Computers, Materials & Continua
DOI: 10.32604/cmc.2023.032245
Article

Tech Science Press

Real Objects Understanding Using 3D Haptic Virtual Reality for E-Learning Education

Samia Allaoua Chelloug[1,*], Hamid Ashfaq[2], Suliman A. Alsuhibany[3], Mohammad Shorfuzzaman[4], Abdulmajeed Alsufyani[4], Ahmad Jalal[2] and Jeongmin Park[5]

[1]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia
[2]Department of Computer Science, Air University, Islamabad, 44000, Pakistan
[3]Department of Computer Science, College of Computer, Qassim University, Buraydah, 51452, Saudi Arabia
[4]Department of Computer Science, College of Computers and Information Technology, Taif University, Taif, 21944, Saudi Arabia
[5]Department of Computer Engineering, Korea Polytechnic University, Siheung-si, Gyeonggi-do, 237, Korea
*Corresponding Author: Samia Allaoua Chelloug. Email: sachelloug@pnu.edu.sa
Received: 11 May 2022; Accepted: 24 June 2022

Abstract: In the past two decades, there has been a lot of work on computer vision technology that incorporates many tasks which implement basic filtering to image classification. The major research areas of this field include object detection and object recognition. Moreover, wireless communication technologies are presently adopted and they have impacted the way of education that has been changed. There are different phases of changes in the traditional system. Perception of three-dimensional (3D) from two-dimensional (2D) image is one of the demanding tasks. Because human can easily perceive but making 3D using software will take time manually. Firstly, the blackboard has been replaced by projectors and other digital screens so such that people can understand the concept better through visualization. Secondly, the computer labs in schools are now more common than ever. Thirdly, online classes have become a reality. However, transferring to online education or e-learning is not without challenges. Therefore, we propose a method for improving the efficiency of e-learning. Our proposed system consists of two-and-a-half dimensional (2.5D) features extraction using machine learning and image processing. Then, these features are utilized to generate 3D mesh using ellipsoidal deformation method. After that, 3D bounding box estimation is applied. Our results show that there is a need to move to 3D virtual reality (VR) with haptic sensors in the field of e-learning for a better understanding of real-world objects. Thus, people will have more information as compared to the traditional or simple online education tools. We compare our result with the ShapeNet dataset to check the accuracy of our proposed method. Our proposed system achieved an accuracy of 90.77% on plane class, 85.72% on chair class, and car class have 72.14%. Mean accuracy of our method is 70.89%.

Keywords: Artificial intelligence; e-learning; online education system; computer vision; virtual reality; 3D haptic

1 Introduction

In this digital age, the mode of information is changing to e-learning. In the past few years, many e-learning methods have been introduced for better understanding of concepts and also training purposes. When it comes to the online education system, there is a major problem regarding explaining 3D objects. 2D objects are easy to explain for the instructors and easy to understand by the students. But for the real view and 3D shapes of the real-world objects, it is very challenging to explain them. One solution was proposed in [1], which argued that immersive and haptic education system is the modern way of training people in virtual workspace. So, the students can better understand and be trained on the specific instrument and they can also recognize the system more precisely. Similarly, [2] argued that using VR in e-learning will enable a deep understanding of any possible concept. This is primarily because of the immersion, interaction and imagination goals of VR [3]. By using 2D to 3D reconstruction, the creation of the virtual world and 3D objects is easier and takes very less time and effort. According to [4], using a 3D virtual world will reduce the gap between learning management system (LMS) and learning theories. The 3D virtual world also improves the interactions of people with the instructors.

The main problem nowadays is the understanding of real 3D objects in e-learning. So, we use 3D haptic VR in e-learning for better understanding of 3D objects. The construction and generation of 3D meshes or objects from 2D images is also one of the challenging domains in this field. We rely on an effective 3D reconstruction technique from the 2D images to resolve this problem. We merge different solutions to resolve our major problem that is learning and understanding of 3D real-world objects using 3D haptic VR. This will reduce the cost of learning and it is easily available to everyone. First, we generate 3D model from its 2D image [5]. Then, this model can interact [6] in a virtual environment. Because of the availability of 2D dataset, we can reconstruct 3D virtual world very easily.

Our proposed system is based on simple image processing filters and uses machine learning techniques to extract features. It uses those for generating 3D output. In the first phase, 2.5D features are extracted from the input images. 2.5D is the combination of silhouette, depth, and surface normal. This is the required data that helps in the estimation of 3D shape of an object. For silhouette extraction, a simple edge detection filter followed by a hole filling filter has been used. Then, a neural network (NN) has been designed for depth estimation and depth results have been further investigated in the computation of surface normal. Then, these 3 features have been fed to a deep neural network for 3D mesh construction. The generated 3D meshes have been used in VR system for the purpose of understanding the real objects in 3D form in e-learning environments.

The article has been organized as follows. Section 2 gives a brief description of the related work. Then, Section 3 specifies the architecture of our proposed 2D to 3D mesh reconstruction system along with the details of each phase. Section 4 shows the performance of our proposed system. Section 5 contributes towards a brief discussion about the use of our system for e-learning and the limitations in our system. In the end, Section 6 presents the conclusion of this article our system and also discusses further research directions that can improve this system.

2 Literature Review

In this research, we are working on improving the e-learning methods using two proposed solutions that is, recovering of 3D from 2D datasets and also visualization of 3D information in virtual environment for e-learning. The traditional method is on board and also physical models of real-world objects are used in education system. Then, with the advancement in technology, we move towards e-learning systems where instructors interact with students using communication technology. It is difficult for students to understand real-world things. So, users will interact using virtual world [6]. Also, it fills the gap in learning.

2.1 Online Education via 3D Objects

According to Han et al. [7], the generation of 3D shapes from 2D images is one of the most challenging tasks for computers. When it comes to human perception, it is one of the easiest and natural tasks. Humans are trained naturally to perceive the objects form inmates in 3D objects. As analyzed by Szegedy et al. [8], it is very easy for humans to detect 3D shapes from images or 2.5D features. The conversion of 2.5D features is possible using simple and fast image processing filters as explained by Fan et al. [9] and convolutional neural network (CNN) methods proposed by Li et al. [10]. Computer use these 2.5D features to estimate and construct 3D models. Human have the capability to perceive 3D from 2D images because of the prior knowledge of different objects. Mathematically, it is impossible to recover the 3D depth according to Saxena et al. [11] from an image because it is flat in 2D form. Human vision can easily perceive depth from image. Häne et al. [12] defined famous hierarchical surface that is used to estimate the geometry of object so we can reconstruct 3D shape of that object. According to Han et al. [7], we can represent 3D objects in many forms in computer graphics that are scalable and are the best standards of 3D visuals. For raw data, we can use point clouds, voxels, polygons, and range images. For surface representation, there are mesh, subdivision, parametric, and implicit forms. For solid objects octree, Binary Space Partitioning (BSP) tree and Constructive Solid Geometry (CSG) are used in the High-end representation scene graphs. However, we use these visual representations in e-learning system where we can explain the concepts in a better way than the previous methods. We can map the behavior of students using behavior mining of students based on the idea presented in [13], and we can also track physics activity as explained in [14].

2.2 Online Education via Virtual Reality Systems

In the distance learning process, information and communication technologies (ICT) introduce new approach for improving learning of students. Learning management systems (LMS) are used nowadays where students communicate with instructors using video based online classes. According to Kotsilieris et al. [6], introducing virtual world in e-learning can change the way of interaction and also help users to interact with their avatars or 3D objects. Students learn better when they have virtual 3D view of objects as compared to the 2D images of different things. Fernandez [15] shows the challenges faced when implementing augmented VR in education system. Kokane et al. [16] implemented a system that is based on 3D virtual tutor using webRTC (Web Real-Time Communications) based application for e-learning.

2.3 E-learning via 3D Haptic Virtual Reality

There are many problems in e-learning when it comes the training of a specific equipment. Grajewski et al. [1], resolved this problem using haptic sensors and VR. We can simulate the equipment without building it. It saves time and cost of buying or building the equipment. Webb et al. [17]

simulated the nanoscale cell in 3D VR that helps biology students to understand the concepts in more depth and visualize it better as compared to 2D diagrams that are used traditionally. According to Edwards et al. [18] haptic VR is very useful in the field of organic chemistry, because of its immersive learning ability. It is used as gamification of chemistry experiments and also simulates the chemistry laboratory. The system is investigated to test the chemical reactions and also it is very safe. The conversion to learning environment as gamified environment will also increase interest in learning. Schönborn et al. [19] explored the tertiary students who used haptic virtual model to understand the structures in biomolecular binding. Students used haptic hardware with its virtual visualization to render the model. By using this system, it was easy to visualize the structure and also interact with them using the haptic sensor. Previously, this was a problem in this field where we could only visualize but not interact with a model. For haptics, we have to use wearable sensors. We can explore activity tracking methods in [20,21] and human interaction recognition in [22,23].

2.4 3D Object Construction for E-Learning Education

3D object reconstruction is one of the most challenging problem. We indicate that for estimating silhouette sequence, the method proposed by Jalal et al. [24–29] is based on depth sensor and it is mostly used for human activity recognition. Different types of human activities are defined in [30–33]. For transformation of feature, we can use hidden Markov model (HMM) and 1D transform features proposed in [34–37]. The latter is based on depth sensors for depth map extraction and use for human detection. Right now, we are working on real-life objects that help in e-learning education. We can employ these methods for different types of 3D models activity tracking. This, will improve the simulation in virtual world and as a result, it will improve the impact on e-learning education. We can traverse in the VR using wearable sensors [38,39] for accurate haptics, full controls and tracking [40].

3 Material and Methods

The proposed system has 4 main phases. Firstly, there is the object boundary estimation phase. We have investigated a simple contour detection algorithm to get the boundary. In the second phase, 2.5D features are extracted from the input images. The 2.5D features have 3 parts including silhouette, depth and surface normal. For silhouette extraction, we have applied a simple sliding window filter [41]. For depth extraction, we have used the indoor New York University (NYU v2) dataset [42] in training. Surface normal is easily extracted by adopting the idea presented in [9]. The simple 3 filters algorithm returns surface normal. Then, we generate 3D meshes utilizing a convolutional neural network (CNN). In the next step, we draw 3D cuboidal bounding boxes on 3D objects. The meshes are then easy to visualize in VR world. Using haptic sensors, we can efficiently feel and operate the model. This system is used for training purpose where students can better visualize the structure of an object and also assemble the object with its component virtually [1]. The complete architecture of our proposed method is shown in Fig. 1.

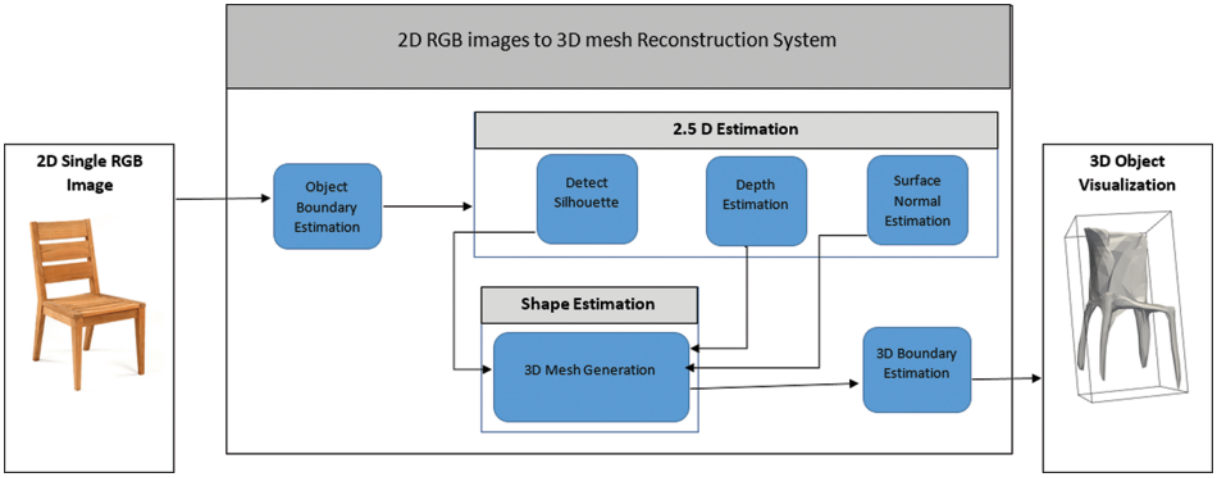


Figure 1: The architecture flow diagram of the proposed 3D reconstruction using a single Red-Green-Blue (RGB) image

3.1 Object Boundary Estimation

The first step is to filter out the objects from the image and then apply the remaining processes only on the objects extracted from the image. Nowadays, different object detection algorithms are commonly used that detect objects very easily based on deep learning methods [8,43,44]. The problem is that we need large dataset to train these models. The accuracy will be high if the volume of the dataset is large but when we have a limited number of images for training, deep learning methods will not provide the expected results. Also, we need graphics processing unit (GPU) or heavy computational resources to run deep learning algorithms. It is worth mentioning that an efficient image processing algorithm was proposed in [45] that uses cascade classifier to detect human faces but it is not very useful for our goal of object detection from images. We also use HoG (Histogram of Oriented Gradient) features [46] that is also used for human detection. Therefore, we resolved our problem by using simple image processing and machine learning algorithms. First, we detected the edges using an edge detection algorithm. Then, we filled the edges using opening and closing morphological operations. This technique has an issue that we need synthetic images and it also removes detail in images. Then, we used a simple machine learning algorithm based on 2D bounding box annotation in images to train the model. This is computationally very feasible to implement on low power devices. For human detection, we can use featured labelled parts of human body features [47] and real-world object detection using [48–50]. Different types of classification methods [51,52] are adopted for current scene classification [53] like semantics [54,55]. Also, Segmentation is used [56–59] to filter out the required portion of the object. Some methods are very useful in human detection and segmentation and also used for 3D real object segmentation.

3.2 2.5D Features Extraction

As suggested by Marr et al. [60], we use 2.5D features. 3D information is retrieved from an image using its 2.5D sketches estimation. By utilizing 2.5D, we can easily generate 3D. The 2.5D feature consists of silhouette, depth and surface normal.

3.2.1 Silhouette Extraction

We benefit from simple image processing filters for edge detection and then we perform simple opening and closing operations for filling the edges. Our method is simple but it requires synthetic images of objects. We use the following equations for silhouette extraction. First, we use RGB values to map the grayscale images and then we employ edge detection. Next, we apply horizontal and vertical gradient filters to the image. We further get the root mean square (RMS) value of both horizontal and vertical values and then mean filter will be adopted to smoothen the result.

$$GrayScale = \frac{R+G+B}{3} \tag{1}$$

$$G_x = \begin{bmatrix} +1 & 0 & -1 \\ 2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * Image\ (I) \tag{2}$$

$$G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * Image\ (I) \tag{3}$$

$$G_{x,y} = \sqrt{{G_x}^2 + {G_y}^2} \tag{4}$$

$$h[i,j] = \frac{1}{9}\sum_{x=i-1}^{i+1}\sum_{y=j-1}^{j+1} G_{x,y} \tag{5}$$

where, R represents the red channel, G represents green and B represents the blue channel in Eq. (1). The matrices indicated in Eqs. (2) and (3) are useful to get horizontal and vertical gradients respectively. After that, Eq. (4) allows combining the horizontal and vertical gradients using RMS. Next, the output is smoothen based on Eq. (5). The object images with simple background gives good results with this method as shown in Fig. 2.

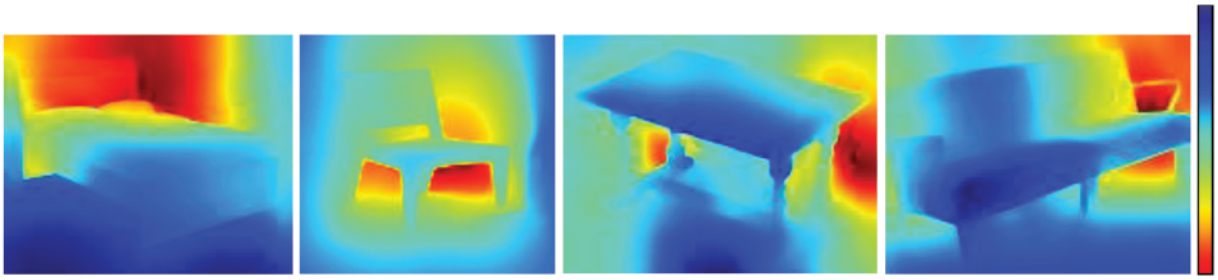
(a) (b) (c) (d)

Figure 2: Silhouette visualization use for object masking. a) Bed class, b) Chair lass, c) Table class and d) Sofa class

3.2.2 Depth Estimation and Surface Normal Extraction

Perceiving depth estimation from the single 2D image is very easy for human but when we approve mathematics to estimate 3D depth, it becomes impossible for computers because 2D images are flat when they are mapped in 2D array of pixels. Our proposed method considers CNN to estimate the depth. Hence, NYU v2 dataset that is a repository of images with its depth images is investigated in our paper. More specifically, the NN trained a model that supports the estimation of depth. We mention that Hu et al. [61] have utilized Squeeze-and-Excitation Networks (SENet-154) which are

an integration of Squeeze-and-Excitation (SE) block with Residual Transformations for Deep Neural Networks (ResNeXt). According to Dai et al. [62], the NYU v2 dataset is used to train this model. Algorithm 1 provides the specification of the proposed CNN model that enables to get the depth of the input image as shown in Fig. 3.

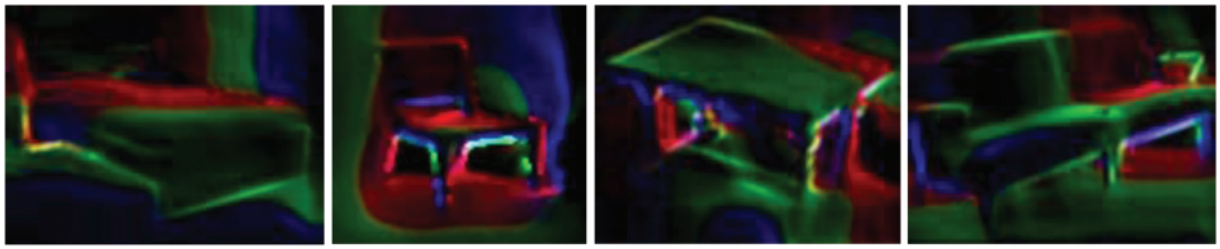


(a) (b) (c) (d)

Figure 3: Depth visualization blue represents the nearest point and red represents the maximum distance where a) Bed b) Chair c) Table and d) Sofa classes

3.2.3 Surface Normal Extraction

Surface normal is the feature that is used to determine the shape and structure of the object. Surface normal differentiates various orientations of the object in an image that further facilitates 3D estimation. According to Fan et al. [9], a simple filter method can be implemented to get the surface normal of object in image. This filter is based on edge detection filters and it has the ability to find angles according to the orientation of the object. More importantly, the role of surface normal is to find the orientation and size of different sides of the object. Wang et al. [63] designed a deep network to detect surface normal in learning stage to separately learn the global process and local process. In local process, red is used for occlusion, green for concave, and blue for convex. This method is also considered in our surface normal extraction filter. During visualization, it will give different colors to different surface orientations. So, it will make it easier for the computer to understand the shape. For computing surface normal, we are using depth image. The Algorithm 1 show the filters and kernel. The final result is shown in Fig. 4.



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

Figure 4: Surface normal image shows the orientation of object in different colors. a) Bed class, b) Chair class, c) Table class and d) Sofa class

Algorithm 1: Depth Estimation and Surface Normal Computation

Input: Img = 2D RGB image from testing dataset
Output: Depth _Img = .jpg image; SurfaceNormal_Img = using .jpg image;
Training Dataset: NYU v2 (Depth Dataset), DepthModel After training using NYU v2(Depth Dataset)
CNN = SENet-154 Architecture;
Testing Dataset: Furniture Detector
Kernalx = [[1, 1, 1],[0, 0, 0],[−1,−1,−1]];
Kernaly = [[−1, 0, 1],[−1, 0, 1],[−1, 0, 1]];
img, img Funtion (Img, Depth_Model, Kernalx,Kernaly)
  Depth_Img = DepthModel(Img)
  Filters: HorizontalFilter = filter2D(Depth_Img,-1, Kernalx);
    VerticalFilter = filter2D(Depth_Img,-1, Kernaly);
    Combine = HorizontalFilter + VerticalFilter;
  Mean = blur(Combine,(3, 3));
Return Depth_Img, SurfaceNormal_Img;

3.3 3D Mesh Generation

Mainly, we extract 2.5D feature and then, we generate mesh using ellipsoidal mesh deformation. We apply CNN and also max pooling for smoothing the edges. Mesh 3D file consists of vertices and edges, so it is scalable and easy for computer to map in 3D environment. According to Pan et al. [64], we can explore CNN to extract features from multiple images and then we can use those features in Graph Convolutional Networks (GCN) for generating 3D shapes in the form of mesh. According to He et al. [65], multiple images of a single object from different angles are needed for generating mesh model of an RGB image. We use adopt this approach to generate 3D meshes using neural network as shown in Fig. 5.

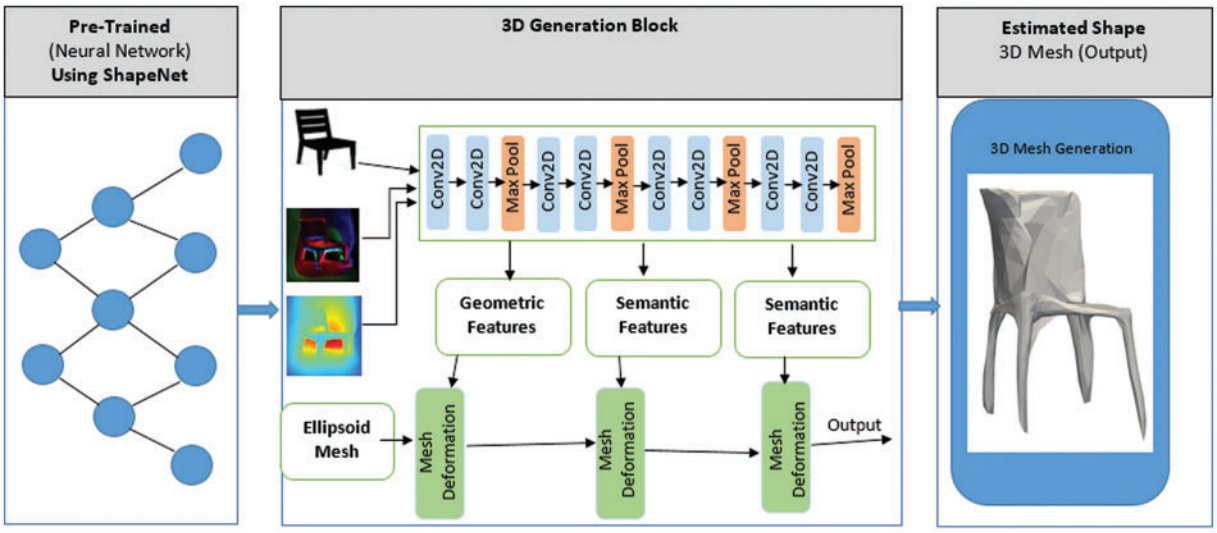


Figure 5: 3D Mesh generation method using neural network

By using different types of filters and methods [66–69] to extract the features we get much information that is useful to reconstruct 3D [70–73]. We investigate that features in the deep neural network to get the 3D mesh model of the object [74–77]. Fig. 6 shows the 3D mesh results of the above examples.

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

Figure 6: 3D Mesh representation of the following classes a) Bed, b) Chair, c) Table and d) Sofa

3.4 3D Bounding Box Estimation

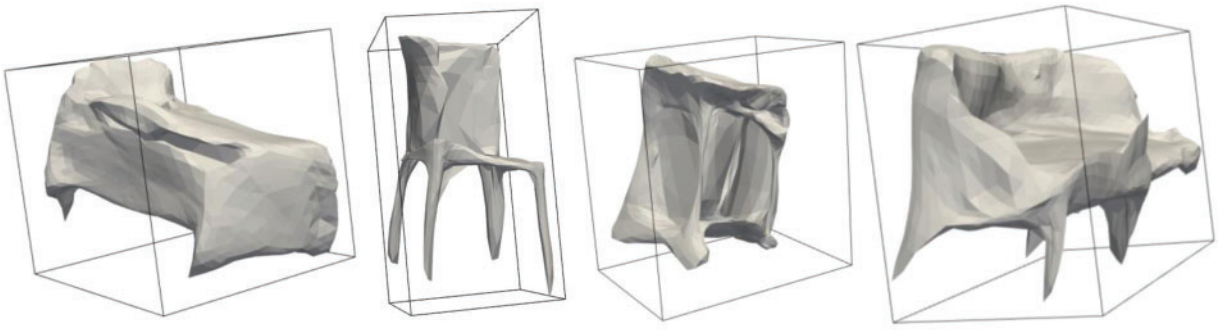
Size estimation is also one of the major challenges in the field of artificial intelligence. We estimated the 3D boundary around the mesh reconstructed based on RGB image and its 2.5D features. The generated 3D mesh was placed on the origin and the bounding box around it using its length, width, and height. The orientation of the 3D object also matters because without orientation, we would need human guidance to place that object in the VR environment. According to Mousavian et al. [78] we can estimate 3D bounding box using 2D bounding box and geometric orientation of object in an image using deep learning methods. We generated 3D bounding box after creating mesh. The 3D bounding box is calculated by these equations.

$$Width = (x_1, y_1 - Mesh_{Center}) + (x_3, y_3 - Mesh_{Center}) \tag{6}$$

$$Length = (x_2, y_2 - Mesh_{Center}) + (x_4, y_4 - Mesh_{Center}) \tag{7}$$

$$Height = (x_5, y_5 - Mesh_{Center}) + (x_6, y_6 - Mesh_{Center}) \tag{8}$$

where, $Mesh_{Center}$ is at the origin, x_1, y_1 are the coordinates of the right most point, and x_3, y_3 are the left most point. So, by adding distances, we can get the width of the mesh. Similarly, if we rotate the origin on x-axis at 90 degree, we get the right most point x_2, y_2 and the left most point x_4, y_4 and we combine these distances to get the length. Then, we rotate the origin at y-axis at 90 degrees to get the right most point x_5, y_5 and the left most point x_6, y_6 and combine the distances to get the height. We get the length, width and height of the bounding box with center its point at origin. As represented in Fig. 7 the bounding box aligns according to the orientation of 3D object.



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

Figure 7: 3D Cuboid boundary of the above classes a) Bed, b) Chair, c) Table and d) Sofa

4 Experiments and Results

In this section, we analyze the results of our proposed system. We divide the main architecture in four modules and based on the results we compare the proposed system with the previously available methods. We use furniture detector dataset for experimentation and testing purpose. Then we use ShapeNet dataset [79] for verification of our system using ground truth values.

4.1 Experimental Setup

For testing our proposed system, we needed a dataset for experimentation and ground truth values to check its accuracy and performance. This section is further divided in to 3 sub-sections. The first sub-section describes the details about the dataset that has been used in experimentation. The second sub-section visualizes the results obtained using the benchmark dataset. Third sub-section compares the proposed method with other state-of-the-art methods for performance evaluation.

4.1.1 Furniture Detector Dataset Description

For experimentation, we have used simple image processing filters to detect boundaries and then we applied morphological operations to refine the shape. So, we needed simple synthetic images to get best results. The above-mentioned techniques work best on synthetic images. In education, we mostly consider synthetic image of the object for description. So, it's very useful in this case. We utilized publicly available furniture detector dataset that we got from Kaggle. This dataset consists of 4 classes (sofa, chair, table and bed). For training, each class has 900 images and table class has 425 images. For validation, there are 100 images from each class and the table class has 23 images. Some sample images from the datasets are shown in Fig. 8.

Figure 8: Furniture detector dataset. a) Bed class, b) Chair class, c) Table class and d) Sofa class

4.1.2 ShapeNet Dataset

For experimentation, we also used ShapeNet dataset [79] for testing the proposed system and performance of our system is also good on this dataset to. We also get the ground truth 3D shape that is very helpful at the end to check the accuracy.

4.2 Visualization and Metrics

Our framework reconstructs a 3D mesh form a single 2D RGB image. There are different types of visualizations for 3D data: mesh, voxels and point cloud. Voxel reconstruction is one of the finest forms of 3D reconstruction that is mostly used in games and 3D environments. In this research, we adopt meshes that consist of edges and vertices and based on the edges in mesh, we can compute the accuracy of our model. Mesh visualization with its ground truth shown in Fig. 9.

We have successfully computed the 3D mesh model from 2D image and we have compared our result with the ShapeNet dataset [79] to check the accuracy of our proposed method. Our proposed system achieved the accuracy of 70.30 on sofa class, 85.72 on chair class, table and bed class have 72.05% and 55.50% respectively. Mean accuracy of our method is 70.89%. The details of accuracy and test error Percentage are shown in Tab. 1.

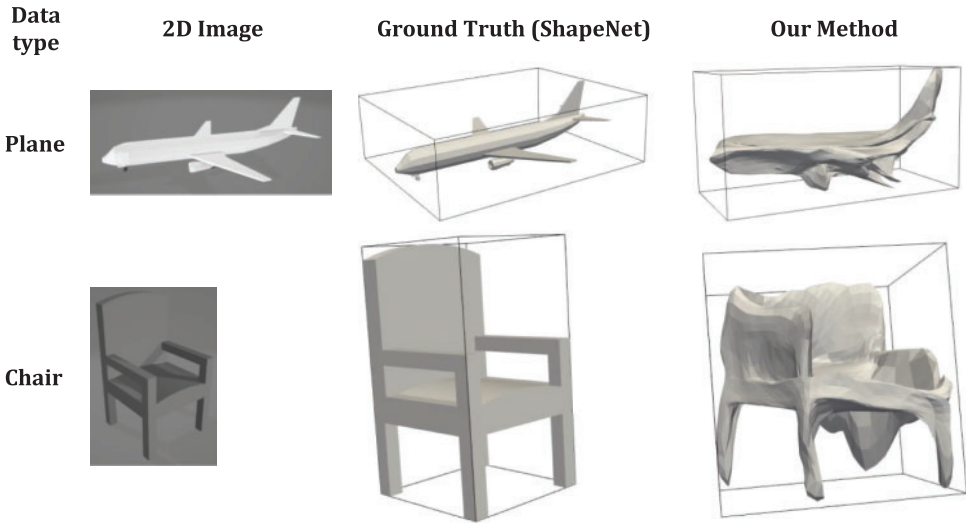


Figure 9: Furniture detector dataset. a) Bed class, b) Chair class, c) Table class and d) Sofa class

Table 1: Accuracy and test-error percentage and other method accuracy

Class	Our method accuracy (%)	Test-Error (%)	Method [12] accuracy (%)	Method [80] accuracy (%)
Sofa	70.30	29.70	45.04	67.37
Chair	85.72	14.28	37.80	49.09
Table	72.05	27.95	60.05	48.29
Bed	55.50	45.50	32.00	-
Mean	70.8925	29.3575	43.7125	60.16

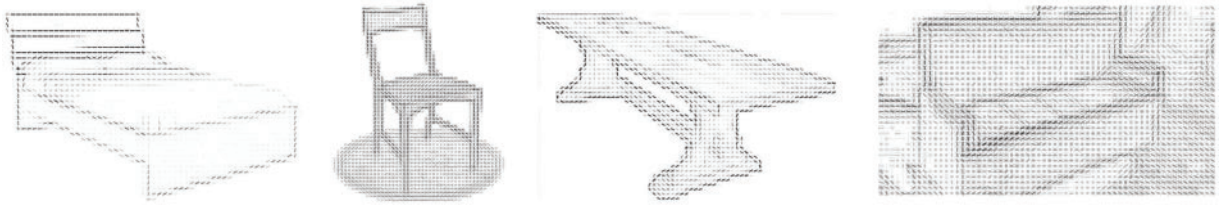
Considering computation time our method automates the process of generation of 3D from 2D images. By using computer-aided design (CAD) integrated software each model of object takes time in designing, development and rendering. The CAD integrated software need powerful system resources to work more efficiently. Tab. 2 shows the required time to construct 3D using integrated software compared with our method.

Table 2: Reconstruction time on integrated cad software and our proposed model

Objects	Integrated software	Our method
Sofa	180 min	5 min
Chair	120 min	5 min
Table	60 min	5 min
Bed	150 min	5 min

5 Discussion

At this stage, we have developed a system that can generate a 3D mesh from a single image so that the object is useable in VR system for e-learning. We can further improve the system by generating components of an object using their images and then these components can be assembled in virtual world using haptic sensor. This type of training is used for engineers and also in the medical field where artificial virtual surgery simulation can be performed for training purpose. We can use demographic factors acceptance VR and relation extraction [81,82] in hybrid evaluation of users in educational games [83] also used to perceive security risk based on modern factors for blockchain technology [84]. We compared our model performance with the state-of-the-art methods that also have very good results. But those methods were based on deep learning. We get better results as compared to the state-of-the-art methods. However, our method has some limitations. We need to use synthetic images as input because our model is based on simple image processing filters that are not good at object segmentation. We also tried to use HoG features shown in Fig. 10. But HoG features are only efficient with human detection. So, HoG feature descriptor didn't work with or system. In future, if we need to reconstruct human model, then we need HoG features that are best for human detection. In future, we are working on human datasets, our goal is to estimate and map the facial and body deformation in 3D for using 2D image dataset.



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

Figure 10: Examples of some failure example cases in a) Bed, b) Chair, c) Table and d) Sofa classes

The compared method [12] use deep neural network for training. Those methods are more complex as compared to our method because need powerful system for training the model.

6 Conclusion

This proposed system is used for understanding and analyzing the 3D real-world object using VR haptic sensors that will improve the overall experience of e-learning. By using this method, there is very less cost on development of 3D virtual world for e-learning system. Because the cost of modeling manually is reduced, we didn't need a heavy system to render the 3D object. So, our system, contributes in saving time and computational resources when building virtual world for e-learning platforms. In physical classes, the educator mostly uses models of different things which are difficult to describe using images and diagrams. The models are mostly deformable and it have more information than a simple 2D image. Hence, this research fills a major gap in the current e-learning education system. In future, we will work on human face and body deformation. Also improve proposed model using deep learning and multiple view dataset.

Funding Statement: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-2018-0-01426) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

In addition; the authors would like to thank the support of the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University, This work has also been supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2022R239), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Also; this work was partially supported by the Taif University Researchers Supporting Project Number (TURSP-2020/115), Taif University, Taif, Saudi Arabia.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] D. Grajewski, F. Górski, A. Hamrol and P. Zawadzki, "Immersive and haptic educational simulations of assembly workplace conditions," *Procedia Computer Science*, vol. 75, pp. 359–368, 2015.

[2] T. Monahan, G. McArdle and M. Bertolotto, "Virtual reality for collaborative e-learning," *Computers & Education*, vol. 50, no. 4, pp. 1339–1353, 2008.

[3] Z. Li, J. Yue and D. A. G. Jáuregui, "A new virtual reality environment used for e-learning," in *IEEE Int. Symp. on IT in Medicine & Education*, Jinan, Shandong, China, pp. 445–449, 2009.

[4] A. Jalal, M. Z. Sarwar and K. Kim, "RGB-D images for objects recognition using 3D point clouds and RANSAC plane fitting," in *Proc. Int. Conf. on Applied Science and Technology*, Islamabad, Pakistan, pp. 518–523, 2021.

[5] A. Ahmed, A. Jalal and K. Kim. "Region and decision tree-based segmentations for multi-objects detection and classification in outdoor scenes," in *Proc. Int. Conf. on Frontiers of Information Technology*, Islambad, Pakistan, pp. 209–2095, 2019.

[6] T. Kotsilieris and N. Dimopoulou, "The evolution of e-learning in the context of 3D virtual worlds," *Electronic Journal of e-Learning*, vol. 11, no. 2, pp. 147–167, 2013.

[7] X. -F. Han, H. Laga and M. Bennamoun, "Image-based 3d object reconstruction: State-of-the-art and trends in the deep learning era," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1578–1604, 2021.

[8] C. Szegedy, A. Toshev and D. Erham, "Deep neural networks for object detection," *Advances in Neural Information Processing Systems*, vol. 26, pp. 2553–2561, 2013.

[9] R. Fan, H. Wang, B. Xue, H. Huang, Y. Wang *et al.,* "Three-filters-to-normal: An accurate and ultrafast surface normal estimator," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5405–5412, 2021.

[10] B. Li, C. Shen, Y. Dai, A. Van Den Hengel and M. He, "Depth and surface normal estimation from monocular images using regression on deep features and hierarchical CRFs," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, San Juan, PR, USA, pp. 1119–1127, 2015.

[11] A. Saxena, M. Sun and A. Y. Ng, "Make3D: Depth perception from a single still image," *Association for the Advancement of Artificial*, vol. 3, pp. 1571–1576, 2008.

[12] C. Häne, S. Tulsiani and J. Malik, "Hierarchical surface prediction for 3D object reconstruction," in *Int. Conf. on 3D Vision (3DV)*, Qingdao, China, pp. 412–420, 2017.

[13] A. Jalal, A. Nadeem and S. Bobasu, "Human body parts estimation and detection for physical sports movements," in *IEEE Int. Conf. on Communication, Computing and Digital Systems*, Jinan, Shandong, China, pp. 104–109, 2019.

[14] A. Ahmed, A. Jalal and A. A. Rafique, "Salient segmentation based object detection and recognition using hybrid genetic transform," in *IEEE ICAEM Conf.*, Taxila, Pakistan, pp. 203–208, 2019.

[15] M. Fernandez, "Augmented virtual reality: How to improve education systems," *Higher Learning Research Communications*, vol. 7, no. 1, pp. 1–15, 2017.
[16] A. Kokane, H. Singhal, S. Mukherjee and G. R. M. Reddy, "Effective e-learning using 3D virtual tutors and webrtc based multimedia chat," in *2014 Int. Conf. on Recent Trends in Information Technology*, Chennai, India, pp. 1–6, 2014.
[17] M. Webb, M. Tracey, W. Harwin, O. Tokatli, F. Hwang *et al.,* "Haptic enabled collaborative learning in virtual reality for schools," *Education and Information Technologies*, vol. 27, no. 1, pp. 937–960, 2022.
[18] B. I. Edwards, K. S. Bielawski, R. Prada and A. D. Cheok, "Haptic virtual reality and immersive learning for enhanced organic chemistry instruction," *Virtual Reality*, vol. 23, no. 4, pp. 363–373, 2019.
[19] K. J. Schönborn, P. Bivall and L. A. Tibell, "Exploring relationships between students' interaction and learning with a haptic virtual biomolecular model," *Computers & Education*, vol. 57, no. 3, pp. 2095–2105, 2011.
[20] A. Jalal, M. Mahmood and A. S. Hasan "Multi-features descriptors for human activity tracking and recognition in indoor-outdoor environments," in *IEEE Int. Conf. on Applied Sciences and Technology*, Islamabad, Pakistan, pp. 371–376, 2019.
[21] A. Jalal, M. A. K. Quaid and A. S. Hasan, "Wearable sensor-based human behavior understanding and recognition in daily life for smart environments," in *IEEE Conf. on Int. Conf. on Frontiers of Information Technology*, Islamabad, Pakistan, pp. 105–110, 2018.
[22] M. Mahmood, A. Jalal and M. A. Sidduqi, "Robust spatio-temporal features for human interaction recognition via artificial neural network," in *Proc. of Int. Conf. on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, pp. 218–223, 2018.
[23] A. Jalal, M. A. K. Quaid and M. A. Sidduqi "A triaxial acceleration-based human motion detection for ambient smart home system," in *IEEE Int. Conf. on Applied Sciences and Technology*, Islamabad, Pakistan, pp. 353–358, 2019.
[24] 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 *Proc. of the Int. Symp. on Sustainable Healthy Buildings*, Seoul, Korea, pp. 91–95, 2012.
[25] 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 the 6th International Symp. on Sustainable Healthy Buildings*, Seoul, Korea, pp. 1–8, 2012.
[26] 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, no. 3, pp. 863–871, 2012.
[27] A. Jalal, Y. Kim and D. Kim, "Ridge body parts features for human pose estimation and recognition from RGB-D video data," in *Proc. of the IEEE Int. Conf. on Computing, Communication and Networking Technologies*, Hefei, China, pp. 1–6, 2014.
[28] A. Jalal, N. Khalid and K. Kim, "Automatic recognition of human interaction via hybrid descriptors and maximum entropy markov model using depth sensors," *Entropy*, vol. 22, no. 8, pp. 1–33, 2020.
[29] 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 Int. Conf. on Advanced Video and Signal-Based Surveillance*, Seoul, South Korea, pp. 119–124, 2014.
[30] A. Jalal and S. Kamal, "Real-time life logging via a depth silhouette-based human activity recognition system for smart home services," in *Proc. of the IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance*, Seoul, South Korea, pp. 74–80, 2014.
[31] A. Jalal, Y. Kim, S. Kamal, A. Farooq and D. Kim, "Human daily activity recognition with joints plus body features representation using kinect sensor," in *Proc. IEEE Int. Conf. on Informatics, Electronics and Vision*, Fukuoka, Japan, pp. 1–6, 2015.
[32] 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 Int. Conf. on Informatics, Electronics and Vision*, Fukuoka, Japan, pp. 1–6, 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, no. 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*, vol. 11, no. 6, pp. 1857–1862, 2016.
[36] 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, no. 4, pp. 1657–1662, 2017.
[37] 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, no. 4, pp. 54–62, 2017.
[38] M. A. K. Quaid and A. Jalal, "Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm," *Multimedia Tools and Applications*, vol. 79, no. 9, pp. 6061–6083, 2019.
[39] A. Nadeem, A. Jalal and K. Kim, "Human actions tracking and recognition based on body parts detection via artificial neural network," in *IEEE Int. Conf. on Advancements in Computational Sciences*, Lahore, Pakistan, pp. 1–6, 2020.
[40] S. Badar, A. Jalal and M. Batool, "Wearable sensors for activity analysis using smo-based random forest over smart home and sports datasets," in *IEEE ICACS Conf.*, Lahore, Pakistan, pp. 1–6, 2020.
[41] A. Jalal, N. Sarif, J. T. Kim and T. -S. Kim, "Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home," *Indoor and Built Environment*, vol. 22, no. 1, pp. 271–279, 2013.
[42] L. He, J. Lu, G. Wang, S. Song and J. Zhou, "SOSD-Net: Joint semantic object segmentation and depth estimation from monocular images," *Neurocomputing*, vol. 440, pp. 251–263, 2021.
[43] R. Huang, J. Pedoeem and C. Chen, "YOLO-LITE: A real-time object detection algorithm optimized for non-gpu computers," in *IEEE Int. Conf. on Big Data (Big Data)*, Seattle, WA, USA, pp. 2503–2510, 2018.
[44] X. Wu, D. Sahoo and S. C. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, 2020.
[45] M. J. Jones and P. Viola, "Robust real-time object detection," in *Workshop on Statistical and Computational Theories of Vision*, Vancouver, Canada, vol. 266, pp. 1–56, 2001.
[46] X. Wang, T. X. Han and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *IEEE 12th Int. Conf. on Computer Vision*, Kyoto, Japan, pp. 32–39, 2009.
[47] A. Jalal, S. Lee, J. Kim and T. Kim, "Human activity recognition via the features of labeled depth body parts," in *Proc. Smart Homes Health Telematics*, Berlin, Heidelberg, pp. 246–249, 2012.
[48] 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*, vol. 14, no. 4, pp. 1733–1739, 2019.
[49] A. Ahmed, A. Jalal and K. Kim, "Region and decision tree-based segmentations for multi-objects detection and classification in outdoor scenes," in *IEEE Conf. on Frontiers of Information Technology*, Islamabad, Pakistan, pp. 209–214, 2019.
[50] M. Mahmood, A. 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*, vol. 79, no. 11, pp. 6919–6950, 2020.
[51] A. Ahmed, A. Jalal and K. Kim, "A novel statistical method for scene classification based on multi-object categorization and logistic regression," *Sensors*, vol. 20, no. 14, pp. 1–20, 2020.
[52] A. Jalal, I. Akhtar and K. Kim, "Humanposture estimation and sustainable events classification via pseudo-2d stick model and k-ary tree hashing," *Sustainability*, vol. 12, no. 23, pp. 1–24, 2020.

[53] Y. Ghadi, I. Akhter, M. Alarfaj, A. Jalal and K. Kim, "Syntactic model-based human body 3d reconstruction and event classification via association based features mining and deep learning," *PeerJ Computer Science*, vol. 7, pp. 1–36, 2021.
[54] A. Jalal, A. Ahmed, A. Rafique and K. "Kim "Scene semantic recognition based on modified fuzzy c-mean and maximum entropy using object-to-object relations"," *IEEE Access*, vol. 9, pp. 27758–27772, 2021.
[55] K. Nida, G. Y. Yazeed, M. Gochoo, A. Jalal and K. Kim, "Semantic recognition of human-object interactions via Gaussian-based elliptical modelling and pixel-level labeling," *IEEE Access*, vol. 9, pp. 111249–111266, 2021.
[56] A. Rafique, A. Jalal and K. Kim, "Statistical multi-objects segmentation for indoor/outdoor scene detection and classification via depth images," in *Proc. Int. Conf. on Int. Bhurban Conf. on Applied Sciences and Technology*, Bhurban, Pakistan, pp. 271–276, 2020.
[57] A. Rafique, A. Jalal and K. Kim, "Automated sustainable multi-object segmentation and recognition via modified sampling consensus and kernel sliding perceptron," *Symmetry*, vol. 12, no. 11, pp. 1–25, 2020.
[58] K. Nida, M. Gochoo, A. Jalal and K. Kim, "Modeling two-person segmentation and locomotion for stereoscopic action identification: A sustainable video surveillance system," *Sustainability*, vol. 13, no. 2, pp. 1–30, 2021.
[59] A. Ahmed, A. Jalal and K. Kim , "Multi-objectsdetection and segmentation for scene understanding based on texton forest and kernel sliding perceptron," *Journal of Electrical Engineering and Technology*, vol. 16, no. 2, pp. 1143–1150, 2020.
[60] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 204, no. 1156, pp. 301–328, 1979.
[61] J. Hu, L. Shen and G. Sun, "Squeeze-and-excitation networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, San Juan, PR, USA, pp. 7132–7141, 2018.
[62] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser *et al.,* "Scannet: Richly-annotated 3d reconstructions of indoor scenes," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 5828–5839, 2017.
[63] X. Wang, D. Fouhey and A. Gupta, "Designing deep networks for surface normal estimation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 539–547, 2015.
[64] J. Pan, X. Han, W. Chen, J. Tang and K. Jia, "Deep mesh reconstruction from single RGB images via topology modification networks," in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, Seoul, Korea, pp. 9964–9973, 2019.
[65] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 770–778, 2016.
[66] A. Jalal and M. Mahmood, "Students' behavior mining in E-learning environment using cognitive processes with information technologies," *Education and Information Technologies*, Springer, vol. 24. no. 5, pp. 2797–2821, 2019.
[67] M. Gochoo, I. Akhter, A. Jalal and K. Kim, "Stochastic remote sensing event classification over adaptive posture estimation via multifused data and deep belief network," *Remote Sensing*, vol. 13, no. 5, pp. 1–29, 2021.
[68] N. Amir, A. Jalal and K. Kim, "Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy markov model," *Multimedia Tools and Applications*, vol. 80, no. 14, pp. 21465–29498, 2021.
[69] I. Akhter, A. Jalal and K. Kim, "Adaptive pose estimation for gait event detection using context-aware model and hierarchical optimization," *Journal of Electrical Engineering and Technology*, vol. 16, no. 5, pp. 2721–2729, 2021.
[70] M. Gochoo, S. Badar, A. Jalal and K. Kim, "Monitoring real-time personal locomotion behaviors over smart indoor-outdoor environments via body-worn sensors," *IEEE Access*, vol. 9, pp. 70556–70570, 2021.
[71] P. Mahwish, G. Yazeed, M. Gochoo, A. Jalal, S. Kamal *et al.,* "A smart surveillance system for people counting and tracking using particle flow and modified SOM," *Sustainability*, vol. 13, no. 10, pp. 1–21, 2021.

[72] M. Gochoo, S. R. Amna, G. Yazeed, A. Jalal, S. Kamal *et al.,* "A systematic deep learning based overhead tracking and counting system using RGB-d remote cameras," *Applied Sciences*, vol. 11, no. 12, pp. 1–21, 2021.
[73] U. Azmat and A. Jalal, "Smartphone inertial sensors for human locomotion activity recognition based on template matching and codebook generation," in *IEEE Int. Conf. on Communication Technologies*, Islamabad, Pakistan, pp. 109–114, 2021.
[74] S. Amna, A. Jalal and K. Kim, "An accurate facial expression detector using multi-landmarks selection and local transform features," in *IEEE ICACS Conf.*, Lahore, Pakistan, pp. 1–6, 2020.
[75] P. Mahwish, A. Jalal and K. Kim, "Hybrid algorithm for multi people counting and tracking for smart surveillance," in *IEEE IBCAST*, Bhurban, Pakistan, pp. 530–535, 2021.
[76] J. Madiha, M. Gochoo, A. Jalal and K. Kim, "HF-SPHR: Hybrid features for sustainable physical healthcare pattern recognition using deep belief networks," *Sustainability*, vol. 13, no. 4, pp. 1–28, 2021.
[77] A. Hira, A. Jalal, M. Gochoo and K. Kim " Hand gesture recognition based on auto-landmark localization and reweighted genetic algorithm for healthcare muscle activities," *Sustainability*, vol. 13, no. 5, pp. 1–26, 2021.
[78] A. Mousavian, A. Arsalan, D. Anguelov, J. Flynn and J. Kosecka, "3D bounding box estimation using deep learning and geometry," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, San Juan, PR, USA, pp. 7074–7082, 2017.
[79] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang *et al., ShapeNet: An Information-Rich 3d Model Repository*, Stanford University, Princeton University, Toyota Technological Institute at Chicago, Technical Report, arXiv, vol. abs/1512.03012, pp. 1–11, 2015.
[80] H. Kato, Y. Ushiku and T. Harada, "Neural 3d mesh renderer," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 3907–3916, 2018.
[81] M. Mustafa, S. Alzubi and M. Alshare., "The moderating effect of demographic factors acceptance virtual reality learning in developing countries in the Middle East," in *Int. Conf. on Advances in Computing and Data Sciences*, Singapore, pp. 12–23, 2020.
[82] H. Sun and R. Grishman, "Employing lexicalized dependency paths based supervised learning for relation extraction," *Computer Systems Science and Engineering*, vol. 43, no. 3, pp. 861–870, 2022.
[83] M. Alshar'e, A. Albadi, M. Mustafa, N. Tahir and M. Al Amri, "Hybrid user evaluation methodology for remote evaluation: Case study of educational games for children during covid-19 pandemic," *Journal of Positive School Psychology*, vol. 6, pp. 3049–3063, 2022.
[84] M. Mustafa, M. Alshare, D. Bhargava, R. Neware, B. Singh *et al.,* "Perceived security risk based on moderating factors for blockchain technology applications in cloud storage to achieve secure healthcare systems," *Computational and Mathematical Methods in Medicine*, vol. 2022, pp. 1–10, 2022.