https://doi.org/10.2352/EI.2024.36.17.AVM-116

An approach to recreate large virtual environments for use in road traffic research

Jack Miller and Eliot Winer; Iowa State University; Ames, Iowa

Abstract

Traffic simulation is a critical tool used by psychologists and engineers to study road behavior and improve safety standards. However, the creation of large 3D virtual environments requires specific technical expertise that traditionally trained traffic researchers may not have. This research proposes an approach to utilize fundamental image processing techniques to identify key features of an environment from a top-down view such as satellite imagery and map. The segmented data from the processed image is then utilized to create an approximate 3D virtual environment. A mesh of the detected roads is automatically generated while buildings and vegetation are selected from a library based on detected attributes. This research would enable traffic researchers with little to no 3D modeling experience to create large complex environments to study a variety of traffic scenarios.

Introduction

The National Highway Traffic Safety Administration reported over 6.1 million traffic accidents in 2021, resulting in nearly two million injuries and forty thousand deaths [1]. These statistics clearly show a need to better understand the roadways and the various entities involved within them. Furthermore, with new technologies and infrastructure being introduced, such as autonomous vehicles and bicycle lanes, it is imperative that traffic safety researchers can understand and create regulations as quickly as possible for the safety of all. However, traffic safety research is often difficult to perform and, in many cases, may take years to complete before any actionable results are found [2], [3]. Traffic simulators are one approach to alleviate many of these issues as they allow researchers to observe participants in a controlled and immersive environment. One major drawback to this approach is that these Virtual Environments (VE) must be created which require expertise in a variety of domains and an extensive amount of time. A researcher may have a top-down image of a given environment and would have to determine what features they would like to replicate and then proceed to take 10s if not 100s of hours modeling the large 3D VE.

To mitigate this issue, this research seeks to substantially reduce the expertise and time required by automating the process from 2D top-down image to 3D VE. A variety of fundamental image processing techniques were implemented to detect key roadway features from a 2D map and satellite image such as roads, buildings, and vegetation. Once the images have been processed, a mesh of the roadways is automatically generated from the segmented data. Vegetation and buildings are then selected and instantiated from a library based on their calculated parameters. Through this process, a large 3D VE is created that recreation of the input 2D images. This research drastically streamlines the creation of VE for use in road traffic research and other applications in the future.

Background

The detection of traffic related entities from aerial imagery has been researched extensively for over two decades. Researchers have explored different techniques in approaches, sensors, and image processing algorithms to achieve this goal. Roadways are central to traffic environments and are the focus of research for many applications. The use of fundamental image processing techniques such as edge detection, Hough transform, morphological operations, and color thresholding have been used extensively for this purpose. He et al. utilized a color-based and edge detection hybrid methodology to detect roads in urban traffic environments [4]. Edge detection is used to detect the boarders of the road while color detection fills in the gaps. Experimental results suggested that the combination of both techniques helped compensate for the limitation in each method. Alternatively, Lin and Saripalli used probabilistic Hough Transform to detect line segments within aerial images that could be roads [5]. The researchers were able to detect roads in 97% of images. However, their research focused on the existence of roadways and not defining their dimensions within the image. More recently, machine learning techniques such as Convolutional Neural Networks (CNN) have become more popular for road detection [6], [7], [8]. Saito and Aoki developed a CNN to detect roads and buildings from large aerial imagery [9]. Despite producing promising results, CNNs required immense data sets and the test images had to be broken into smaller pieces to be processed and later stitched back together for the full environment. Similarly, Henry et al. used a fully CNNs for satellite imagery to detect roads. While they also found promising results, the network had to be tuned based on the type of environment the roads were located in. Despite CNNs showing potential for the detection of roads, they often require large datasets or are highly sensitive to new environments. More traditional image processing techniques do not reach the same level of accuracy; however, they tend to be more generalizable in terms of environments and do not require large amounts of data.

In addition to roadways, building detection is also a major area of research for image processing. Like roadways, a variety of approaches of been implemented for use cases related to traffic applications [10]. Shu et al. explored the use of CNNs to detect buildings of high density from aerial imagery [11]. The researchers were able to locate buildings with a high degree of accuracy but could only determine the center point. For this research, it is key to detect additional information such as the orientation and size of the building. Vegetation is another key component within traffic related environments. Despite far less research exploring the detection of vegetation for traffic related applications, various environmental science domains have explored the use of detecting trees and vegetation from aerial imagery for conservation purposes. Polewski et al. explored the use of color thresholding to detect dead trees from aerial imagery. Despite successfully segmenting out the dead trees, they relied on infrared imagery which would not be widely available to traffic researchers. Liu and An utilized a Hue, Saturation, Value

(HSV) color space with maximum likelihood classification and support vector machine to detect and classify trees with ~75% accuracy depending on the conditions [12]. These results are promising since the focus of this research is to simply detect vegetation and not classify it further.

The creation of VEs has been researched for as long as 3D rendering has existed. Despite advances in the creation of virtual content, it often requires expertise in a range of activities such as 3D modeling, animation, and user interface design [13]. Further, researchers have found that the creation of 3D environments require skills in 3D modeling and computer graphics which can be difficult to teach for those outside of technical fields [14]. Other approaches such as utilizing GIS data have been attempted [15], [16], [17]. However, this requires an additional skillset as GIS databases may be difficult to interface with or not accessible to traffic researchers. Finally, 3D reconstruction approaches such as Google Earth may be effective for many applications but visual artifacts such as cars and people baked into the roads and building textures will be detrimental for traffic simulation research [18], [19].

From this background research, it was discovered that there is a significant research gap in the accessible, fast, and intuitive creation of large 3D environments and models for traffic safety research. Fundamental image processing techniques may not be as accurate as more modern CNNs, however their application is more accessible and does not require immense data sets. Furthermore, depending on the implementation, they may be more robust to a variety of environment types. In addition, the creation of 3D environments is also not feasible unless you have a background in a related technical field. Streamlining this process would be greatly beneficial for those who utilize 3D VE but do not have the skillset to create them.

Methodology

This approach can be broken into two major steps. First, key environmental details such as roads, buildings, and trees were segmented from top-down 2D images using various image processing techniques. To achieve this, the OpenCV library for Python was utilized for the vast image processing support. The second major step was to then recreate the VE from data extracted from the 2D images. The Blender Python API was utilized for the creation of the road due to its various operations to create and modify meshes [20]. Other environmental details were instantiated from a library based on various parameters. The following section will detail the approach taken to achieve the end goal.

Input Data

Since the scope of this research was to simplify the process of a traffic safety researcher to go from a 2D top-down image to a 3D VE, the input data was based on publicly available imagery. This included a map view and satellite view of a given location from applications such as Google Maps or Open Street Map, see Figures 1 and 2. Alternatively, the researcher may use a sketch of a novel environment if the location does not actually exist. The images are processed to extract the necessary information to then recreate the environment in 3D.

Figure 1. *Example input satellite image.*

Figure 2. *Example input map image.*

Road Detection

The first image processing step is to identify and segment the roads. This is accomplished by leveraging the map image and extracting out the specific color of the road. The map is converted to greyscale and then a thresholding operation is applied. The resulting image captures a large majority of the roadways included in the image. However, a small amount of noise may exist depending on the location and image quality. Therefore, an erosion operation is applied to remove various artifacts followed by a dilation operation to undo the effects of the erosion on the remaining roadway. The final segmented image, shown in Figure 3, is then exported as a Scalable Vector Graphic (SVG) to be used later in the creation of the road mesh.

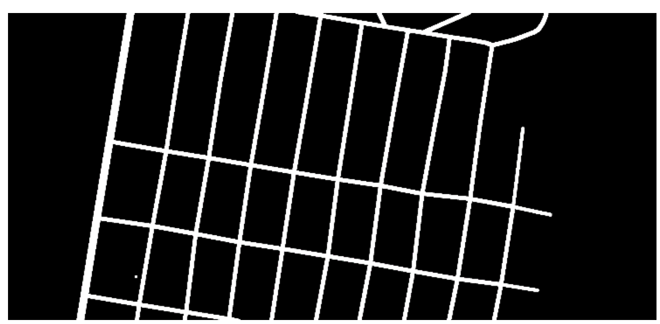

Figure 3. *Resulting road segmentation of example images.*

Building Detection

The next step is to identify buildings within the images so the appropriate model may be selected from a library. Like the detection of the roadways, the map image is leveraged by converting to greyscale and thresholding for the building color. Since there is less contrast between the buildings and the remaining image, additional noise is captured but most of the buildings are still identified. Again, morphological operations of erosion and dilation are used to clean up the image. From the resulting image, contours are detected, and the location, orientation, and areas are calculated. Figure 4 shows an example where the detected contours are highlighted in green. The orientation is determined by calculating the vector to the nearest roadway. To further refine the results of this step, contours with area values less than 50 square feet are removed as they are likely too small to be a building.

Figure 4. *Detected buildings outlined in green.*

Vegetation Detection

Vegetation and trees are additional environmental details that are critical to the immersion of the VE. For the detection of vegetation, satellite images are primarily used. First, the image is thresholded based on a color range in the HSV color space. This allows the thresholding process to be less sensitive to different lighting conditions as the hue and saturation is separated from the value of the light intensity. The specific color range used was based on prior research and experimental results [12], [21]. Once the image is thresholded, the output is the segmented area of vegetation candidates. Since the image is a top-down view, vegetation such as trees and bushes will appear circular. Therefore, Hough transform is then used to detect circular objects in the candidate locations. Parameters for minimum radius, maximum radius, and minimum distance from each other are applied to further refine the results. At this point in the process, the location and size of potential vegetation is calculated. The final refinement step is to apply a mask of the road and buildings segmentation results and remove any vegetation detected in those locations. Effectively, buildings are roads are prioritized over trees as they are more critical to the usage of the final VE. Figure 5 shows an example of the results. While not every instance of vegetation is detected, this method provides a consistent representation of the distribution of vegetation in each area.

Figure 5. *Detected trees circled in green.*

The final image processing step was to calculate the pixel to distance ratio of the image and estimate the type of environment. The pixel to distance ratio is used to properly scale parameters and the mesh output. Tesseract OCR was implemented to detect and extract the text information of the scale, e.g., 100ft. The contour next to the text is thresholded and the width is measured. The ratio between distance and pixels can then be established based on the number value detected, and the width of the contour in pixels. For example, 100 ft per 1000 pixels. Environment types for this research were broken into three major groups: grass, desert, and city. K-Means clustering was implemented to automatically cluster the colors within the image and determine the dominant color. The dominant color is then sorted into the three environment type groups based on its value in the HSV color space. Again, HSV was used to limit the effects of lighting conditions in satellite imagery.

Once the major image processing operations are complete and the key features are extracted from the image, the second major step involves recreating the 3D VE. To achieve this, the Blender API for Python was utilized for various functionalities related to 3D modeling and the manipulation of meshes. The segmented road exported as an SVG is used by the Blender API to convert the curves of the SVG into a flat mesh object. The flat mesh is then extruded upwards to complete the full 3D model of the road, see Figure 6. However, the conversion from a segmented raster image to a vector image and then into a mesh result in numerous artifacts. The main error tends to be in the triangulation of the mesh, often in the form of extremely dense locations of polygons in corners and small curves. Therefore, a decimation operation is used to substantially reduce the polygon count (up to 95%) to smooth out the locations of erroneously high polygons while maintaining the same overall geometry. A retriangulation operation is then applied to clean up any remaining noise. Figure 7 shows an example of the processed mesh object. The polygon count in this case was reduced to 4,356 for the entire roadway. This example shows the efficient triangulation of the mesh; however, it should also be noted that some artifacts do remain such as intersecting triangles. Although, since the mesh is relatively well optimized at this point, the resulting issues should be negligible. The segmented road and tree data is then used to select an appropriate model from a pre-defined library of models. The shape, size, orientation, and location of the model is matched appropriately. Similarly, the terrain is textured based on the results of the K-Means clustering. The road model, buildings, trees, and terrain are then brought together in a single scene to recreate the full environment in 3D, shown in Figure 8.

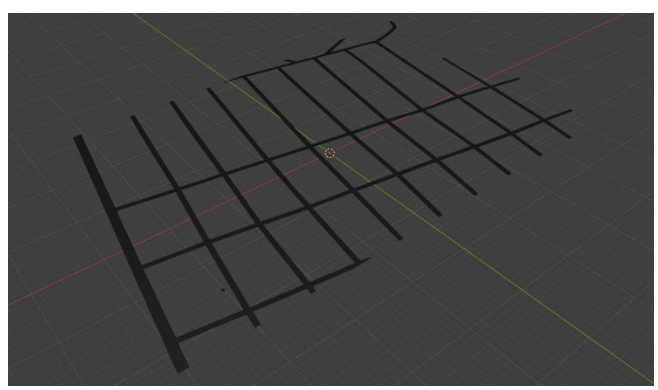

Figure 6. *Mesh model of roadways.*

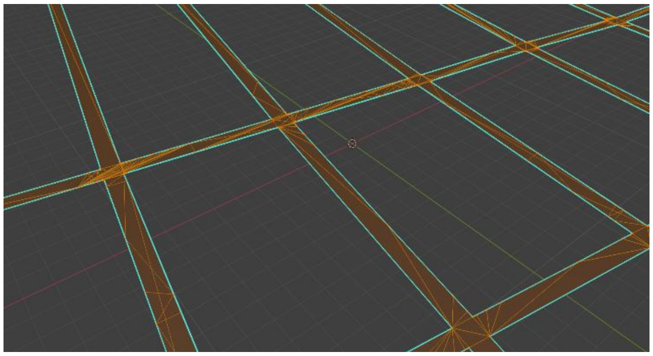

Figure 7. *Example triangulation of road mesh.*

Discussion

The approach described in this research results in the recreation of large 3D environments from 2D images. Roads, buildings, and trees are identified from the segmented image and represented in a virtual environment with appropriate attributes. The example given in this paper is only one such environment created. Figures 1-8 show the various processes that are undergone throughout the creation of the 3D VE. This approach can be applied to several real-world environments and novel sketches. Through this research, traffic safety researchers will be able to create large 3D environments for use in traffic simulation.

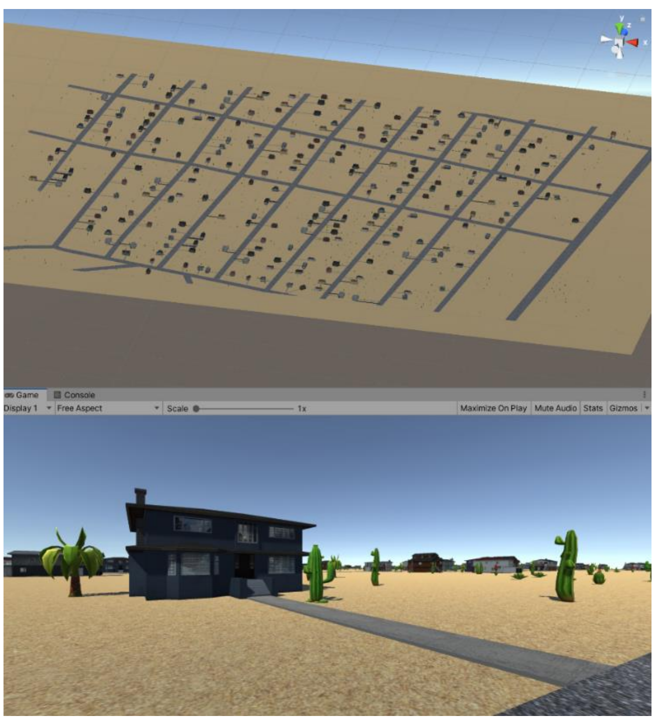

Figure 8. *Final output of VE creation.*

Conclusion and Future Work

This research focuses on an approach to empower traffic researchers to create large 3D environments without the need for expertise in 3D modeling and related software skills. The process described in this paper streamlines the workflow for traffic research in creating novel VEs by automating the detection of key environmental features and recreation of those features for use in a VE. Further, by avoiding heavy machine learning methodologies, the requirement for immense training data sets is eliminated and the computational demand of the described processes remains low. This will enhance usability and potentially allow future researchers to make additional modifications and enhancements as needed.

Future work of this research will expand on the current capabilities of this approach from both an image processing and computer graphics perspective. It would be ideal to eliminate the need for a map image and use only a satellite or high-altitude image. This would enable more use cases as there may be areas or data sets that do not include a map representation. Similarly, it would also be advantageous to detect other environmental features such as road markers and other scenery. Finally, once the VE is created, it would be useful to have a basic GUI that allows the user to make tweaks and adjustments to the environment itself. Currently, the user would have to go back into 3D modeling software to do so, this would further streamline the creation processing.

Acknowledgements

The authors would like to acknowledge the Turner-Fairbanks Federal Highway Research Center for providing funding support on this project.

References

[1] National Highway Traffic Safety Administration, “Early Estimates of Motor Vehicle Traffic Fatalities and Fatality Rate by Sub-Categories 2021,” May 2022. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813298

[2] D. Shinar, *Traffic Safety and Human Behavior: Second Edition*. Emerald Group Publishing, 2017.

[3] L. Zheng, T. Sayed, and F. Mannering, “Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions,” *Anal. Methods Accid. Res.*, vol. 29, p. 100142, Mar. 2021, doi: 10.1016/j.amar.2020.100142.

[4] “Color-based road detection in urban traffic scenes | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Dec. 19, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1364007

[5] Y. Lin and S. Saripalli, “Road detection from aerial imagery,” in *2012 IEEE International Conference on Robotics and Automation*, May 2012, pp. 3588–3593. doi: 10.1109/ICRA.2012.6225112.

[6] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, “Building detection in very high resolution multispectral data with deep learning features,” in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2015, pp. 1873–1876. doi: 10.1109/IGARSS.2015.7326158.

[7] T. Li, M. Comer, and J. Zerubia, “Feature Extraction and Tracking of CNN Segmentations for Improved Road Detection from Satellite Imagery,” in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 2641–2645. doi: 10.1109/ICIP.2019.8803355.

[8] C. Henry, S. M. Azimi, and N. Merkle, "Road Segmentation in SAR Satellite Images With Deep Fully Convolutional Neural Networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1867–1871, Dec. 2018, doi: 10.1109/LGRS.2018.2864342.

[9] S. Saito and Y. Aoki, "Building and road detection from large aerial imagery," in *Image Processing: Machine Vision Applications VIII*, SPIE, Feb. 2015, pp. 153–164. doi: 10.1117/12.2083273.

[10] N. Paparoditis, M. Cord, M. Jordan, and J.-P. Cocquerez, "Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery," *Comput. Vis. Image Underst.*, vol. 72, no. 2, pp. 122–142, Nov. 1998, doi: 10.1006/cviu.1998.0722.

[11] Z. Shu, X. Hu, and J. Sun, "Center-Point-Guided Proposal Generation for Detection of Small and Dense Buildings in Aerial Imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 7, pp. 1100–1104, Jul. 2018, doi: 10.1109/LGRS.2018.2822760.

[12] H. Liu and H. An, "Urban Greening Tree Species Classification Based on HSV Colour Space of WorldView-2," *J. Indian Soc. Remote Sens.*, vol. 47, no. 11, pp. 1959–1967, Nov. 2019, doi: 10.1007/s12524-019-01028-z.

[13] R. Dörner, M. Kallmann, and Y. Huang, "Content Creation and Authoring Challenges for Virtual Environments: From User Interfaces to Autonomous Virtual Characters," in *Virtual Realities: International Dagstuhl Seminar, Dagstuhl Castle, Germany, June 9-14, 2013, Revised Selected Papers*, G. Brunnett, S. Coquillart, R. van Liere, G. Welch, and L. Váša, Eds., in Lecture Notes in Computer Science. , Cham: Springer International Publishing, 2015, pp. 187–212. doi: 10.1007/978-3-319-17043-5_11.

[14] T. K. Coşkun and G. F. Deniz, "The contribution of 3D computer modeling education to twenty-first century skills: self-assessment of secondary school students," *Int. J. Technol. Des. Educ.*, vol. 32, no. 3, pp. 1553–1581, Jul. 2022, doi: 10.1007/s10798-021-09660-y.

[15] S. Zhang and A. B. Moore, "The Usability of Online Geographic Virtual Reality for Urban Planning," in *Innovations in 3D Geo-Information Sciences*, U. Isikdag, Ed., in Lecture Notes in Geoinformation and Cartography. , Cham: Springer International Publishing, 2014, pp. 225–242. doi: 10.1007/978-3-319-00515-7_14.

[16] J. Keil, D. Edler, T. Schmitt, and F. Dickmann, "Creating Immersive Virtual Environments Based on Open Geospatial Data and Game Engines," *KN - J. Cartogr. Geogr. Inf.*, vol. 71, no. 1, pp. 53–65, Mar. 2021, doi: 10.1007/s42489-020-00069-6.

[17] Z. Lv, T. Yin, X. Zhang, H. Song, and G. Chen, "Virtual Reality Smart City Based on WebVRGIS," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1015–1024, Dec. 2016, doi: 10.1109/JIOT.2016.2546307.

[18] D. Anguelov *et al.*, "Google Street View: Capturing the World at Street Level," *Computer*, vol. 43, no. 6, pp. 32–38, Jun. 2010, doi: 10.1109/MC.2010.170.

[19] N. Bruno and R. Roncella, "ACCURACY ASSESSMENT OF 3D MODELS GENERATED FROM GOOGLE STREET VIEW IMAGERY," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLII-2-W9, pp. 181–188, Jan. 2019, doi: 10.5194/isprs-archives-XLII-2-W9-181-2019.

[20] B. Foundation, "blender.org - Home of the Blender project - Free and Open 3D Creation Software," blender.org. Accessed: Apr. 10, 2023. [Online]. Available: https://www.blender.org/

[21] M. H. Almeer, "Vegetation Extraction from Free Google Earth Images of Deserts Using a Robust BPNN Approach in HSV Space," vol. 1, no. 3.

Author Biography

Jack Miller is a Ph.D. candidate in Computer Engineering and Human-Computer Interaction at Iowa State University's Virtual Reality Applications Center. His current research interests include exploring multi-user experiences in virtual and augmented reality environments, specifically for traffic simulation.

Eliot Winer is the director of VRAC and professor of Mechanical Engineering, Electrical and Computer Engineering, and Aerospace Engineering at Iowa State University. Dr. Winer has over 20 years of experience working in virtual reality and 3D computer graphics technologies on sponsored projects for the Department of Defense, Air Force Office of Scientific Research, Department of the Army, National Science Foundation, Department of Agriculture, Boeing, John Deere, and the Federal Highway Administration.