International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

3D Animation Using YoloV3

S. Thejaswini[1], Dr. T. Sreenivasulu Reddy[2]
[1]Student, Department of ECE, SVU, Tirupati, Andhra Pradesh, India
[2]Professor, Department of ECE SVU, Tirupati, Andhra Pradesh, India

ARTICLEINFO

Article History:

Accepted: 05 May 2023
Published: 16 May 2023

Publication Issue
Volume 10, Issue 3
May-June-2023

Page Number
159-164

ABSTRACT

Simulating and demonstrating a dynamic system can be done in a variety of ways. Our emphasis is on fabricating, comprehending and dragging the positive and negative attributes of three distinct methods that have been proportionate uncomplicated to implement with software that is both inexpensive and freely available, ranging from amalgam of MATLAB, A few well-known illustration players are Simulink, In conjunction with the preloaded animation function in MATLAB, their peers animation tools include Simulink 3D Animation, SolidWorks (basic), SolidWorks (Motion Manager), as well as Windows (Live) Media Maker. A MATLAB/Simulink Motion Manager-based animation registry may be used for animation creation. In this regard, the final SolidWorks data in the "Simulink 3D Animation" must include information ingested throughout the MATLAB environment and altered using the VRML Editor integrant in order to initiate the creation of geometric constraints that will be represented as an animation cyberspace sink block within the Simulink model of the dynamic system. Every scenario may be addressed with a You Only Look Once (YOLO) Version 3 prompt. These three distinct methods will be juxtaposed and appraised, a benchmark challenge was formulated: A tandem parking vehicle with four wheels and frontal steering.

Keywords—3D Animation, Simulink, VRML Editor.

I. INTRODUCTION

The automated processing, analysis, and comprehension of digital images and movies are under the combined responsibility of two associated fields called computer vision and image processing. The fields are closely related and share a lot of techniques and algorithms. The primary difference between the two is that computer vision aims to extract information from images to understand the world around us, while image processing is concerned with manipulating and improving images for better visualization or interpretation.

Computer Vision

The main objective of the branch of research known as computer vision is to make it possible for machines to understand and evaluate the visual environment people live in. It entails creating algorithms and methods that give computers the ability to detect and understand images and videos much like people do. Applications for computer vision systems include robotics, surveillance, autonomous vehicles, medical imaging, and more. Computer vision is a difficult field. Field that involves a combination of techniques from various areas of study, including mathematics, physics, statistics, and computer science. Image segmentation, spotting objects, image classification, and image recognition represent a few of the pursue degrees methods utilised within computer vision.

Image Processing

The specialty of image processing, on the other conjunction, emphasises improving and altering digital images to facilitate improved viewing or elucidation. It involves developing algorithms and techniques that can remove noise, enhance contrast, and sharpen images. Medical imaging, digital photography, and computer graphics are just a few of the applications that make use of image processing techniques. Enhancing the visual quality of an image is the main objective of image processing or make it easier to interpret. Some of the key techniques used in image processing include image filtering, image enhancement, and image restoration.

Computer Vision vs. Image Processing

Despite the fact that image processing and computer vision are remarkably comparable, there are also some key differences between the two fields. The primary difference is that computer vision is concerned with extracting information from images to understand the world around us, while image processing is concerned with manipulating and improving images for better visualization or interpretation. Another difference is the type of input data used in each field. Computer vision typically uses image or video data captured by cameras or sensors, while image processing can also work with other types of data, such as signals from medical equipment or satellite imagery. Finally, computer vision tends to use more advanced techniques and algorithms, such as machine learning and deep learning, to extract information from images, while image processing typically uses more traditional signal processing techniques.

The method of producing animated three-dimensional visuals in a computer setting is known as 3D animation. These images were produced using 3D software, which enables animators to produce digital things that appear three-dimensional despite being on a two-dimensional surface. Animation professionals can make anything appear to be moving evenly three dimensions, from a video game character to an automobile in an advertisement. This is done via the use of visual effects and exact timing.

YOLOv3 (You Only Look Once version 3) is primarily used for object detection and tracking in 2D images and videos. While YOLOv3 itself is not designed for 3D animation, it can be used as a tool to help automate the process of object tracking and motion capture, which are important components of 3D animation. In order to use YOLOv3 for 3D animation, you would need to first detect and track the movements of real-world objects or people in a video using YOLOv3. This can be done using YOLOv3's object detection and tracking capabilities, which can identify and track objects in a video frame-by-frame. Once you have tracked the movements of the objects or people in the video, you can use this data to drive the animation of corresponding 3D objects in a virtual space. This is typically done using motion capture techniques, which involve mapping the movements of real-world objects onto digital 3D models.

II. RELATED WORKS

A novel method of object detection was used in YOLOv1. Leveraging classifiers to detect objects has been done before. Using bounding boxes, the fact that objects reside miles apart and are matched with the correct class probabilities, beyond an object identification component, is done as a regression issue. A single neural network employs bounding boxes, class probabilities, and just one analysis of all the data to create predictions. The unified paradigm or a single-stage detector are both as simple as YOLOv1. You Only Look Once version 3 (YOLOv3) is primarily used for object recognition and tracking in 2D pictures and movies. A single neural network predicts multiple bounding boxes and class probabilities. While YOLOv3 itself is not designed for 3D animation, it can be used as a tool to help automate the process of object tracking and motion capture, which are important components of 3D animation. In order to use YOLOv3 for 3D animation, you would need to first detect and track the movements of real-world objects or people in a video using YOLOv3. This can be done using YOLOv3's object detection and tracking capabilities, which can identify and track objects in a video frame-by-frame. Once you have tracked the movements of the objects or people in the video, you can use this data to drive the animation of corresponding 3D objects in a virtual space. This is typically done using motion capture techniques, which involve mapping the movements of real-world objects onto digital 3D models. Those boxes alongside other tasks, negating the need for a convoluted pipeline with numerous steps [1].

YOLOv1 takes the whole image for detection and after scaling the model divides the input image into N equal-sized grids. The object that requires the centre of the object must be positioned inside a specific grid is determined by each cell or grid. The bounding box for every recognised object is composed of 5 variables, including the bounding box's coordinate location, its height and width, and two additional values. the most important value is the probability of the object belonging to the detected class. Probability determines the likeliness whether the discovered object belongs to a specific class. The identified object is classified as a stored in the form of the number. The output of the detected bounding box is stored in an array involving the 4 values of location along with detected class number and class probability. Thus, YOLO prediction has the total shape of N * N * no of boxes * (5+Probability).

YOLOv2 performs localization and classification in a single stage. Different to SSD, only one feature map is used for prediction and anchor box regression is modified by using a sigmoid function to predict the center coordinates. Instead of VGG-16, the authors propose Darknet-19 as base network. Darknet-19 comprises less parameters, which results in decreased inference time while the classification accuracy is comparable to VGG-16. To determine the anchor box dimensions, k-means clustering is employed [2].

The YOLO neural network integrates the candidate boxes extraction, feature extraction and objects classification methods into a neural network. The YOLO neural network directly extracts candidate boxes from images and objects are detected through the entire image features. Traffic sign detection refers to applying the candidate box extraction technique to the input image to determine whether it contains traffic signs and output their location. The data set we used in this paper for training and evaluation comes from ImageNet. None of the images are tagged and they are independent from the ones used in pre-training. We selected 1,652 images containing traffic signs as the data set. Then, 1,318 images were selected which form the standard training set, and the remaining 334 images were used as the test set. All traffic signs in the images are labelled [3].

To recognize the objects in real world images, shape features are used. Since most of the objects are better explained by their shape rather than texture, shape features are compared to local features like SIFT. Shape features are the supplement for the local features since the local features cannot avoid large amount of background mess [4].

III.METHODOLOGY

The approach used in the field for achieving animated objects and people in three dimensions (three-dimensional animation) occupies computer-generated imagery (CGI) and three-dimensional models. It involves designing and building 3D models, rigging them for animation, creating textures, lighting, and animating them in order to give the impression of depth and motion.

3D animation production can be a labor-intensive process broken down into several stages:

1. Pre-production: This stage involves conceptualizing and planning the project, including developing the story, characters, and environment. This stage also includes creating storyboards and animatics, which are rough visual representations of the animation.
2. Modeling: This stage involves creating 3D models of the characters, objects, and environment using specialized software. The models can be created from scratch or from pre-existing templates.
3. Texturing: This stage involves adding colour, texture, and other details to the 3D models. Texturing can be done using various techniques, including painting, sculpting, or applying pre-existing textures.
4. Rigging: This stage involves adding a digital skeleton or rig to the 3D models, which allows them to be animated. The rigging process involves creating bones, joints, and controls for the model.
5. Animation: This stage involves bringing the 3D models to life by animating them. The animator creates keyframes, which are the main poses or actions of the character, and then fills in the in-between frames to create smooth motion.
6. Lighting: This stage involves adding lighting and shadows to the scene to create the desired mood and atmosphere.
7. Rendering: This stage involves exporting the animation from the software into a final video file.

The graphics all appear flat because they only span the x- and y-axes in a 2D universe. The z-axis, which adds depth, is a crucial third axis in 3D animation. 2D animation may be seen in classic Disney animated films like "Sleeping Beauty" and "Bambi". In comparison, examples of 3D animation include "Frozen" and "Ice Age". This essential distinction can be attributed to the technical procedures used to produce 2D vs 3D animation. What distinguishes 3D animation from 2D animation? The animator draws a series of images on a flat surface to produce a 2D animation. By slightly changing the position of the animation figure between frames, which are then quickly played back to produce a dynamic image, the illusion of movement is produced.

In a variety of fields, such as film, television, video games, and advertising, 3D animation has emerged as a crucial tool. Toy Story, Shrek, and The Incredibles are just a few of the 3D animated films and TV shows that have achieved great popularity. In the world of video games, realistic and compelling surroundings and characters are made possible via 3D animation. Animations for cutscenes, trailers, and in-game play are all made in 3D by video game developers. The production of captivating and interesting commercials and promotional movies in the advertising sector uses 3D animation.3D animation allows advertisers to create highly realistic or stylized visuals, which can be used to demonstrate the features and benefits of a product or service. Animation is possible because of a biological phenomenon known vision of perseverance. The ability of the human eye to keep an image for a little length of time after it has vanished is known as the persistence of vision, which is a biological phenomenon. This phenomenon is a crucial component of how we perceive motion in the world around us, and it is the basis for many visual technologies, including film, television, and animation.

When an image is projected onto Upon impact with the retina of the eye, it causes chemical and electrical reactions in the retinal cells, which convey messages to the brain. The viewer experiences an image as a result of these impulses. However, even after the image is no longer being projected onto the retina, the cells continue to send signals to the brain for a fraction of a second. This creates a brief afterimage that appears to linger in the viewer's mind.

This afterimage is what allows us to perceive motion in a collection of still photos, such those found in a movie or television show. The afterimages that result from showing a series of images quickly one after another combine to provide the appearance of motion. This effect is called the phi phenomenon, and it is the foundation of animation.

The persistence of vision has been studied extensively by scientists and artists throughout history. Some of the earliest experiments were conducted by Dutch scientist Christiaan Huygens in the 17th century, who made the discovery that a spinning disc with numerous slits would appear to have a continuous image when viewed through the openings. The methodology was later enhanced and tapped in the beginning film projectors, which gave an illusion of motion by whirling the shutter mechanism. The longing of vision is a fundamental concept in many fields today, such as aesthetics, psychology, and neuroscience. It is also a fundamental part of several contemporary technologies, such as digital projectors, LED displays, and virtual reality equipment.

rendering enlivened illustrations that appear to be in two dimensions is known as "2D space animation." This type of animation is often used in traditional hand-drawn animation, as well as in digital animation software such as Adobe Flash and Toon Boom Harmony. In 2D space animation, images are created on a flat plane, with no apparent depth or volume. However, through the use of various techniques such as perspective, shading, and layering, the images can appear to have depth and movement.

Pseudo Algorithm:

- Creating one Virtual world
- Examine the Virtual World Properties
- Creating the YOLOv3 Version
- Performing the YOLOv3 real time virtual world
- Develop a VRNODE object connected to the VRML node "Automobile" that symbolises a representation of a vehicle on a road.
- In order to reach the next section of the road, we'll slightly turn the car.
- The 'rotation' property of the 'Automobile' node is set to achieve this.
- Accessing the Neural Networks based on the YOLO real time objects.

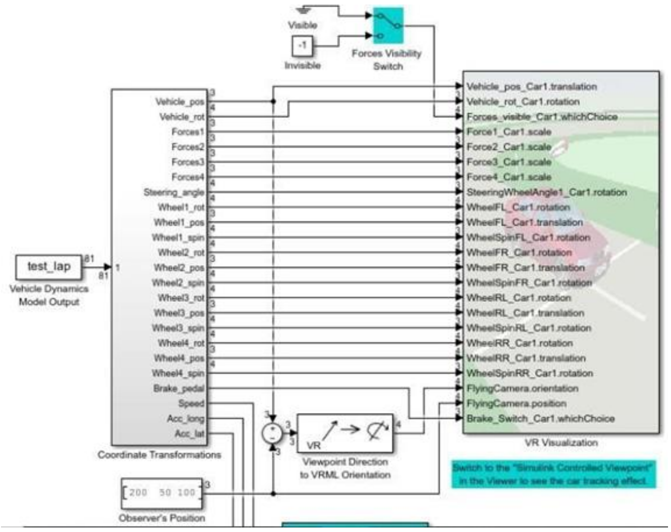
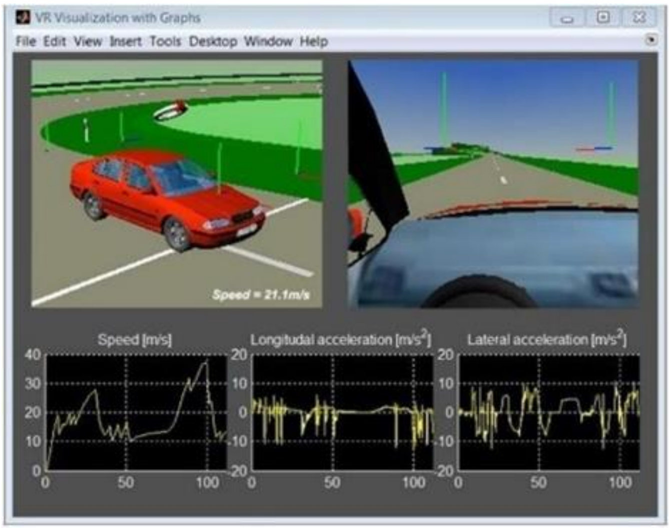


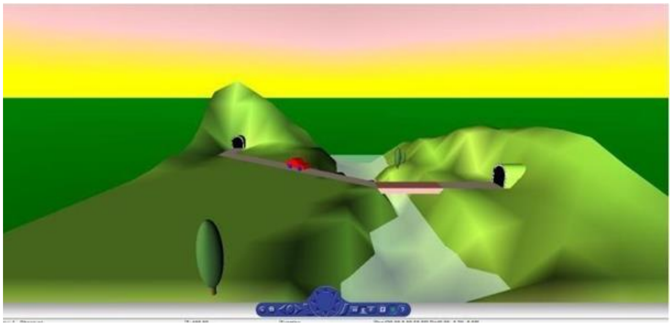
Fig 1: Vehicle Dynamics Visualization

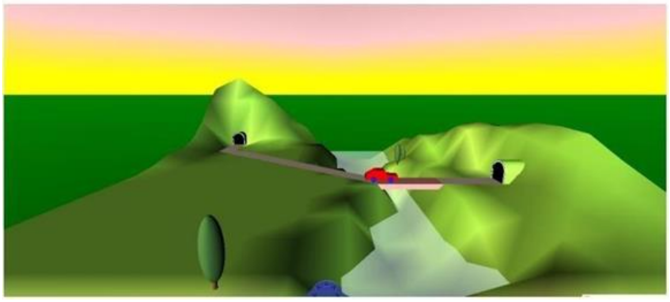
IV. RESULTS AND DISCUSSIONS

An average precision was used to measure and estimate the model. The object detection results are obtained by using YOLOV3.



Simulation in vehicles dynamics that 3D output in a synthetic camera attached to the car is processed using a video processing. That space mouse is used to control a manipulator in the virtual scene.





In this process, the VRML nodes are used to rotate the video in multiple directions. The 3D animation project involves mainly in the YOLOV3, it detects the visual real time objects by seeing and inspire audiences of all the ages. Finally importing the CAD packages such as solidworks and PTC creo for developing the 3D animations.

V. CONCLUSION

A combination of technical proficiency and artistic talent is needed to create 3D animation, which is a sophisticated and dynamic art form. Concept creation, modelling, texturing, rigging, animation, lighting, and rendering are just a few of the stages that go into making a 3D animation project. Each step necessitates meticulous preparation, close attention to detail, and teamwork skills. A successful 3D animation project requires a strong concept, engaging characters, and a compelling story that resonates with the audience. It also requires a deep understanding of the principles of animation, such as timing, spacing, and anticipation, as well as an ability to convey emotion and expression through movement and gesture. Advancements in technology and software have made 3D animation more accessible and efficient than ever before, but it remains a highly competitive field that demands a high level of skill and dedication. However, for those who are willing to put in the time and effort, the rewards can be great, both creatively and professionally. Overall, 3D animation is an exciting and constantly evolving field that offers endless possibilities for artistic expression and storytelling. Whether it's creating characters for a blockbuster movie or designing a virtual reality experience, 3D animation has the power to captivate and inspire audiences of all ages and backgrounds.

VI. REFERENCES

[1]. R. L. Simpson, "Computer vision: an overview," in IEEE Expert, vol. 6, no. 4, pp. 11-15, Aug. 1991, doi: 10.1109/64.85917.

[2]. E. Kutlugün and C. Eyüpoğlu, "Artificial Intelligence Methods Used in Computer Vision," 2020 5th International Conference on Computer Science and Engineering (UBMK), Diyarbakir, Turkey, 2020, pp. 214-218, doi: 10.1109/UBMK50275.2020.9219385.

[3]. M. Vashisht and B. Kumar, "A Survey Paper on Object Detection Methods in Image Processing," 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 2020, pp. 1-4, doi: 10.1109/ICCSEA49143.2020.9132871.

[4]. A. K. Shetty, I. Saha, R. M. Sanghvi, S. A. Save and Y. J. Patel, "A Review: Object Detection Models," 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-8, doi: 10.1109/I2CT51068.2021.9417895.

[5]. O. Acatay, L. Sommer, A. Schumann and J. Beyerer, "Comprehensive Evaluation of Deep Learning based Detection Methods for Vehicle Detection in Aerial Imagery," 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 2018, pp. 1-6, doi: 10.1109/AVSS.2018.8639127.

[6]. C. Liu, Y. Tao, J. Liang, K. Li and Y. Chen, "Object Detection Based on YOLO Network," 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 2018, pp. 799-803, doi: 10.1109/ITOEC.2018.8740604.

Cite this Article

S. Thejaswini, Dr. T. Sreenivasulu Reddy, "3D Animation Using YoloV3", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 3, pp. 174-180, May-June 2023.
Journal URL : https://ijsrst.com/IJSRST52310329