Model Retrieval by 3D Sketching in Immersive Virtual Reality

Daniele Giunchi & Stuart James & Anthony Steed
University College London

Introduction

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, colour 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 coloured 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.

System Description

In our system (showed in Figure 1), the user is immersed in a virtual reality display. We provide an example model of the class to act as a reference for the user. The user can then make free-form coloured sketches on and around this base model. A neural net system analyzes the sketch and retrieves a set of matching models from a database. The user is then able to iterate, making further correctional sketches (e.g. adding new pieces to the model or refinements) until they find their target object.

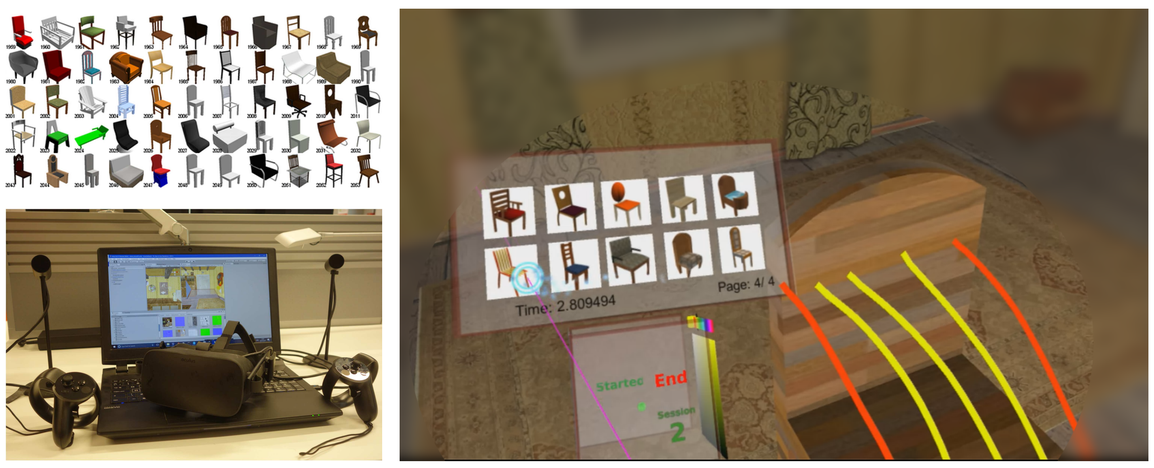

Figure 1: Top left: a collage of the *chairs* class models. Bottom left: laptop, Oculus Rift and Touch. On the right the interface implemented for sketching

CNN

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. [2], which we briefly outline in the Figure 2.

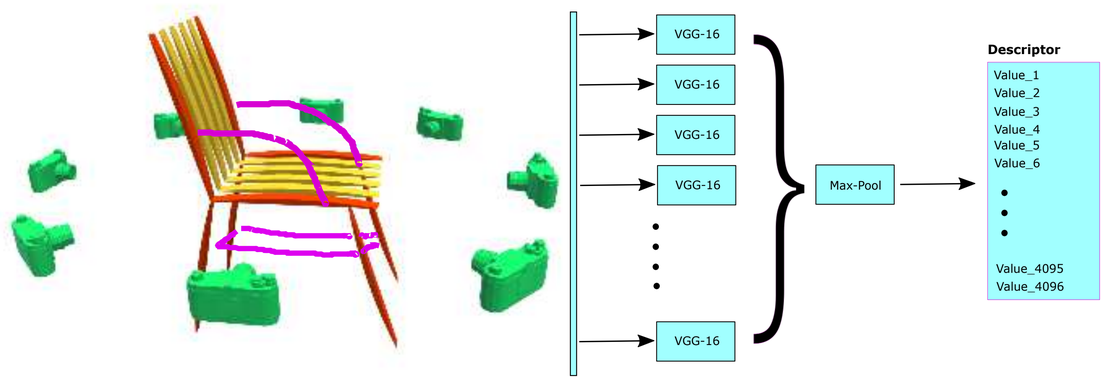

Figure 2: CNN can be triggered with snapshots with both sketch and chair model.

To generate a single descriptor of a model the chair is projected into 12 distinct views. Each view is then described by an independent CNN model. The standard VGG-M network is applied from [1], consisting of five convolutional layers and three fully connected layers. As in [2] the model is trained on ImageNet then fine-tuned on the 2D images of the 3D Shapes dataset. Each view of the model is applied to the convolutional layers of the VGG-M where the resulting descriptors are aggregated by element-wise max pooling. The VGG-M network is shared amongst views.

Evaluation

We designed an experiment to compare two methods: the proposed sketch-based method, and a naive scrolling panel method. For each session of the test, we first showed the participant the twelve views of a target chair as generated for the descriptor. We then asked the participant to retrieve the chair from the database (showed in Figure 3), using one of the two methods. We tracked the success rate, the time to complete the task and a subjective evaluation of the user experience through a questionnaire. The number of successful task completions for the scroll method was 119 out of 180 (66%) and for the sketch method 171 out of 180 (95%).

Discussion

Two different user strategies emerged from the experiment: sketch only and sketch with a model. 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 second approach is to 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 sketch is usually started again after each query. The advantage of this method is that the quick response from the system (2 seconds) enables fast iterative refinement. This method requires more experience from the user, but after few iterations we observed several participants starting to adopt it.

Figure 3: Our interface lets the user draw coloured sketches in a virtual environment and select iteratively one model from the retrieved set.

Conclusions

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 (as showed in Figure 4) to search for a target model. An experiment collected information about the time taken to complete the task and user experience rating. We thus believe that sketch-based queries are a very promising complement to existing immersive sketching systems.

Figure 4: Leftmost column: target chairs. The other columns show the visual queries.

References

[1] 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.

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