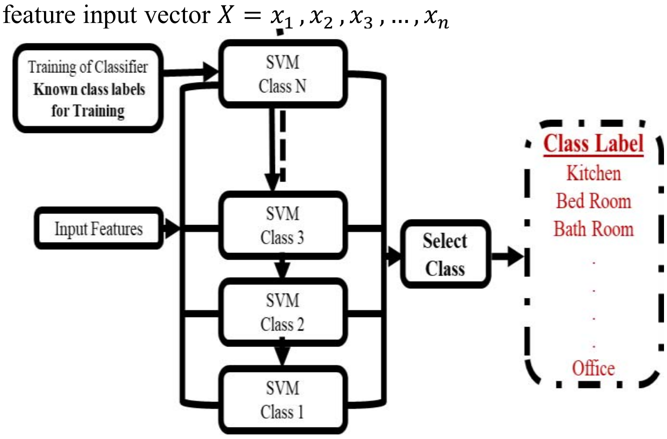

Fig. 5. Flow diagram of linear SVM

D. *Scene Recognition Using The Perceptron Classifier*

As for as supervised learning is concerned, the perceptron [53, 54] classifier is known to be one of the best binary classifiers. Fig. 6. Demonstrates the flow of multilayer perceptron. It takes features vector [55-59] as input combined with weights. A multilayer perceptron is class of feedforward artificial neural network. The nodes of the perceptron consists of three layers. Layer which takes feature vector and weights/bias as input is called input layer, output layer where the output is provided and the hidden layer between input and output layers. Hidden layers may vary from one to many according to the problem specifications. The process for any perceptron may be defined as under:

1. Generate an input output pair x *and* y_{pred} respectively.
2. Present the network with x and the y_{org} output to be generated.
3. To compute the error, calculate difference of y_{pred} and y_{org}.
4. Adjust weight w_{new} to reduce error
5. Repeat 2-4 multiple times

Now learning rule may be represented as under:

1. Randomly initialize the weights
2. Compute the following for each training pair x, y_{pred}

where x is a set of input features and y_{pred} is predicted class label.

a. Compute output y_{pred}

b. Compute error, $\delta = (y_{pred} - y_{org})$

c. Update weights by using δ as under:

$$\Delta w = w_{new} - w_{old} \text{ or } \Delta w = \gamma * \delta * x$$

where γ represents learning rate which defines the smoothness of learning process.

3. Repeat 2 until minimum value of δ (zero) /convergence

where
$$w_{new} = w_{old} + \gamma * \delta * x$$
$$\delta = (y_{pred} - y_{org})$$

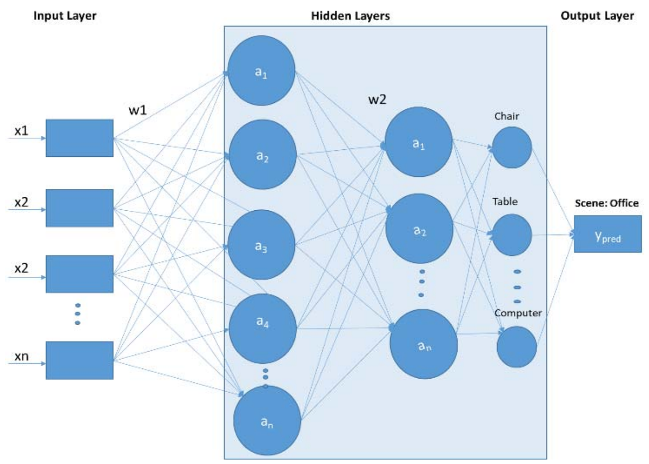

Fig. 6. Office scene recognition using multilayer perceptron

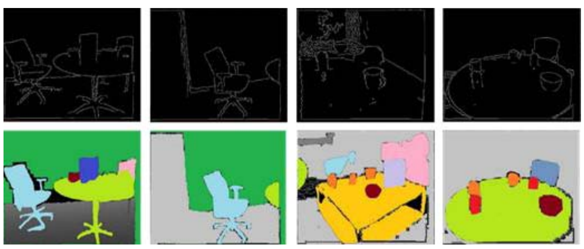

Fig. 7. Labeling results on some example images of RGB-D dataset

III. PERFORMANCE EVALUATION

The experimental studies are carried out at NYUDv1 and RGB-D Scenes datasets v2 to access the accuracy and efficiency of our proposed model. Datasets description and comparison with other state of the art methods is provided in the following subsections.

A. Dataset Descriptoin of NYUDv1 Dataset

NYUDv1 dataset [24] consists of 64 different indoor scenes of 7 types having 2347 unique & labeled frames while 1, 08,617 unlabeled frames. These unique frames and scenes belongs to the following classes*: bathroom, bedroom, bookstore, café, kitchen, living room* and *office*. Both the raw and labeled datasets are available. Labeled dataset is the subset of the raw dataset. Fig. 8 shows few examples of NYUDv1 datasets.

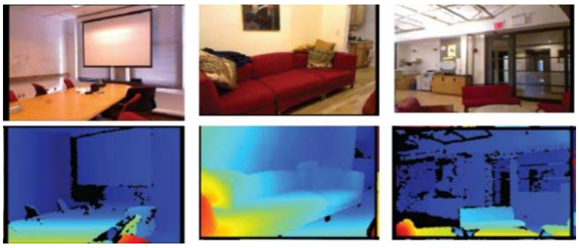

Fig. 8. Some examples of NYUDv1 dataset

B. Dataset Descriptoin of RGB-D 7-Scenes Dataset

Newly developed dataset on RGB-D [23] images having seven scenes. The scene labels are as under: *Chess, Fire, Heads, Office, Pumpkin, RedKitchen* and *Stairs*. The dataset consists of 26K training and 17K test frames. Fig. 9 shows few examples of RGB-D 7-scenes datasets.

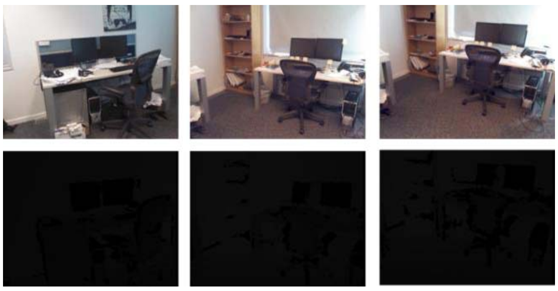

Fig. 9. Some examples of RGB-D 7-Scenes dataset (office scene)

C. Dataset Descriptoin of RGB-D Scenes v2 Dataset

The RGB-D Scenes Dataset v2 [60] consists of 14 scenes containing furniture (*chair, coffee table, sofa, table*) and a subset of the objects in the RGB-D Object Dataset (*bowls, caps, cereal boxes, coffee mugs,* and *soda cans*). Fig. 10. Illustrates examples of RGB-D Scenes dataset.

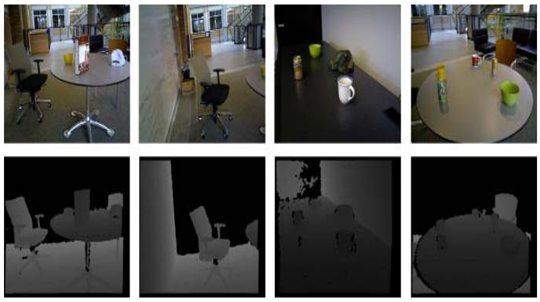

Fig. 10. Some examples of RGB-D Scenes dataset

D. Performance Evaluation of RGB-D Scenes V2 dataset and RGB-D 7-Scenes dataset

Table I presents the segmentation accuracy compared with the annotated benchmark datasets of RGB-D Scenes v2 and NYU-Dv1 dataset.

TABLE I. SEGMENTATION ACCURACY ON RGB-D SCENES V2 DATASET AND NYU-D V1 DATASET

Object Class	**Segmentation Mean Accuracy % (NYUD v1)**	**Object Class**	**Segmentation Mean Accuracy % (RGBD v2 Scenes)**
Chair	85.75	Chair	86.29
Table	80.27	Coffee Table	79.71
Monitor	85.88	Sofa	63.25
Cupboard	47.52	Table	75.88
		Bowl	57.71
		Caps	49.78
		Cereal Box	49.97
		Coffee Mug	73.76
		Soda Cans	49.87
		Wall	88.27
		Floor	87.51
Overall accuracy	**74.85**	**Overall accuracy**	**69.27**

Table II shows the evaluation results as a confusion matrix for the accuracy of RGB-D Scenes dataset.

TABLE II. CONFUSION MATRIX OF SEGMENTATION ACCURACY ON RGB-D SCENES DATASET

Object Classes	**CH**	**CT**	**SO**	**TB**	**BO**	**CP**	**CB**	**CM**	**SC**	**WL**	**FL**
CH	**86**	00	00	00	00	00	01	00	00	00	00
CT	00	**80**	00	00	03	00	00	00	06	00	00
SO	09	00	**63**	00	00	00	20	00	00	00	00
TB	00	30	00	**76**	00	00	00	30	04	00	00
BO	00	70	00	03	**58**	00	00	00	00	00	00
CP	03	00	05	07	00	**50**	03	00	00	00	00
CB	03	00	00	00	00	00	**50**	00	00	00	00
CM	04	00	02	00	00	00	02	**74**	00	00	00
SC	00	10	00	03	02	00	00	00	**50**	00	00
WL	00	00	00	00	00	00	00	00	00	**88**	09
FL	00	00	00	00	00	00	00	00	00	08	**87**

*CH = chair, CT = coffee table, SO = sofa, TB = table, BO = bowl, CP = cap, CB = cereal box, CM = coffee mug, SC = soda can, WL = wall, FL = floor.

Table III shows the comparison of the proposed method with the other state-of-the-art methods and depicts a significant improvement over other methods.

TABLE III. COMPARISON OF OUR METHOD WITH OTHER STATE OF THE ART METHODS

Method	NYUD v1	RGB-D Scenes
RANSAC [61]	62.7%	-----
RF+Scene [62]	64.97%	-----
SVM+Scene [62]	64.90%	-----
Proposed Method	**74.85%**	**69.27%**

IV. CONCLUSION

We have introduced an effective approach for scene understanding on depth images by applying statistical multi object segmentation. The segmentation process reveals the modified edge detection technique followed by morphological operations to get the clear segments. A hybrid of HOG and local geometrical features over segmented depth images are computed to detect the objects in the image. The computed feature vector is then passed to linear SVM for object recognition. Finally, the simple perceptron is applied to predict the scene label.

In future, we are devoted to apply Convolutional Neural Network for semantic segmentation on RGB, depth and RGB-D datasets.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2018R1D1A1A02085645).

REFERENCES

[1] M. Mahmood, A. Jalal, and H. Evans, "Facial Expression Recognition in Image Sequences Using 1D Transform and Gabor Wavelet Transform, *Proc. IEEE ICAEM*, 2018.

[2] A. Jalal, S. Kamal and D. 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] S. M. Abid Hasan and K. Kon, "Depth edge detection by image-based smoothing and morphological operations", Journal of Computational Design and Engineering, pp. 191-197, Feb. 2016.

[4] S. Kamal, A. Jalal and D. Kim, "Depth Images-based Human Detection, Tracking and Activity Recognition Using Spatiotemporal Features and Modified HMM, JEET, pp. 1921-1926, 2016.

[5] 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.

[6] A. Jalal and S. Kamal, "Improved behavior monitoring and classification using cues parameters extraction from camera array images," International Journal of Interactive multimedia and Artificial Intelligence, vol. 5(2), 2018.

[7] 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), 2017.

[8] 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.

[9] 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.

[10] A. Jalal, S. Kamal, and D. Kim, "Depth Map-based Human Activity Tracking and Recognition Using Body Joints Features and Self-Organized Map," in Proc. IEEE ICCCNT, pp. 1-6, 2014.

[11] A. Jalal, S. Kamal and D. Kim, "Depth Silhouettes Context: A new robust feature for human tracking and activity recognition based on embedded HMMs," in Proc. IEEE URAI, pp. 294-299, 2015.

[12] A. Jalal and Y. Kim, "Dense Depth Maps-based Human Pose Tracking and Recognition in Dynamic Scenes Using Ridge Data," in Proc. of the IEEE AVSS, pp. 119-124, 2014.

[13] A. Jalal, S. Kamal and D. Kim, "Individual Detection-Tracking-Recognition using depth activity images," in Proc. IEEE URAI, 2015.

[14] S. Kamal and A. Jalal, "A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors, AJSE, vol. 41(3), pp. 1043-1051, 2016.

[15] 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.

[16] R. Kirti and A Bhatnagar, "Image Segmentation Using Canny Edge Detection Technique", in porc. of international journal of techno-management research, pp. 8-14, Mar. 2017.

[17] A. Jalal, S. Kamal, A. Farooq and D. Kim, "A spatiotemporal motion variation features extraction approach for human tracking and pose based action recognition," in Proc. IEEE ICIEV, 2015.

[18] A. Farooq, A. Jalal and S. Kamal, "Dense RGB-D Map-Based Human Tracking and Activity Recognition using Skin Joints Features and Self-Organizing Map ," KSII, vol. 9(5), pp. 1856-1869, 2015.

[19] S. Gupta, P. Arbelaez, R.Girshick and J. Malik, "Indoor Scene Understanding with RGB-D Images Bottom-up Segmentation, Object Detection and Semantic Segmentation", International Journal Computer Vision, pp. 133-149, Nov. 2014.

[20] X. Ren, L. Bo and D. Fox, "RGB-(D) Scene Labeling: Features and Algorithms", IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, Jul. 2012.

[21] D. Lin, S. Fidler and R. Urtasun, "Holistic Scene Understanding for 3D Object Detection with RGBD cameras", IEEE International Conference on Computer Vision, pp. 1-8, Mar. 2014.

[22] K. Lai, L.Bo and D.Fox, "Unsupervised Feature Learning for 3D Scene Labeling", IEEE International Conference on Robotics & Automation, pp. 3050-3057, Jun. 2014.

[23] J. Shotton, B. Glocker, C. Zach, S. Izadi, A. Criminisi, A. Fitzgibbon, "Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images", IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, Oct. 2013.

[24] N. Silberman and R. Fergus, "Indoor Scene Segmentation using a Structured Light Sensor", IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp.1-8, Nov. 2011.

[25] A. Jalal, S. Kamal, and D. Kim, "Shape and motion features approach for activity tracking and recognition from Kinect video camera," in Proc. IEEE AINA, pp. 445-450, 2015.

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

[27] A. Jalal and S. Kim, "Algorithmic implementation and efficiency maintenance of real-time environment using low-bitrate wireless communication," in Proc. IEEE STFEUS, 2006.

[28] G. Gupta, "Algorithm for Image Processing Using Improved Median Filter and Comparison of Mean, Median and Improved Median Filter", International Journal of Soft Computing and Engineering (IJSCE), pp. 304-311, Nov. 2011.

[29] A. Jalal, Maria Mahmood, and M. A. Sidduqi, "Robust spatio-temporal features for human interaction recognition via artificial neural network,

IEEE conference on International Conference on Frontiers of information technology, 2018.

[30] A. Jalal and Y. Rasheed, “Collaboration achievement along with performance maintenance in video streaming,” in Proceedings of the IEEE conference on Interactive computer aided learning, pp. 1-8, 2007

[31] 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.

[32] A. Jalal, “IM-DailyDepthActivity dataset,” imlab.postech.ac.kr/databases.htm, 2015, [Online; accessed 14 April- 2015].

[33] A. Jalal and S. Kamal, “Real-Time Life Logging via a Depth Silhouette-based Human Activity Recognition System for Smart Home Services,” in Proc. AVSS, pp. 74-80, 2014.

[34] A. Jalal, MZ. Uddin, J. T. Kim, and T. Kim, “Daily Human Activity Recognition Using Depth Silhouettes and R Transformation for Smart Home,” in Proc. ICOST, pp. 25-32, 2011.

[35] Hongcheng Wang, Yunqiang Chen, Tong Fang, Jason Tyan, and Narendra Ahuja, "Gradient adaptive image restoration and enhancement", in Proc. of ICIP, pp. 2893-2896, 2006.

[36] A. Jalal, M. Maria and A. Hasan, “Multi-features descriptors for human activity tracking and recognition in Indoor-outdoor environments,” in proc. on IBCAST, 2019.

[37] 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.

[38] A. Jalal, M. Mahmood and K. Kim, “Vision-based Human Activity Recognition System Using Depth Silhouettes: A Smart Home System for Monitoring the Residents,” Journal of Electrical Engineering & Technology, 2019.

[39] A. Jalal, M. Quaid and M. Sidduqi, “A Triaxial acceleration-based human motion detection for ambient smart home system,” in Proceedings of IBCAST, 2019.

[40] A Jalal, Md. Zia Uddin and T.-S. Kim, “Depth Video-based Human Activity Recognition System Using Translation and Scaling Invariant Features for Life Logging at Smart Home”, IEEE Transaction on Consumer Electronics, vol. 58, pp. 863-871, 2012.

[41] 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 ICCCNT, pp. 1-6, 2014.

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

[43] 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.

[44] J. Konecny and M. Hagara, “One-Shot-Learning Gesture Recognition using HOG-HOF Features”, Journal of Machine Learning Research, pp. 1-20, Dec. 2013.

[45] 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.

[46] A. Jalal, and M. A. Zeb, “Security and QoS Optimization for distributed real time environment,” In: Proc. of the IEEE conference on Computer and Information Technology, pp. 369-374, 2007.

[47] A. Jalal, J. T. Kim, and T.-S. Kim, “Human activity recognition using the labeled depth body parts information of depth silhouettes,” in Proc. of ISHB, pp. 1-8, 2012.

[48] A. Jalal, S. Lee, J. Kim, and T. Kim, “Human activity recognition via the features of labeled depth body parts,” in *Proceedings Smart Homes Health Telematics*, pp. 246-249, 2012.

[49] J. Song, T. Wu and P. An, “Cascade Linear SVM for Object Detection”, The 9th International Conference for Young Computer Scientists, pp. 1755-1759, Dec. 2008.

[50] Y. Lee. Y. Lin and G. Wahba, “Multicategory Support Vector Machines”, in proc. of NSF, pp-1-15, 1999.

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

[52] A. Jalal and A. Nadeem and S. Bobasu, “Human body parts estimation and detection for physical sports movements,” IEEE International Conference on Communication, Computing and Digital Systems, 2019.

[53] A. Jalal, M. Mahmood and K. Kim, “Vision-based Human Activity Recognition System Using Depth Silhouettes: A Smart Home System for Monitoring the Residents,” Journal of Electrical Engineering & Technology, 2019.

[54] M. Liu *et al*., “Scene Recognition for Indoor Localization Using a Multi-Sensor Fusion Approach”, Sensors(Basel), pp. 1-20, Dec. 2017.

[55] A. Jalal, S. Kim and B. J. Yun, "Assembled algorithm in the real-time H.263 codec for advanced performance," ” in *Proceedings of the IEEE workshop on enterprise networking and computing in healthcare industry*, pp.295-298, 2005.

[56] 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.

[57] A. Jalal, J. T. Kim, and T.-S Kim, “Development of a life logging system via depth imaging-based human activity recognition for smart homes,” in *Proceedings of the International Symposium on Sustainable Healthy Buildings* , pp. 91-95, 2012.

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

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

[60] S. Song, S.P. Lichtenberg and J. Xiao, “SUN RGB-D: A RGB-D Scene Understanding Benchmark Suite”, Proceedings of 28th IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-10, 2015.

[61] N. Silberman, D. Hoiem, P. Kohli and R. Fergus, “Indoor Segmentation and Support Inference from RGBD Images”, 2012.

[62] S. Gupta, P. Arbel´aez, and J. Malik, “Perceptual Organization and Recognition of Indoor Scenes from RGB-D Images”, 2013.