Model Retrieval by 3D Sketching in Immersive Virtual Reality

Daniele Giunchi* University College London
Stuart James† University College London
Anthony Steed‡ University College London

ABSTRACT

We describe a novel method for searching 3D model collections using free-form sketches within a virtual environment as queries. As opposed to traditional Sketch Retrieval, our queries are drawn directly onto an example model. Using immersive virtual reality the user can express their query through a sketch that demonstrates the desired structure, color and texture. Unlike previous sketch-based retrieval methods, users remain immersed within the environment without relying on textual queries or 2D projections which can disconnect the user from the environment. We show how a convolutional neural network (CNN) can create multi-view representations of colored 3D sketches. Using such a descriptor representation, our system is able to rapidly retrieve models and in this way, we provide the user with an interactive method of navigating large object datasets. Through a preliminary user study we demonstrate that by using our VR 3D model retrieval system, users can perform quick and intuitive search. Using our system users can rapidly populate a virtual environment with specific models from a very large database, and thus the technique has the potential to be broadly applicable in immersive editing systems.

1 INTRODUCTION

In recent years, with the rapid growth of interest in 3D modeling, repositories of 3D objects have ballooned in size. While many models may be labeled with a few keywords and/or fields that describe their appearance and structure, these are insufficient to convey the complexity of certain designs. Furthermore, in many existing databases, these keywords and fields are incomplete. Thus query-by-example methods have become a very active area of research. In query-by-example systems, the user typically sketches elements of the object or scene they wish to retrieve. A search system then retrieves matching elements from a database. In our system, the user is immersed in a virtual reality display. We provide a base example of the class of object to act as a reference for the user. The user can then make free-form colored sketches on and around this base model. A neural net system can analyze this sketch and retrieve a set of matching models from a database. The user can then iterate by making further correctional sketches until they find an object that closely matches their intended model. This leverages the strengths of traditional approaches while embracing new interaction modalities uniquely available within a 3D virtual environment. The main challenge in sketch-based retrieval is that annotations in the form of sketches are an approximation of the real object and may suffer from being a subjective representation and over-simplifications. For image retrieval, methods focus on enhancing lines through gradients, GF-HOG [4] and Tensor Structure [3], with more recent approaches based on convolutional neural nets (CNNs) [1, 7]. To match 3D models, it is typical to normalize the models to have the same orientation, so that a standard set of images at set orientation can be rendered to compare the sketch to. We adopt this view-based method as it allows an interactive experience where users can get responses with little delay. So far, sketching within a virtual environment as a retrieval method has received little attention. There are various tools to allow the user to sketch (e.g. Tiltbrush, or Quill), but these focus on the sketch itself as the end result. Other systems allow free-form manipulation of objects by simple affine manipulation through drag points [5]. In contrast, we instead are interested in how a user can utilize sketch as a method of retrieval. We therefore performed a user study to compare sketch-based retrieval to a naive linear browsing to demonstrate that sketching is an effective and usable method of exploring model databases. We present a novel approach to searching model collections based on annotations on an example model. This example model represents the current best match within the dataset and sketching on this model is used to retrieve a better match. Our system is the first of its type to work online in an immersive virtual environment. This model retrieval technique can be broadly applied in editing scenarios, and it enables the editor to remain immersed within the virtual environment during their editing session.

*e-mail: d.giunchi@ucl.ac.uk
†e-mail: stuart.james@ucl.ac.uk
‡e-mail: a.steed@cs.ucl.ac.uk

2 3D SKETCH-BASED RETRIEVAL

Searching for a model in a large collection using 2D sketches can be tedious and require an extended period of time. It also requires a particular set of skills, such as understanding perspective and occlusion. By using virtual reality this experience can be improved because ambiguity between views is greatly reduced and the user no longer has to hallucinate the projections from 2D to 3D. A novel aspect of our method is that we allow users to make sketches directly on top of existing models (see in Fig. 2). The users can express color, textures and the shape of the desired object.

2.1 Pre-Processing

Drawing from current state-of-the-art model descriptions approaches we apply a multi-view convolutional neural network architecture to describe the content of the model. We apply the Multi-View CNN model proposed by Su et al. [6], which we briefly outline below and can be seen in Fig. 1.

To generate a single vector description of a model the chair is projected into 12 distinct views as show in Fig. 1 (a). Each view is then described by an independent CNN model. The standard VGG-M network is applied from [2], consisting of five convolutional layers and three fully connected layers. We apply this network to our dataset, generating the 12 images through a virtual camera that takes snapshots from different point of views.

2.2 Online Queries

At query time, 12 images are generated from the users' sketches and, optionally the current 3D model that is the best match, and a forward pass through the network returns the descriptor. After comparing the descriptor with the descriptor collection the system replies with the K-nearest models that fit the input sent. We provide the user two ways to perform the query: sketch-only query or both sketch and model query. This is achieved by enabling or disabling the visualization of the model (see in Fig. 1). After the system proposes results, if the target model is not present the user can edit the sketch or conversely can replace the current model with a new one that better matches represents the desired target. This facilitates a step by step refinement to navigate through the visual feature space of

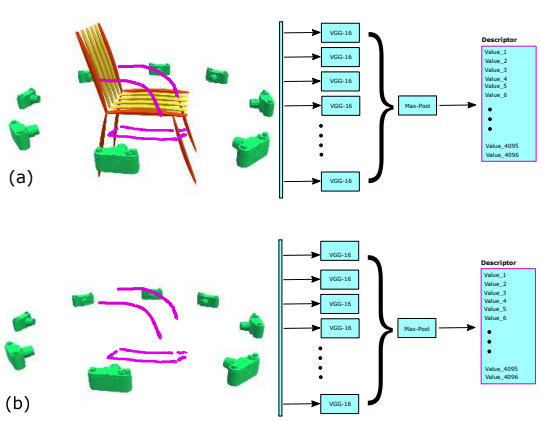

Figure 1: (a) CNN can be triggered with snapshots with both sketch and chair model. (b): CNN can be triggered with snapshots with only sketch present.

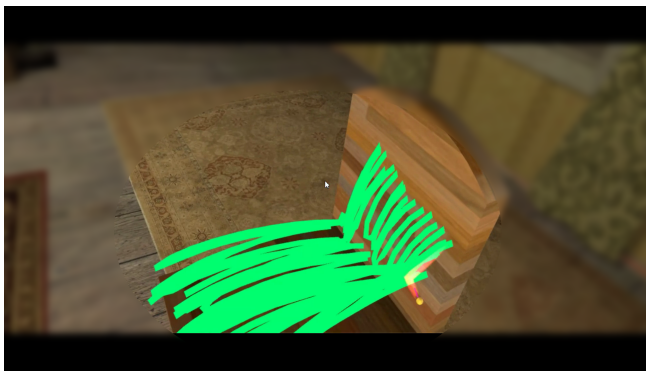

Figure 2: An example of a users sketch within the sketch interface.

the collection, commonly achieving the target model within a few iterations as shown in Fig. 3.

3 Discussion

In tests with users we found that it is possible, through an iterative process of sketching and model selection, to perform an effective search for a model in a large database while immersed in a virtual environment. Two different strategies emerged from the experiment. The first and more intuitive approach is to make a single sketch and detail it step by step until most features of the chairs are resolved without replacing the model. The user can interrogate the system to have a feedback but essentially will continue to sketch. The downside is that the user can waste time on detailing a sketch and, in addition, can depict features that are not so relevant. Determining whether features are relevant or not is not a trivial task for two reasons. The first one is that different users will over-rate the saliency of the feature. The second one is the possibility that the specific feature is common to many database objects. In both cases the participant experiences an unsatisfactory answer from the system as it proposes a chair set without that feature or conversely many chairs containing it. The second approach is to only model differences to the current object: that is the user queries the system and then only adds features that are different in the target object. The advantage of this method is that the quick system's response (~2 seconds) enables fast iterative refinement. When the system receives a different combination of sketch and model it will retrieve a different set of chairs.

Figure 3: Examples of users that successfully triggered the system using a combination of sketches and model. The left column contains the target chairs, while the other columns contain a subset of the snapshots used by the system.

4 Conclusion

The benefits of the virtual reality in the field of scene modeling have been investigated for several years. Previous research has focused on free-form modeling rather than developing a way to retrieve models from a large database. Current strategies for navigating an existing dataset use queries on tags or show to the user the entire set of models. In addition, large collections can suffer from a lack of meta-information which hampers model search and thus excludes part of the dataset from query results. We proposed a novel interaction paradigm that helps users to select a target item using an iterative sketch-based mechanism. We improve this interaction with the possibility of combining sketches and a background model together to form a query to search for a target model. An experiment collected information about the time taken to complete the task and user experience rating. We believe that sketch-based queries are a very promising complement to existing immersive sketching systems.

References

[1] T. Bui, L. Ribeiro, M. Ponti, and J. Collomosse. Compact descriptors for sketch-based image retrieval using a triplet loss convolutional neural network. *Computer Vision and Image Understanding*, 2017.

[2] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In *British Machine Vision Conference*, 2014.

[3] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa. Sketch-based image retrieval: Benchmark and bag-of-features descriptors. *IEEE Transactions on Visualization and Computer Graphics*, 17(11):1624–1636, 2011.

[4] R. Hu and J. Collomosse. A performance evaluation of gradient field hog descriptor for sketch based image retrieval. *Comput. Vis. Image Underst.*, 117(7):790–806, July 2013.

[5] T. Santos, A. Ferreira, F. Dias, and M. J. Fonseca. Using sketches and retrieval to create lego models. In *Proceedings of the Fifth Eurographics Conference on Sketch-Based Interfaces and Modeling*, SBM'08, pp. 89–96, 2008.

[6] H. Su, S. Maji, E. Kalogerakis, and E. G. Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proc. ICCV*, 2015.

[7] Q. Yu, F. Liu, Y. Z. Song, T. Xiang, T. M. Hospedales, and C. C. Loy. Sketch me that shoe. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 799–807, June 2016.