Northumbria Research Link

3D Car Shape Reconstruction from a Single Sketch Image

Naoiki Nozawa
s112800563@akane.waseda.jp
Waseda University
Tokyo, Japan

Hubert P. H. Shum
hubert.shum@northumbria.ac.uk
Northumbria University
Newcastle upon Tyne, UK

Edmond S. L. Ho
e.ho@northumbria.ac.uk
Northumbria University
Newcastle upon Tyne, UK

Shigeo Morishima
shigeo@waseda.jp
Waseda Research Institute for Science and Engineering
Tokyo, Japan

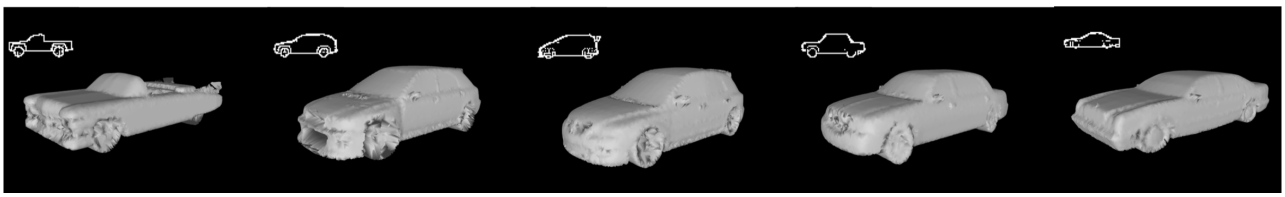


Figure 1: The 3D car shapes generated by our system from 2D sketch inputs.

ABSTRACT

Efficient car shape design is a challenging problem in both the automotive industry and the computer animation/games industry. In this paper, we present a system to reconstruct the 3D car shape from a single 2D sketch image. To learn the correlation between 2D sketches and 3D cars, we propose a Variational Autoencoder deep neural network that takes a 2D sketch and generates a set of multi-view depth & mask images, which are more effective representation comparing to 3D mesh, and can be combined to form the 3D car shape. To ensure the volume and diversity of the training data, we propose a feature-preserving car mesh augmentation pipeline for data augmentation. Since deep learning has limited capacity to reconstruct fine-detail features, we propose a lazy learning approach that constructs a small subspace based on a few relevant car samples in the database. Due to the small size of such a subspace, fine details can be represented effectively with a small number of parameters. With a low-cost optimization process, a high-quality car with detailed features is created. Experimental results show that the system performs consistently to create highly realistic cars of substantially different shape and topology, with a very low computational cost.

CCS CONCEPTS

• **Computing methodologies** → **Computer graphics**; *Shape modelling*; • **Computer vision problems** → Reconstruction.

KEYWORDS

Deep Learning, Lazy Learning, 3D Reconstruction, Sketch-based Interface, Car

ACM Reference Format:

Naoiki Nozawa, Hubert P. H. Shum, Edmond S. L. Ho, and Shigeo Morishima. 2019. 3D Car Shape Reconstruction from a Single Sketch Image. In *MIG '19: ACM SIGGRAPH Conference on Motion, Interaction and Games, Oct 28–30, 2019, Newcastle upon Tyne, UK*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

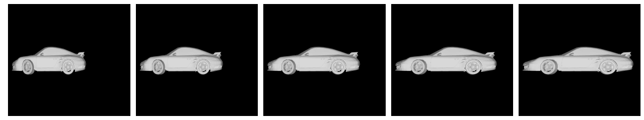
1 INTRODUCTION

Car shape design is a common area in automotive manufacturing, computer animation and games. The design process is time-consuming and labour intensive, as it is a combination of arts and engineering. In this paper, we propose a new 3D car design interface that is based on a single 2D sketch, which contains only the outline information on the cars' shape. Since a single outline sketch cannot provide enough information on 3D car reconstruction, our new framework estimates such missing information from a 3D car shape database. The major contributions of this paper are summarized as follows:

- We propose a feature-preserving mesh augmentation framework to construct a large car database with pairwise 3D mesh and 2D sketch, based on the small number of car meshes in ShapeNet [2].
- We propose a Variational Autoencoder (VAE) [3] deep learning network to learn the correlation between a 2D sketch and the corresponding rough 3D shape.
- We propose a lazy learning algorithm to learn a local subspace to reconstruct the fine detail features of the car from the rough shape.

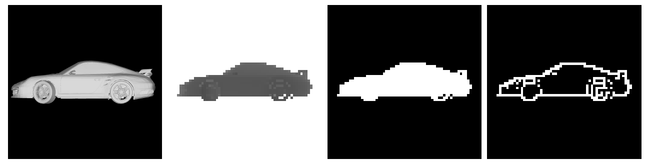
2 DATABASE CREATION

As a data-driven approach, the diversity and quality of the database is key to our system. With a novel feature-preserving data augmentation techniques[4], we create a large variety of logically correct car meshes as indicated in Figure 2. We set the resizing parameter as ±20%, ±15%, ±10% and ±5% of each scalable direction. They are converted into two sets of representations: (1) side-view 2D sketch, multiple depth and mask images (side, top, front and rear views) for shape reconstruction, and (2) registered 3D point clouds for details synthesis. For the registration, we pick a random point cloud of a car as a template, and then evaluate the Earth Mover's Distance (EMD) between the template and the rest of the cars. Since we have the same number of points for all point clouds, the optimal flow from the template to the target point cloud of a car has a 1-to-1 correspondence, which therefore is considered as the registration result.



Z-axis −20% **Z-axis −10%** **Original** **Z-axis +10%** **Z-axis +20%**

Figure 2: Examples of resized 3D car meshes.



(a) 3D mesh **(b) Depth image** **(c) Mask image** **(d) Sketch**

Figure 3: Different types of images generated from 3D mesh.

3 ROUGH CAR SHAPE RECONSTRUCTION

We adapt Variational Autoencoder (VAE) [3] for getting the depth and mask images as indicated in Figure4. While VAE is inferior to Generative Adversarial Network (GAN) in terms of the appearance of the output, it is difficult to control the image synthesis process in GAN to create multiple outputs. Also, it takes much longer to train GAN and to guarantee network convergence. In our situation, we prefer VAE as it produces results with high enough quality to generate a car shape, and the details of the car are introduced as a second stage process. The loss function is expressed as:

$$
\begin{aligned}
E = & \left((Depth^{ref} - Depth^{rec}) \circ Mask^{ref}\right)_{L1} \\
& + \left(Mask^{ref} - Mask^{rec}\right)_{BinaryCrossEntropy} \\
& + \left(\Delta(Depth^{ref} \circ Mask^{ref}) - \Delta(Depth^{rec} \circ Mask^{ref})\right)_{L1} \\
& + KLLoss
\end{aligned} \tag{1}
$$

where $Depth^{ref}$ and $Mask^{ref}$ are ground truth images, $Depth^{rec}$ and $Mask^{rec}$ are reconstructed images, the subscripts *L1* and *BinaryCrossEntropy* represent the calculation metrics, Δ means Laplacian filtering and $\circ$ is the Hadamard product, *KLLoss* is the standard KL loss function.

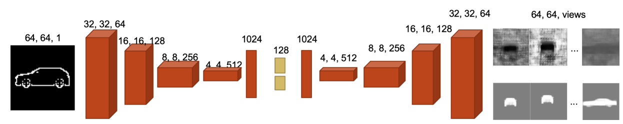


Figure 4: The encoder-decoder network structure.

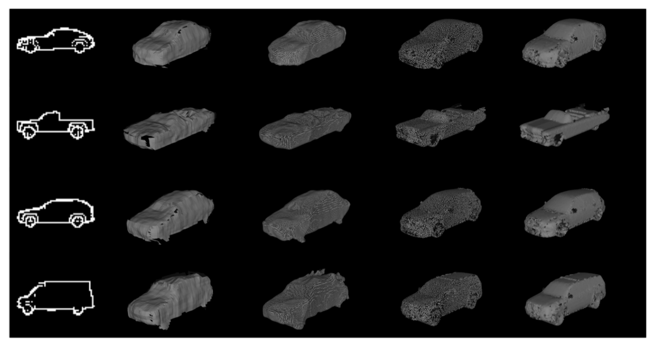


Figure 5: From left to right: sketches, meshes from generated depth images, registered point clouds, detailed added point clouds, final output meshes.

4 LAZY LEARNING FOR FINE DETAILS

Given a car shape generated in Section 3, we search for the k nearest samples from the database. We first convert the generated car shape into a point cloud using the same process we described in Section 2. We apply Principal Component Analysis (PCA) onto the z-score representation of the point clouds to generate a search space, instead of using the Cartesian space. With the k nearest neighbours selected from the database, we can then learn a small subspace with PCA. In such a subspace, we optimize a set of eigenvalues with [1] to construct a car shape that is as similar as possible to the rough shape. We then back project the eigenvalues to formulate a car shape with shape details such as the headlight, which is served as our final output showed in Figure 5.

REFERENCES

[1] Volker Blanz and Thomas Vetter. 1999. A Morphable Model for the Synthesis of 3D Faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '99)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 187–194. https://doi.org/10.1145/311535.311556

[2] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. 2015. *ShapeNet: An Information-Rich 3D Model Repository.* Technical Report arXiv:1512.03012 [cs.GR]. Stanford University — Princeton University — Toyota Technological Institute at Chicago.

[3] Durk P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. 2014. Semi-supervised learning with deep generative models. In *Advances in neural information processing systems*. 3581–3589.

[4] Vladislav Kraevoy, Alla Sheffer, Ariel Shamir, and Daniel Cohen-Or. 2008. Non-homogeneous Resizing of Complex Models. In *ACM SIGGRAPH Asia 2008 Papers (SIGGRAPH Asia '08)*. ACM, New York, NY, USA, Article 111, 9 pages. https://doi.org/10.1145/1457515.1409064