Title:
**A Method for the Automatic Generation of 3D models based on Artificial Intelligence**

Authors:
Laura Loredana Micoli, laura.micoli@polimi.it, Politecnico di Milano, Italy
Gabriele Guidi, gabguidi@iu.edu, Indiana University, Bloomington, USA
Giandomenico Caruso, giandomenico.caruso@polimi.it, Politecnico di Milano, Italy

Keywords:
Artificial Intelligence, 3D modeling, Automatic Modeling, Process Optimization, 3D Object Database

DOI: 10.14733/cadconfP.2023.149-153

Introduction:
Virtual Reality (VR) is a widespread technology to develop an interaction between humans and simulated environments for multiple purposes, ranging from gaming to advanced industrial applications. The need for a real-time response of VR systems imposes strict constraints on the virtual content, which is usually manually optimized. Building VR scenarios with thousands of 3D models makes this aspect crucial. However, the manual modeling of numerous items is time-consuming and includes significant issues, especially if the 3D models must replicate real objects in photorealistic quality. In this context, a indicative case study relates to the virtual simulation of entire commercial retail areas and their consumer goods to support marketing and management design. It is necessary that the 3D models replicating consumer goods have the same level of detail and fidelity as their physical counterparts. Such visually sophisticated models involve not only a certain complexity but also need to be frequently updated according to marketing needs. These needs make reducing the development time for a single virtual replica even more important [5].

This work proposes a method to automate the modeling process of a large commercial area (e.g., a grocery store) involving a massive set of products whose shapes can be classified according to predefined geometrical categories (archetypes) [7]. In particular, the dataset for this study relates to "Personal Care" products (e.g., bottles, tubes, boxes, jars, etc.), which present a wide variety of shapes attributable to just a few archetypes. The method uses orthographic images of an object to be evaluated by Convolutional Neural Networks (CNN)s to assign it to a specific archetypal category [13]. Based on this assignment, Computer-Vision (CV) algorithms analyze the same images to extrapolate the characteristic dimensions related to the assigned archetypes. Once the category and the parameters are defined, a virtual replica is automatically generated through a scripted procedure running on 3D modeling software. The experimental results obtained with the method demonstrate its effectiveness for recognizing the appropriate archetype, extracting relevant parameters, and creating a 3D model comparable with those generated with other high-fidelity processes but much lighter in terms of model size.

Proposed method:
Developing VR environments populated by thousands of items poses some recurring critical obstacles: i) optimization of the number of polygons (high-fidelity shape vs. real-time rendering); ii) production time of the single 3D model. If these 3D models vary in their formal macro-characteristics, the hypothesis is to use an automatic procedural modeling approach based on a codified and semi-automatic process [4], [6], [10-12], [14], [15]. The method was implemented through an in-depth preliminary and iterative study to define each step of the process.

Proceedings of CAD'23, Mexico City, Mexico, July 10-12, 2023, 149-153

The proposed method develops the following steps:

- definition of the catalog of archetypes;
- identification of the archetype category through CNNs;
- measurement of the identified object with CV algorithms;
- automatic 3D modeling according to the archetype category.

*Definition of the catalog of archetypes*

A morphological macro-analysis was carried out on about 1600 objects to codify the common geometrical features of heterogeneous products with archetypes that were then collected in a catalog. For this research, products belonging to the "Personal care" category were selected because they present great geometrical variety, useful for studying a generalized approach to implement an automatic 3D modeling process. The packaging of these products can be extremely simple and devoid of details, such as boxes, or very complex, like the ones defined by free-form geometries, minute details, and asymmetrical elements. It emerges that many items share evident morphological characteristics, even between different brands and product subcategories; they are diverse in size, shape, and texture but attributable to formal families that require the same 3D modeling steps [2], [3], [9]. The morphological analysis on the 1600 items taken into consideration highlighted that most of them (86%) could be modeled with double symmetry surfaces with respect to the principal vertical planes (i.e., the two planes containing the object's vertical axis, orthogonal to the horizontal plane, and identifying the fontal and the lateral views), while 14% had none, being characterized by sculptural surfaces. According to this analysis, three macro-families have been identified: box, linear bottle, and curvilinear bottle.

They have been subsequently codified within a catalog used as a reference to implement dedicated CNNs for shape recognition of new products. The catalog is designed on the geometric features of different shape families: i) the type of fundamental curves/points on which the modeling will be based and ii) how these geometries are connected to each other. Moreover, each family will have subtypes grouped in a hierarchical subdivision based on different properties, for example, the number of curves along the vertical axis and how it is shaped in the apical part of the object.

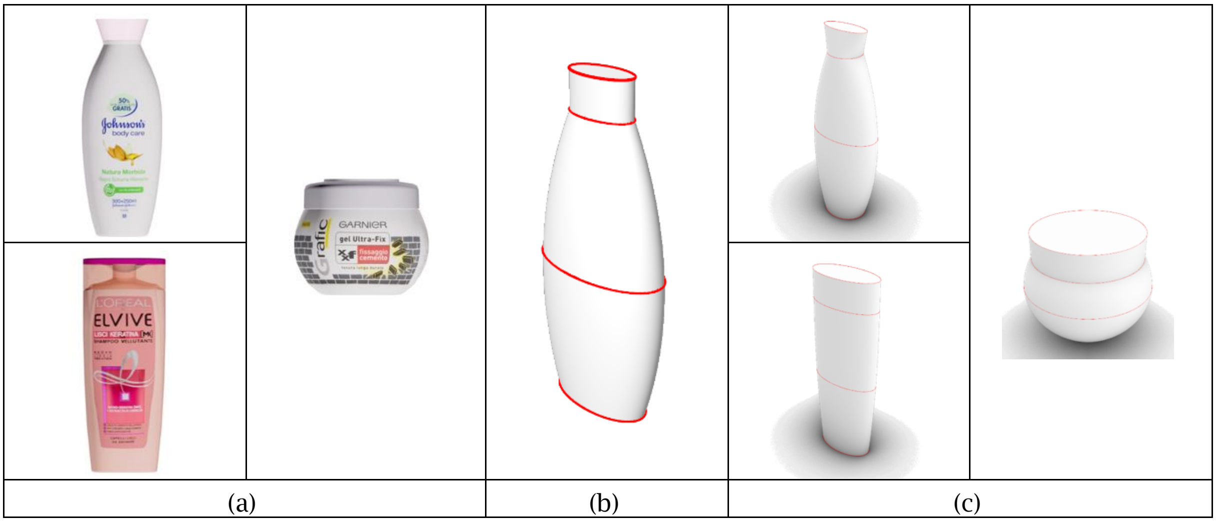



Fig. 1: Example of the approach: (a) pictures of three different items, (b) reference archetype, (c) digital replica of the products obtained based on the identified archetype and the main dimensions of the products.

A unique identification code has been defined for each archetype, consisting of 6 digits comprised by three pairs that identify the salient formal features; the catalog is scalable and can be widened by including further families and archetypes. Thanks to this approach, starting from the same archetype, it is possible to generalize the digital replica of apparently very different products that share the same formal scheme (Fig. 1).

*Identification of the archetype category through CNNs*
The CNNs used for the automatic identification of the archetype have been implemented with MobileNet [8], which allows the implementation of efficient CNNs widely used for image recognition. The results obtained with MobileNet were satisfactory even if compared with sophisticated CNNs like VGG16 [13], which also required more time for training. Our CNNs were trained with heterogeneous datasets, one for each archetype category, including real and synthetic orthographic images, i.e., generated by parametric 3D models. This training approach was useful, as shown in the work of Su et al. [16], where a similar task proves that synthetic images in place of real images (as usually used in such tests) required a greater volume. Thanks to the reduced size of MobileNet and the development of different CNNs, one for each archetype category, we could train each CNN without requiring wide datasets. By analyzing the object's orthographic front and side views, the first CNN verifies the required double symmetry of the object. If the object is symmetric, the second CNN analyzes the top and bottom views to discriminate the main archetype category. If necessary and according to the specific category, a third CNN analyzes the side view again to identify the second-level category.

*Measurement of the identified object with CV algorithms*
After identifying the archetype category, dedicated CV algorithms process the same orthographic images of the object to retrieve information for the subsequent automatic modeling. These algorithms have been developed with OpenCV (https://opencv.org/), a cross-platform library for real-time CV applications. A CV algorithm extracts the silhouette curve for objects with axial symmetry. For the other archetype categories, dedicated algorithms are used to identify the second-level category, if necessary, and to retrieve dimensions of the reference geometries of the object. To extract the dimensions, the algorithm makes a derivative analysis of the silhouette.

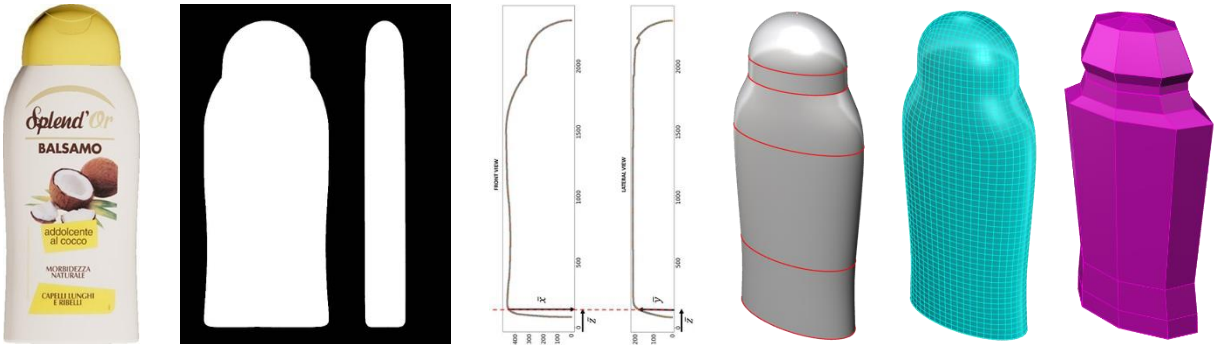
*Automatic 3D modeling according to the archetype category*
The geometric modeling for VR applications is normally carried out with polygonal modeling techniques. Despite this, according to the proposed methodology, it was decided to reproduce the objects starting with modeling using curves and surfaces and tessellate them afterward with the proper polygonal density. In this way, with a limited number of parameters, it is possible to obtain better control of the shape. For this reason, it was necessary to identify a software application, among those available on the market, capable of managing automatized operations to generate suitable geometries for VR applications. The identified software for testing the method was Rhinoceros, (https://www.rhino3d.com/), which allows: i) modeling curves and surfaces and the subsequent Sub-D conversion and check; ii) automated data input and modeling through a scripting language [1], [6], [4], [10-12].

The code used for scripting was Python (https://www.python.org/), a cross-platform language used to execute a series of commands as a script or to make links between different applications, also used in this study for image-based shape analysis. The processing pipeline to be carried out to realize each archetype was outlined using the chosen modeling software and scripting language, with the aim of identifying a sequence of actions suitable for optimizing the quality of the result and prioritizing computational efficiency. This phase considered the subsequent conversion of the operations into scripts aimed at generalizing the parameters and the repeatability of the approach.

The procedure for most of the archetypes, except for boxes and objects with rotational symmetry, was developed with the following steps: a) definition of dimensional parameters; b) drawing of reference curves (ellipses, lines, arcs, etc.) so that they all have a coherent and functional orientation for the construction of the surfaces; c) modeling of vertically developed surfaces based on the drawn curves (extrusion, striped loft, normal loft, uniform loft, sweep2rail); d) base modeling (planar surface); e) lid modeling (planar, ellipsoid, lofted end point surface); f) Surface normal check. Once the surfaces are completed, the model is converted to polygonal geometry as QUAD and SubD surfaces.

QUAD surfaces are optimized for: i) preserving the shape of curved surfaces, the representation of edges, and the correct transition between surfaces; ii) obtaining a visually satisfactory smoothed model; iii) limiting the number of polygons. This solution is ideal for items that do not need more details and are ready to be texturized. SubD surfaces are extracted to obtain a polygonal model composed of a very limited number of faces coherent with the subdivision of the previously defined surface patches. This solution provides a polygonal reference mesh, which is optimal for adding shape details using the typical Box Modeling technique (i.e., sculpting the shape from an elementary 3D primitive).

Once the optimal modeling path of the archetypes had been defined, this was translated into a scripting language using specific functions implemented in the modeling software. To generate the 3D model of a product, it is necessary to know the archetype to be used and based on this, to have the related dimensional parameters, which typically are taken in correspondence of the peaks of the surface curvature gradients. The execution of a single script requires about 5-10 seconds, ensuring high efficiency compared to a manual modeling approach. Fig. 2 presents the representative images for the main steps of the proposed method.



(a) (b) (c) (d) (e) (f)

Fig. 2: Main steps of the proposed method: (a) real orthographic image of the product; (b) orthographic images of the product processed with CNN and CV; (c) derivative analysis of the silhouette; (d) 3D surface model obtained from the proposed modeling; (e) model conversion to Quad Mesh (f); model conversion to SubD surfaces.

Conclusion:

This work presents a new method to automate the creation of 3D models of a massive set of products sharing the same shape structure, identified by a catalog of archetypes, and extracting the main data from CNN analysis of orthogonal views of the real object. The method, conceived for the shapes typical of the "Personal Care" product packages, is enough general to be applied to any shape with symmetry with respect to a central axis or a couple of orthogonal planes. Such symmetry, found in 86% of the 1600 analyzed product category, is typical of a large percentage of many other product packages and can be used to quickly create 3D models of large product sets to populate virtual scenes. We have shown that the use of CNNs and CV can enable a quick modeling of accurate and optimized 3D models that can be applied in different fields. The comparison made with our gold standards reveals an accurate reconstruction of the object and shows promising perspectives for future works. Such works could apply this method starting only from orthophotos or developing a pipeline that can work directly with the 3D scan of an object.

Acknowledgements:

The authors sincerely thank INVRSION S.r.l. (https://www.invrsion.com/), which supported the research by providing data and highlighting the real issues for the reproduction of large scenarios, and Matteo Agostini, for his valuable contribution.

References:

[1] Aliaga, D.G.; Rosen, P.A.; Bekins, D.R.: Style Grammars for Interactive Visualization of Architecture, IEEE Transactions on Visualization and Computer Graphics, 13 (4), 2007, 786-797. https://doi.org/10.1109/TVCG.2007.1024

[2] De Luca, L.; Véron, P.; Florenzano, M.: A generic formalism for the semantic modeling and representation of architectural elements, The Visual Computer journal, 23, 2007, 181-205. https://doi.org/10.1007/s00371-006-0092-5

[3] Demir, İ.; Aliaga D.G.: Guided proceduralization: Optimizing geometry processing and grammar extraction for architectural models, Computers & Graphics, 74, 2018, 257–267. https://doi.org/10.1016/j.cag.2018.05.013

[4] Ebert, D.S.; Musgrave, K.; Peachey, D.; Perlin, K.; Worley, S.: Texturing and Modeling: a Procedural Approach., 3nd edn. Morgan Kaufman Publishers, San Francisco, 2003. https://doi.org/10.1016/B978-155860848-1/50029-2

[5] Guidi, G.; Micoli, L.L.; Casagrande, C.; Ghezzi, L.: Virtual reality for retail, 16th International Conference on Virtual Systems and Multimedia, Seoul, Korea (South), 2010, 285-288. https://doi.org/10.1109/VSMM.2010.5665949

[6] Guidi, G.; Frischer, B.; Lucenti, I.: Rome Reborn - Virtualizing the Ancient Imperial Rome, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. ISPRS, Zurich, Switzerland, 2007, 1–4 https://www.isprs.org/proceedings/XXXVI/5-W47/pdf/guidi_etal.pdf

[7] Guidi, G.; Micoli L.L.: A semi-automatic modeling system for quick generation of large virtual reality models. In: Proceedings of the ASME 2011 World Conference on Innovative Virtual Reality. ASME, Milan, Italy, 2011, 1–8. https://doi.org/10.1115/WINVR2011-5512

[8] Howard, A. G.; Zhu, M.; Chen, B., Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017. https://doi.org/10.48550/arXiv.1704.04861

[9] Martin, I.; Patow, G.: Ruleset-rewriting for procedural modeling of buildings, Computers & Graphics, 84, 2019, 93–102. https://doi.org/10.1016/j.cag.2019.08.003

[10] Mas, A.; Martin, I.; Patow, G.: Simulating the Evolution of Ancient Fortified Cities, Computer Graphics Forum 39(1), 2020, 650–671. https://doi.org/10.1111/cgf.13897

[11] Müller, P.; Wonka, P.; Haegler, S.; Ulmer, A.; Van Gool, L.: Procedural modeling of buildings, ACM Trans. Graph., 25(3), 2006, 614–623. https://doi.org/10.1145/1179352.1141931

[12] Parish, Y.I.H.; Müller, P.: Procedural modeling of cities. Proceedings of the 28th annual conference on Computer graphics and interactive techniques - SIGGRAPH '01. ACM Press, New York, New York, USA, 2001, 301–308. https://doi.org/10.1145/383259.383292

[13] Simonyan, K.; Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014. https://doi.org/10.48550/arXiv.1409.1556

[14] Stiny, G.; Gips, J.: Shape grammars and the generative specification of painting and sculpture, Inf Process 71 Proc IFIP Congress, 2(3), 1972, 1460-1465.

[15] Stiny, G.; Mitchell, W.J.: The Palladian grammar, Environ Plan B Plan Des, Environment and Planning B: Planning and Design, 5(1), 1978, 5–18. https://doi.org/10.1068/b050005

[16] Su H, Qi CR, Li Y, et al. Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model Views, 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, 2686-2694, https://doi.org/10.1109/ICCV.2015.308