



Enhancing the Customer Experience by Mixed Reality in the Retail Industry

Yirui Jiang

Cranfield Centre for Competitive Creative Design, School of Water, Energy and Environment Cranfield University, Bedfordshire, UK

Trung Hieu Tran*

Cranfield Centre for Competitive Creative Design, School of Water, Energy and Environment Cranfield University, Bedfordshire, UK

Leon Williams

Cranfield Centre for Competitive Creative Design, School of Water, Energy and Environment Cranfield University, Bedfordshire, UK

Jaime Palmer

OrangeLV, Inc., Seattle WA, USA

Edgar Simson

OrangeLV, Inc., Seattle WA, USA

Daniel Benson

OrangeLV, Inc., Seattle WA, USA

Michael Christopher

OrangeLV, Inc., Seattle WA, USA

Daila Christopher

OrangeLV, Inc., Seattle WA, USA

ABSTRACT

*Nowadays, customization by mixed reality to enhance the customer experience plays an important role in the retail industry. Customers can choose and customize products with their images and labels in a virtual reality environment. However, the existing asset creation pipelines are labor-intensive and time-consuming to display the images and labels (aka logos) on 3D product models, and cannot be easily customized by customers in real-time. In this paper, we thus propose a real-time 3D logo mapping framework for converting 3D logo mesh from a specified image and fitting it to the 3D product models. In the framework, Convolutional Neural Network (CNN) is adopted to reconstruct 3D logo/product models from their images. The detailed 3D information and the logo location provided by a customer are used to select the effective sampling points to mesh deformation. This method can preserve both the visual quality and details of 3D product models. Experimental results, carried out on various sizes of logos and types of products, show that our method can produce accurately and quickly customized logos on 3D product models.

CCS CONCEPTS

• **Human-centered computing**; • **Mixed/augmented reality**; • **Virtual reality**; • **Graphical user interfaces**; • **Visualization systems and tools**;

KEYWORDS

Mixed reality, Convolutional neural network, Customization, Logo, Retail industry

*Corresponding Author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

AIVR 2021, July 23–25, 2021, Kumamoto, Japan

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8414-8/21/07...\$15.00

<https://doi.org/10.1145/3480433.3480438>

ACM Reference Format:

Yirui Jiang, Trung Hieu Tran, Leon Williams, Jaime Palmer, Edgar Simson, Daniel Benson, Michael Christopher, and Daila Christopher. 2021. Enhancing the Customer Experience by Mixed Reality in the Retail Industry. In *2021 5th International Conference on Artificial Intelligence and Virtual Reality (AIVR) (AIVR 2021), July 23–25, 2021, Kumamoto, Japan*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3480433.3480438>

1 INTRODUCTION

Customization plays an important role in the retail industry, from fashion to home décor. With the boom of Augmented reality (AR) applications, AR-based customization has the potential to transform the customer experience. It can allow customers to select their preferences and tailor them for specific needs. Though many brands have hopped onto the AR bandwagon, a few of them focus on logo customization. Logos are an integral part of brand identity, and intangible assets to show customers' unique taste. Manually reconstructing a 3D logo from customized images is a labor-intensive and time-consuming process. Several companies adopt photogrammetry methods [1] and reconstruct the 3D model from hundreds of photos. Some companies simply overlap the customer-designed flat images with product photos, or even not providing customization services. Both methods are reducing consumer believability and conversion rates. In addition, with the large numbers of input photos and strict shooting angle, it is challenging to learn the mapping between 2D logo and 3D product model in real-time. Designers generally need to take several hours to complete 3D reconstruction with carefully selected photos. Each of those asset creation pipelines requires extraordinary human work and cannot be easily self-serviced by the customers, making logo customization more costly.

In this paper, we propose a new 3D logo mapping framework for converting 3D logo mesh from a specified image and fitting it to the 3D product models, which can be successful in a few seconds. The key idea of our framework is that we use an end-to-end deep learning architecture to generate 3D product mesh representation from a few color images with good details [2]. Though the generated 3D product model is closely related to the 3D logo surface, it can only provide a weak representation of user-selected point clouds. Therefore, it is challenging to design an effective approach to sampling

points for further mesh deformation and logo reconstruction. We provide customers a user interface to set the location, sizing, and pattern of their logo. For mapping the logo with a specified size, we adopt a K-D Tree range search algorithm [3] to judge whether a point on the 3D product model is plausible for reconstruction or not, which greatly reduces the computation cost in next step. In the surface fitting module, we re-sample particular physical dimensional points centered at a user-selected point from the range of searched point clouds by using a classic yet effective quadric surface fitting algorithm [4]. The sampled point clouds are then used to fit a plane of a specified size in the mesh deformation module by using an as-rigid-as-possible algorithm [5]. Finally, we project the customized image onto the reconstructed 3D logo mesh. The user can easily alter the position and orientation of these images. To the best of our knowledge, the proposed system is the first one to real-time display the customized logos on 3D product models from a single image. Experimental results show that our framework can directly reconstruct logo from a few color images with high visual quality, and demonstrate the applicability of our approach across different product categories on the ShapeNet dataset [6] and our industrial partners’ physical products. Compared to other existing asset creation approaches, our framework has shown state-of-the-art of results with great time saving and enable self-service by customers.

In summary, we make the following contributions:

- We propose a new 3D logo mapping framework for converting 3D logo mesh from a specified image and fitting it to the 3D product models.
- This is the first system for real-time self-serviced logo customization in the retail industry.
- It can be applicable for a variety of product categories including the ones in the ShapeNet dataset [6] and the real-world products with a textured surface.

2 RELATED WORKS

In this section, we review brands that hop onto the AR bandwagon to reimagine customer experiences. We also study multi-view 3D mesh generation and mesh deformation for 3D logo mapping framework.

2.1 AR in Retail Industry

In recent years, we have witnessed the fast boom of AR in various applications. Many big brands are experimenting with immersive technology. For example, Converse, Nike, H&M, Amazon, Wayfair, IKEA, and Ashley are using AR in the physical retail environment for future shopping to achieve a better customer experience [7]. However, most of the brands only focus on the whole physical products, but rarely consider the logo customization. They use simpler visualization methods, which is lacking visual quality and accuracy, needless to say, details of the products. In fact, logos are everywhere, on the clothes we wear or on each product we buy, and very important to companies and customers. Companies use logos for sufficient brand recognition, and customers use them to show their unique style. However, existing logo asset creation pipelines are costly in both human labor and time-consuming.

2.2 3D Mesh Generation from Color Images

CNN-based 3D reconstruction has demonstrated an impressive performance in industrial applications. A variety of deep learning network architectures have been utilized to generate 3D models with an output of volumetric representations [8-10], mesh representations [11-12], or point cloud representations [13]. To be able to apply widely for industries with memory efficiency, we adopt deformation-based mesh representations to reconstruct our 3D product model from color images. Multi-view shape generation uses fewer input views for improving the reconstructed model quality. Choy *et al.* [6] proposed a sequentially object generation framework. Kar *et al.* [14] built a multi-view stereo machine. Gwak *et al.* [10] studied shapes from silhouettes by Generative Adversarial Network (GAN). Chao *et al.* [2] provided a coarse to fine strategy. Our approach uses the method of Chao *et al.* [2], a cross-view graph convolutional network (GCN), which can produce accurate details with good generalization capability.

2.3 Mesh Deformation

For mesh deformation, various energy formulations have been proposed to keep transformation as rigid as possible. The classical elastic energy formulation, e.g., shell energy, defines the surfaces with local shape preservation. Kobbelt *et al.* [15] and Botsch *et al.* [16] proposed a linearized energy formulation, which deforms smooth surface with later added high-frequency details. In this paper, we use the method of Olga *et al.* [5], an iterative minimization scheme, related to Laplacian surface deformation techniques, for mesh deformation in the proposed framework.

3 METHODOLOGY

The 3D logo mapping framework proposed in this paper consists of two main components: (i) an end-to-end multi-view shape generation module and (ii) a mesh deformation module. Figure 1 is an overview of our framework. The multi-view shape generation module generates a 3D model of the product from multiple input images. The output is directly shown on our user interface. The graphical user interface is created for the customers to set the location and size of the customized logo. Every point on the 3D product model within a specified range is detected and used for the surface fitting module, where quadric surface fitting is applied to keep details of the product surface. The new sampled points exposed on limited areas are passed to the mesh deformation module. We use the method of [5] to keep the transformation for the logo surface as rigid as possible. The pattern selected by the user is then mapped on self-customized logo mesh as the texture in the 3D product model, which helps retain the details and curvatures on the product’s surface. The details of each part of our system will be elaborated in the following sub-sections.

3.1 Multi-view Shape Generation Network

In this sub-section, we introduce a multi-view shape generation network, using the cross-view information to train GCN based model [2]. We compare and analyze some state-of-the-art 3D shape generation methods, e.g. 3DR2N2 [6], LSM [14], P2M [12], and P2M++ [2] on a subset of the ShapeNet dataset provided by Choy *et al.* [6]. The comparison is based F-score where $\tau = 10^{-4}$. Figure 2 shows F-score

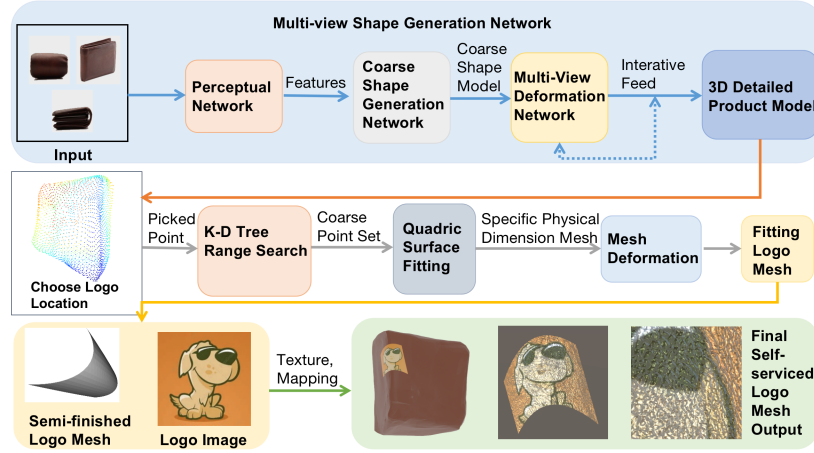


Figure 1: An overview of the 3D logo mapping framework.

on each semantic category (where better performance is for a higher F-score). We refer P2M++ to reconstruct a 3D product model from several color photos, which achieves better surface details. In the P2M++ method, a coarse-to-fine strategy is adopted to generate a rough shape at first, and then the details of the model are added iteratively. Each vertex is sampled of deformation hypotheses, and the VGG-16 architecture [2] is used to extract perceptual features with known camera intrinsic and extrinsic. Concatenation and statistics are used to learn from cross-view feature correlations, and a 1347 dimension feature vector is computed. A differentiable network component is then designed to infer a desirable deformation for each vertex. The detailed structure of the multi-view deformation network is shown in Figure 3. Originated from Pixel2Mesh [12], our loss function adopts the Chamfer loss [13] to constrain the location of mesh vertices, $l_c = \sum_p \min_q \|p - q\|_2^2 + \sum_q \min_p \|p - q\|_2^2$, where p is a vertex in the predicted mesh, and q is a vertex in the ground truth mesh. The loss function uses normal loss l_n to characterize high order properties. Laplacian regularization l_{lap} to handle local minimum stuck problem, and edge length regularization l_{loc} to solve long edge limitations. The overall loss is a weighted sum of all the four losses $l_{all} = l_c + \omega_1 l_n + \omega_2 l_{lap} + \omega_3 l_{loc}$, where $\omega_1=1.6e-4$, $\omega_2=0.3$ and $\omega_3=0.1$. A mesh with 2,466 vertices is an output of the network. The network structure implementation is shown in Table 1. Adam algorithm is used for optimization. During training stage, the model is divided into two parts. During the testing, the 3D results of various items are reconstructed within a few seconds.

3.2 Point Selection and Range Search

We allow customers to customize the logo’s location and size in a graphical user interface. The location of the logo is used in the range search module as the center point. A classical data structure K-D Tree [3] is used to query nearest neighbor points within customer-defined range. To accomplish a range search, we call two functions: (i) ‘IN_RANGE’ is to identify whether a point is within the range, and (ii) ‘BOUNDS_INTERSECT_RANGE’ is to distinguish bounds array. We visit as less points as possible to reduce the algorithm’s running time.

Table 1: Experimental Setup

Parameters	Attributes
Platform	Tensorflow
Optimize	Adam
Weight decay	1e-5
Mini-batch size	1
Epochs	50
Learning rate	1e-5 (first 30 epochs) 1e-6 (another 20 epochs)
Equipment	NVIDIA Titan Xp

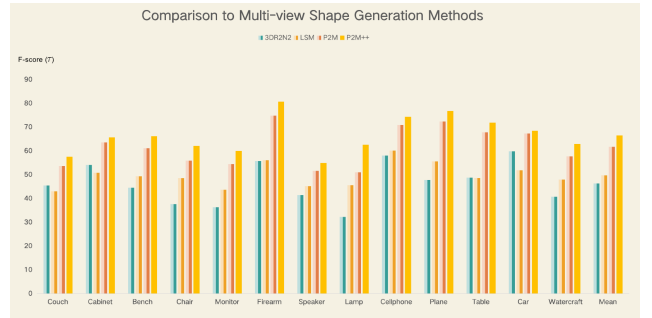


Figure 2: Comparison of multi-view shape generation methods.

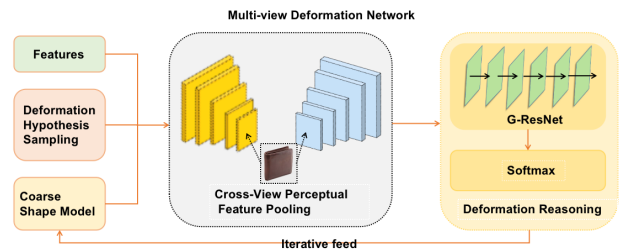


Figure 3: Multi-view deformation network structure.

3.3 Quadric Surface Fitting

The result of the range search module is used for quadric surface fitting, as well as removing outliers that are points within the specified range but out of the reconstruction. Quadric surface fitting is a low-degree algebraic surface fitting approach, with a fast average running time, which is used for commercial applications [4]. In this paper, we resample searched points to fit the curvatures of mesh. The equation of a quadric surface $f(x)=0$, which is equivalent to, $f(x)=C^T$, where $x = [x, y, z]^T \in \mathbb{R}^3$ is a 3D point, $C=[c_0, c_1, \dots, c_9]^T$ is the vector of unknown coefficients of the quadric surface. We approximate the Euclidean distance from a point to a quadric surface by $\frac{f(x)}{|\nabla f(x)|}$ and take the squared Euclidean distance into account $E_d(t, f) = \int_t \frac{f(x)^2}{|\nabla f(x)|^2} d\sigma$, where E_d measures the squared Euclidean distance L^2 , t is a triangle, and $f(x)$ is a quadric surface [4]. We use the classical gradient descent method to solve the non-linear least square optimization problem.

3.4 Mesh Deformation

We implement a mesh deformation module based on the effective and notably detail-preserving techniques, to deal with specific size of reconstructed customized logos [5]. The module consists of the following steps: (i) transforming the surface with the principle of local rigidity, (ii) estimating the corresponding points on the initial and the deformed surfaces, (iii) performing a rigid transformation to the original shape, and (iv) measuring the deviation to the deformed shape. We use an as-rigid-as-possible technique and implement a simple iterative approach to minimize the non-linear energy function. Denote a mesh by S , a deformed mesh by S' , and the set of vertices connected to vertex i by $N(i)$. In addition, we define that vertex positions $p_i \in \mathbb{R}^3$ is geometric embedding of S , $p'_i \in \mathbb{R}^3$ is geometric embedding of S' , R_i is the approximating rotation, ω_{ij} are weights, the cell C_i is corresponding to vertex i , and the C'_i is a deformed version. We then analyze the rigid transformations between two cells using the following equation:

$$E(C_i, C'_i) = \sum_{j \in N(i)} \omega_{ij} \left\| (p'_i - p'_j) - R_i (p_i - p_j) \right\|^2 \quad (1)$$

By summation of the deviations from rigidity per cell to measure the rigidity of a deformation of the whole mesh [5], the energy function is given:

$$E(S') = \sum_{i=1}^n \omega_i E(C_i, C'_i) = \sum_{i=1}^n \omega_i \sum_{j \in N(i)} \omega_{ij} \left\| (p'_i - p'_j) - R_i (p_i - p_j) \right\|^2 \quad (2)$$

where ω_i and ω_{ij} are fixed cell and edge weights, respectively. This as-rigid-as-possible method leads to an efficient solution of the mesh deformation with the predefined size.

4 EXPERIMENTS

In this section, we evaluate the proposed 3D logo mapping framework. We carry out experiments on 13 categories of the ShapeNet dataset provided by Choy *et al.* [6] (sofa, plane, chair, etc.), and on the set of real product photos provided by our industrial partner (wallet, clothes, etc.). With the AR headsets or mobile phones, the

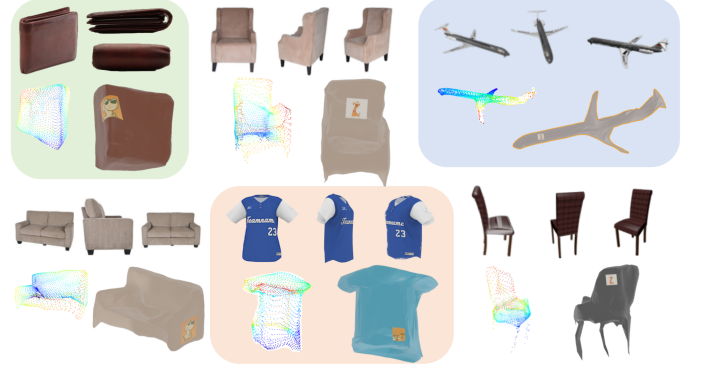


Figure 4: Experimental results.

customers can take photos (e.g., three photos) of their customized product, and our framework will reconstruct a 3D product model within a few seconds. In our graphical user interface, the customers can pick a point on the 3D product model as a customized logo central point, and set preferred size or pattern of their logo in real-time. The results are shown in Figure 4. The traditional asset creation pipelines may require a few hours of human labor and cannot be easily self-served by their customers. Compared to the previous works, our 3D logo mapping framework performs the state-of-the-art by producing 3D logos with good visual quality and accurate surface details in real-time. Although our framework has positive performance, there are some limitations. For example, the untrained product categories may produce a coarse 3D model, and the detailed large scale of points may impede logo generation speed. Overall, however, our method provides a pathway to automatically reconstruct a customized logo from a few color images and customer-preferred patterns.

5 CONCLUSION AND FUTURE WORK

This paper introduces a real-time 3D logo mapping framework for converting 3D logo mesh from a specified image and fitting it to the 3D product models. In the framework, we adopt a 3D mesh generation network to reconstruct 3D product, and a mesh deformation algorithm to reconstruct customer-defined logo mesh. It achieves significantly effective logo customization on both benchmark and physical industrial products. The framework (implemented via a graphical user interface) can enhance the customer experience in the retail industry. Our work can be used both in the retail industry and other fields with MR. It makes products better tailored to match consumers' needs. In the future, we will continue to find a balance between visual quality and time consumption for the framework.

ACKNOWLEDGMENTS

The authors would like to thank the editor and the anonymous reviewers for their valuable comments. This work has been supported by OrangeLV, Inc.

REFERENCES

- [1] Furukawa, Y., & Ponce, J. 2009. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 32(8), 1362-1376.

- [2] Wen, C., Zhang, Y., Li, Z., & Fu, Y. 2019. Pixel2mesh++: Multi-view 3d mesh generation via deformation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1042-1051).
- [3] Bentley, J. L. 1975. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9), 509-517.
- [4] Yan, D. M., Liu, Y., & Wang, W. 2006. Quadric surface extraction by variational shape approximation. In International conference on geometric modeling and processing (pp. 73-86). Springer, Berlin, Heidelberg.
- [5] Sorkine, O., & Alexa, M. 2007. As-rigid-as-possible surface modeling. In Symposium on Geometry processing (Vol. 4, pp. 109-116).
- [6] Choy, C. B., Xu, D., Gwak, J., Chen, K., & Savarese, S. 2016. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. In European conference on computer vision (pp. 628-644). Springer, Cham.
- [7] Smink, A. R., Frowijn, S., van Reijmersdal. 2019. Try online before you buy: How does shopping with augmented reality affect brand responses and personal data disclosure. *Electronic Commerce Research and Applications*, 35, 100854.
- [8] Tulsiani, S., Kar, A., Carreira, J., & Malik, J. 2016. Learning category-specific deformable 3d models for object reconstruction. *IEEE transactions on pattern analysis and machine intelligence*, 39(4), 719-731.
- [9] Tulsiani, S., Efros, A. A., & Malik, J. 2018. Multi-view consistency as supervisory signal for learning shape and pose prediction. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2897-2905).
- [10] Gwak, J., Choy, C. B., Garg, A., Chandraker, M., & Savarese, S. 2017. Weakly supervised generative adversarial networks for 3d reconstruction. arXiv preprint arXiv:1705.10904, 2.
- [11] Pumarola, A., Agudo, A., Porzi, L., Sanfeliu, A. 2018. Geometry-aware network for non-rigid shape prediction from a single view. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4681-4690).
- [12] Wang, N., Zhang, Y., Li, Z., Fu. 2018. Pixel2mesh: Generating 3d mesh models from single rgb images. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 52-67).
- [13] Fan, H., Su, H., & Guibas, L. J. 2017. A point set generation network for 3d object reconstruction from a single image. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 605-613).
- [14] Kar, A., Häne, C., & Malik, J. 2017. Learning a multi-view stereo machine. arXiv preprint arXiv:1708.05375.
- [15] Kobbelt, L., Campagna, S., Vorsatz, J., & Seidel, H. P. 1998. Interactive multi-resolution modeling on arbitrary meshes. In Proceedings of the 25th annual conference on Computer graphics and interactive techniques (pp. 105-114).
- [16] Botsch, M., & Kobbelt, L. 2004. An intuitive framework for real-time freeform modeling. *ACM Transactions on Graphics (TOG)*, 23(3), 630-634.