Explorative Study on Asymmetric Sketch Interactions for Object Retrieval in Virtual Reality

Daniele Giunchi
University College London
London, United Kingdom
d.giunchi@ucl.ac.uk

Stuart James
Istituto Italiano di Tecnologia
Genoa, Italy
stuart.james@iit.it

Riccardo Bovo
Imperial College London
London, United Kingdom
r.bovo19@imperial.ac.uk

Donald Degraen
German Research Center for Artificial Intelligence (DFKI))
Saarbrücken, Germany
donald.degraen@intel-vci.uni-saarland.de

Anthony Steed
University College London
London, United Kingdom
a.steed@cs.ucl.ac.uk

2022 International Conference on Interactive Media, Smart Systems and Emerging Technologies (IMET) | 978-1-6654-7016-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/IMET54801.2022.9929445

***Abstract*—Drawing tools for Virtual Reality (VR) enable users to model 3D designs from within the virtual environment itself. These tools employ sketching and sculpting techniques known from desktop-based interfaces and apply them to hand-based controller interaction. While these techniques allow for mid-air sketching of basic shapes, it remains difficult for users to create detailed and comprehensive 3D models. Our work focuses on supporting the user in designing the virtual environment around them by enhancing sketch-based interfaces with a supporting system for interactive model retrieval. An immersed user can query a database containing detailed 3D models and replace them with the virtual environment through sketching. To understand supportive sketching within a virtual environment, we made an explorative comparison between asymmetric methods of sketch interaction, i.e., 3D mid-air sketching, 2D sketching on a virtual tablet, 2D sketching on a fixed virtual whiteboard, and 2D sketching on a real tablet. Our work shows that different patterns emerge when users interact with 3D sketches rather than 2D sketches to compensate for different results from the retrieval system. In particular, the user adopts strategies when drawing on canvas of different sizes or using a physical device instead of a virtual canvas. While we pose our work as a retrieval problem for 3D models of chairs, our results can be extrapolated to other sketching tasks for virtual environments.**

***Index Terms*—Sketch, Virtual Reality, CNN, HCI**

I. Introduction

Sketching is a very intuitive form of communication that has been successfully exploited in the digital world, firstly on 2D displays([1]–[3]) and then in 3D within Virtual Reality(VR)([4], [5]). It gives complete freedom to depict shapes and objects using different colours, and in VR, it can access the additional third dimension expanding the usual 2D canvas. Therefore, while sketch benefits from VR characteristics, and vice versa VR can take advantage of the spontaneous and familiar form of interaction that sketch represents. To aid users in the creation of complex 3D designs, existing methods are often supported by a retrieval algorithm capable of finding complex designs based on a simple sketch made by the user by searching a model database. Common approaches can be divided into methods focusing on gestural interaction [6], [7] or techniques allowing to freely draw sketches in either 2D [8] or 3D space [5]. Gestural interaction techniques are widely used to execute an action as a trigger mechanism or depict a simple trajectory in the design space. While gestures are generally easy to use, they are usually not suitable for characterizing detailed features of an object. Gestures additionally limit the ability of the user to express their desires freely. This is especially so for the task of retrieval where flexibility is key to finding the relevant content, i.e., the so-called *needle in a hay stack* problem. However, both 2D and 3D sketches allow the user to convey complex structures, including their details. These techniques extend the scope of potential designs to many objects within a collection with significant variations in terms of shape, colour, and texture. Despite the growing interest in methods for Sketch-based Retrieval [3], [9]–[12], only few examples of such systems for VR have been proposed [5], [13]. The reason for this is twofold: (1) the ability to sketch descriptions of an object to express the model adequately, and (2) the ability for users to get confidence in depicting structural elements. Prior work by Giunchi *et al*. [5] utilized a Multi-View CNN to solve for the first and introduced model sketching in VR to solve for the second. While Giunchi *et al*. [5] showed how a combination of sketch and model improves the user's retrieval query, they only explore a single modality of interaction, i.e., 3D mid-air sketching, and its level of effectiveness. However, it is essential to understand how different interaction methodologies can impact user performance and experience. In addition, it is crucial to separate the model component from the sketch component because the first one tends to have supremacy over the second one when combined. Therefore, we present a study to understand how users interact with physical and virtual devices framed in a retrieval context. Our work investigates different techniques for users to provide initial sketch designs as input for sketch-based retrieval algorithms in virtual environments. Therefore our contributions are as follows:

- Four methods of Sketch-based Retrieval interaction in VR:

- 3D Mid-Air Sketching, based on the method of Giunchi *et al*. [5] using mid-air drawing using a controller;
  - 2D Sketching on a VR Tablet, using a 2D tablet within the virtual environment;
  - 2D Sketching on a VR Whiteboard, using a VR plane to annotate the model;
  - 2D Sketching on a physical tablet, using a real-world tablet tracked in VR to annotate the model with strokes.
- A pilot study over the four methods identifying the advantages and disadvantages of methods with regards to the user, and possible emerging strategies.

II. Related Work

Using *Sketches* has a long history for image retrieval [1]–[3] interaction in 2D [8], or more recently for interaction in 3D [5]. We focus on core techniques in 2D sketching (sec. II-A) and 3D sketching in AR/VR (sec. II-B).

A. 2D Sketching for Retrieval

Sketch-based retrieval techniques using 2D sketches query rely on a predefined database to obtain either a resulting image (Sketch-Based-Image-Retrieval, SBIR) or video (Sketch-Based-Video-Retrieval ,SBVR). SBIR approaches can be classified into blob-based techniques that try to extract features for shape, colour or texture for the blob or contour-based techniques that characterize the image with curves and lines. Blob-based SBIR methods try to describe images using descriptions of the segments within the image, for example, Query-By-Image-Content (QBIC) [14], which creates separate descriptors for the modalities. Alternatively, topology models [15] can describe the blob characteristics. Contour-based methods include elastic matching [16] and grid and interest points approaches [17]. More recently, neural networks and, in particular, triplet convolutional neural networks have gained interest as they have the capacity to deal with deep embedding spaces [9]. All the approaches above focus on 'Drawing on a whiteboard' with a generally singular or interactive search. While improvements in the field of model retrieval increase accuracy, supporting the user during sketching is imperative to increase performance. One example is ShadowDraw [18], where the user is provided with feedback related to texture, colour, and shapes to improve the image retrieval task. The system provides the user with a real-time shadow image to help free-form drawing and achieve a better final sketch of the object during a sketching activity. Sketch-to-Collage [19] uses a query-by-sketch to generate collages from an image collection and deploys an indexing mechanism based on colours and shapes. With the rapid growth of 3D model collections ([20]–[23]), the retrieval of a 3D model from a collection has received attention. Two classes of descriptors were introduced: model-based descriptors and view based descriptors. For model-based there consists three types of descriptors: geometric moment [24], surface distribution [25] and volumetric descriptors [26]. The common goal for all of them is to extract the features that describe the shape of the objects. View-based extracts feature from a collection of the 2D projections of the 3D models. Su [27] implemented a stack of convolutional neural networks that takes multiple 2D projections of the 3D model generating a single compact descriptor for the object.

B. 3D Sketching in AR/VR

Rather than transforming 2D sketches into 3D representations, immersive modelling environments allow users to sketch and design 3D content using mid-air interaction techniques. Such techniques utilizing 3D space are intuitive to learn regardless of the user's expertise with the VR system [7]. A very early example of such an approach called HoloSketch [6] combined head-tracked stereoscopic shutter glasses and a CRT monitor with a six-axis mouse or *wand* for mid-air freehand drawing. While 3D freehand drawing can support expert 3D artists [28] and shows improving accuracy and uniformity of sketched objects over time [29], interaction techniques for sketching in 3D are still limited in terms of accuracy and user fatigue. When users are unable to determine their active drawing location and spatial relationships with other contents inside the scene, sketching errors occur. For *Sketch metaphor* Wacker *et al*. [30] designed and implemented a tool called ARPen, whose real-world position is tracked by a smartphone app that lets the user interact via mid-air sketch. In addition, they evaluated different techniques to select and move virtual objects with such a tool through a user study. 3D sketching is used in the works of Giunchi *et al*. [4], [5] where the task is retrieving a target model from a large collection of chairs by combining sketch input with 3D models by using a multi-view CNN to improve accuracy during the search. We develop our study considering the 3D sketch method based on the same freehand interaction but focusing specifically on the aspect of sketch generation.

C. Sketch on Surfaces in Virtual Reality

In recent years in VR environments, the use of 2D (Virtual) surfaces, has been explored to replace the inaccuracy brought by 3D tracking.Arora *et al*.developed SymbiosisSketch AR [31] that combines 3D drawing in mid-air with freehand drawing on a surface. They equipped the user with an Hololens, a tracked stylus, and a tablet and created a hybrid sketching system suitable for professional designers and amateurs. A virtually rendered tablet in VR can not provide the same latency-free response that a 2D tool or *Physical surfaces* can have. Different attempts of integrating a real physical 2D tablet have been made. Arora compared traditional sketching on a physical surface to sketching in VR, with and without a physical surface to rest the stylus on [32]. Wacker *et al*. [33] studied how accurately visual lines and concave/convex surfaces let users draw 3D shapes attached to physical/virtual objects in AR and Dorta [34] draws on virtual planes using a tracked tablets. A relevant study that couples pen with touch and gaze interactions on mobile surfaces within VR is PoVRPoint [35]. Such prototype for authoring presentations shows the improvement achieved with VR in tasks such as visual search and shape manipulation.

III. Sketch Modalities for VR

We describe our implemented model retrieval system, made of two parts: a VR application where users can interact by sketching and a backend that retrieves 3D objects from the dataset. In particular, we detail our four sketch interaction methods and the neural network used in the backend.

A. Interaction Methodologies

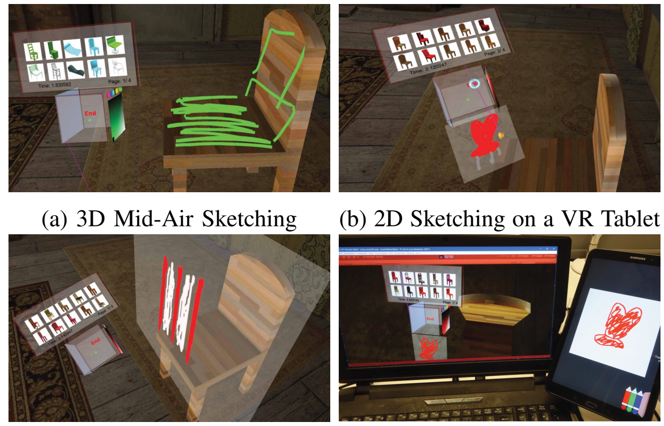
Four different interaction methods are examined below; three of them use 2D sketches generated on a different canvas and with different actions, and only one uses 3D sketch.

3D Mid-Air Sketching: This method, shown in Figure 1a, is similar to existing systems for sketching in VR and is directly based on the method for 3D mid-air sketching described in [5]. The user directly sketches in 3D space using a hand-held controller. While holding down the trigger button, a virtual stroke is applied in the air at the current position. By dragging the controller, strokes are extended towards a continuous line. Once the trigger button is released, the active stroke is considered to be completed, and a new stroke can be initiated. There is no theoretical limit to the volume the sketch can occupy and the user can create the sketch in any position.

2D Sketching on a VR Tablet: With this method, we mimic a natural way of sketching but are placed within VR. A 2D panel is attached to the user's non-dominant hand controller, see Figure 1b, referencing the familiar painting palette. As this method aims to simulate sketching on a portable tablet, we designed the 2D panel with a similar size to a commonly used tablet device with the largest dimension as the side of the squared panel. The actual sketching of lines is done using the laser coming out from the controller in the user's dominant hand. This makes the interaction technique a bi-handed approach as both hands are involved in the process of sketch creation,i.e., one hand performs the sketch while the other hand stabilises the drawing canvas.

2D Sketching on a VR Whiteboard: Similar to VR Tablet, the whiteboard method places a panel used to sketch in 2D, see Figure 1c. A familiar design paradigm, the whiteboard is larger than the tablet to provide additional space for drawing. Still, the user draws with a laser coming out from the controller and colliding with the canvas. The dimension of the whiteboard is linearly five times the size of the virtual tablet, with the same pixel resolution. As the whiteboard is positioned in a fixed location (room centre) this method only requires the use of the user's dominant hand.

2D Sketching on a Physical Tablet: We used a real-world tablet (Galaxy Tab A 10.1") that provides a physical surface to perform 2D sketching and tactile feedback while immersed in the virtual environment, see Figure 1d. This mimics the most common technique used by digital artists. The tablet is positioned on a table and requires a short registration procedure before the start of a sketching session. While the tablet is still limited in drawing space, the physical feedback



(a) 3D Mid-Air Sketching (b) 2D Sketching on a VR Tablet (c) 2D Sketching on a VR Whiteboard (d) 2D Sketching on a Physical Tablet

Fig. 1: Overview of the four implemented interaction modalities for sketch-based retrieval.

provided from the actual device aims to improve the stability during sketching. The user is able to sketch using her finger, and this approach does not require the use of a controller. The user's hands are tracked using a LEAP motion device as the virtual environment needs to visualise the correct position of the hands of the user. This additional tracking is necessary as we noticed during the first implementation that the absence of visual feedback for the finger position led to an unpleasant experience. This was mainly due to the user being unable to find the right contact point between her finger and the tablet. To track the tablet, we fixed an Oculus Touch controller via a Velcro strip attached to a Oculus Rockband VR Touch Guitar Mount Attachment. We calibrated the tablet, and created a digital panel in the virtual environment with its exact size.

B. Sketch, 3D collection and model retrival

The 3D sketch is implemented as a coloured strip with a wider section that is exposed to the current camera. Consequently, all the virtual cameras involved in the multi-view generation render the sketch-oriented along the wider section independently from the stroke's length or trajectory. The user can change the colour of the next sketch using the palette without modifying the strokes already generated. The 2D sketches are simply implemented as sequences of connected points on a canvas. As chairs database, we used a set of 3370 chairs that are part of ShapeNet [20]. The sketch retrieval process is provided by a backend that hosts a pre-loaded CNN model. This backend answers the sketches' visual queries by producing a list of the 40 models that are considered more similar to the input sketch. We used the VGG-M Matlab implementation of Su [27]. This implementation provides a single visual descriptor after elaborating the snapshots taken by VR software. The system generates a set of snapshots from virtual cameras positioned around the models. The CNN processes such images, and the final outcome is a unique descriptor. The encodings from each view are then merged

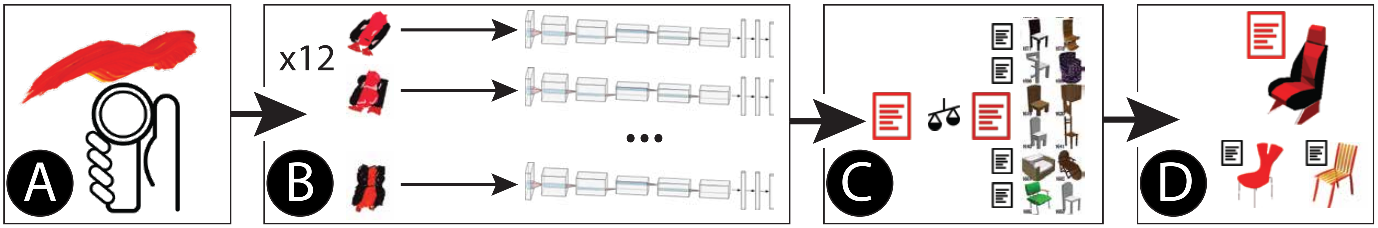


Fig. 2: Overview of the system's model retrieval mechanic. (A) the sketch created by the user results in a total set of 12 images (B) which are processed by 12 versions of the same CNN. One descriptor is generated and (C) compared through Euclidean distance with the descriptors previously calculated for all the chairs of the collection. The result of the search is (D) a small subset of the most similar chairs from which the user can select.

through view-pooling using a max function over the views, using the final 3 fully connected layers of VGG-M to create a descriptor. The response time of the system when the user triggers to retrieve the models is an average of 2.5 seconds. This system can be generalised for different types of objects, such as a table as stated in [5]. An overview of the entire system is depicted in Figure 2. Thus, the user interacts in VR by sketching (2D or 3D strokes) to find a target chair (Figure 2, A). When the user considers the sketch reliable to represent valid information, he triggers the system that creates a view-representation of the sketch itself that feeds a neural network (Figure 2, B). In the case of a 3D representation, multiple views are exploited, while for the 2D representation, only one view is used. The model processes the inputs and creates a descriptor (Figure 2, C) that is compared with the feature vectors of the target chairs (Figure 2, D). The user is presented with the most similar 40 chairs.

IV. Experiment

To evaluate our interaction methods for sketching in VR, we designed and conducted a user study performed in our lab. This user study is explorative because of the low number of participants but helps us understand sketching strategies.

***Participants**:* We used a within-subjects experiment to help to reduce the number of participants and errors associated with individual differences. To counterbalance possible carryover effects, the methods were randomised between the users. As our independent variable, we distinguish the methods used to sketch, 3D sketch, 2D sketch on a virtual tablet, 2D sketch on a fixed virtual whiteboard and 2D sketch on a real tablet. A total of 5 participants (4 male, 1 female, $25-43$ age range with avg. 34) volunteered for our study. All participants had previous experiences with VR and already used an Oculus RIFT, and Touch. We opted for a within-subjects study with a high number of searching tasks to minimise the possible fluctuations in drawing skills emerging from a study with many users on a few chairs. A total of 32 searches were performed by each user, eight for each of the four methods. In total, we have 160 search sessions. We recruited participants from the Computer Science department. We only recruited participants with average drawing abilities (self-reported) to demonstrate that users without particular drawing skills can interact proficiently with the proposed VR sketching tool.None of the participants were affected by colour-blindness and they were all compensated with £10. To avoid unbalanced results, all the users were not professional artists, and all self-reported to be not particularly skilled in drawing.

***Apparatus**:* The rendering of the Virtual Environment was performed in Unity 2018.2.13 using an Oculus RIFT DK1 headset with a connected laptop computer. The specification of the laptop is Intel i7 CPU, 64 GB RAM with Nvidia GeForce GTX 980M graphics card. The interaction with the 3D environment was provided by both the Oculus Touch paired with the headset and hand-tracking using a LEAP Motion device. We use a Galaxy A6 tablet for the real tablet session, tracked within Unity 2018.2 application via an Oculus Touch controller attached on the top right corner with an Oculus Rockband VR Touch guitar mount attachment.

***Procedure**:* Before starting the experiment, each participant signed a consent form and was instructed on the searching task. A period of 15 minutes was dedicated to training the user to develop confidence with the controllers, all four sketch mechanisms and the virtual environment. During the experiment, users had 3 minutes of rest between each method, and they could perform the task seated or standing up. Upon completion of the introduction, the experiment commenced. Users were placed inside a virtually furnished room with a chair model in the centre inside the virtual environment. The user searched the target chair by sketching and triggering the system. Results were displayed on a floating panel. The panel was positioned on the left controller, and the selection was made via the right-hand controller. The query results consisted of a selection of 40 chairs that are displayed in four panels. The user could navigate such panels with simple controls. We decided to display only ten models per panel to reduce the occlusion of the scene while providing enough variety for the user to choose from. Participants were asked to perform sketch searches for a given set of 8 different chairs in shape and textures for each method. For each session, the participant started with a randomly selected sketch interaction method and performed the search for each target chair of the 8 proposed in random order. Using the selected method, the participant started sketching to initiate the search for the presented target chair. Upon confirmation, the system provided the user with a set of potential chairs considered to be most similar to the created sketch. The participant could refine the search results by editing or detailing the sketch. When the participant was satisfied with the search, he selected the chair from the panel replacing the current 3D chair in the scene. This selection concluded the search task and would replace the presented 3D model inside the scene. Each session was given a time limit of 3 minutes, after which the search was considered terminated without a successful result. Additional functions available to the participants during sketching with all the different methods were: *undo* function to erase the sketch either entirely or partially, and a *colour palette* to characterise the sketch with a chosen colour. The experimenter recorded the success rate and completion time for each task. The accuracy is determined by counting the number of successful searches over the total number of searches in Table I. We also measured the average time to complete the tasks and the number of tries.

	Accuracy	Completion Time (avg)	N. of Tries (avg)
3D Sketch	**92.5%**	**71**	**2**
2D Virtual Pad	15%	156	5
2D Whiteboard	22.5%	169	6
2D Physical Pad	12.5	166	3

TABLE I: Table showing accuracy, average completion time and number of tries for each method.

V. Discussion

We investigated the differences between the four methods of interaction using sketches within a virtual environment. We asked each participant to find eight different chairs, providing four methods of searching. We analysed the emerging strategies from a qualitative point of view and measured the accuracy of the returned model, the time, and the number of queries required to terminate the search. In our work, we are aware that the backend is designed to work for 3D sketches, as its input is configured to receive up to 12 images from different points of view in the 3D scene. When 3D sketch is used, all the generated snapshots are used to calculate the final descriptor, while in the 2D interactions, only one image is used to generate the feature vector. Despite this unbalance result, we can compare fairly the 2d sketch-based conditions. Such comparison highlights different users' strategies that tend to mitigate the efficiency gap between the methods. These strategies try to take advantage of properties that 2D sketch exhibits, unlike 3D sketch. Firstly, we are trained to write, draw, and sketch on a bidimensional surface in our lives. 3D sketch is still a novel interface, and it is difficult to achieve proper daily training in real-life even if using a 3Doodler (3D pen) [36]. We notice during our experiment that users in a 3D sketch session users tend to sketch on a 2D surface instead of exploiting depth information, where they begin to gain confidence with this interaction over the session. Secondly, 2D sketch is quicker to achieve, and the user does not deal with depth information, which sometimes can be confusing. Moreover, when the user *Undo* his sketch, 3D sketch feedback is harder to follow due to possible occlusions or to the current point of view of the user that culls the removed stroke. Thirdly, while with 3D sketches, the user is stimulated to move around, sketch in different environment positions, and with different body poses, the 2D sketches do not ask for such a commitment.

We identified some emerging techniques and strategies within the 32 searches that each user did in total. The users explored how the input impacts the neural network response and exploited patterns that perform better for each method. We analyzed the sketches, reviewing them after recording each user session, and we explained the reasons for specific user behaviours. As mentioned above, 3D sketch interaction generates fully exploits the retrieval system, while each 2D interaction provides only a fraction of the input. Thus with 2D sketch, each user quickly developed the idea of depicting from their favourite view angle. They identify the projection of the model that delivers more information and starts to sketch on the provided surface (real or virtual). Also, the 3D sketch does not require the same accuracy required by 2D sketches for two reasons. Firstly, the backend gives good results with inaccurate 3D sketches, and the user finds the target before starting to detail more the sketch. Secondly, detailing a 2D sketch is simpler than 3D sketch. We noticed that for each 2D method, the user tended to detail the drawing to increase the probability of finding the target. This caused a longer execution time when users sketch in 2D, spending it drawing more than querying the system. In such a case, the user preferred to accurately fill areas between the edges on the 2D canvas. This disposition is not necessary for 3D where the system can interpret a few quick strokes as a filled surface. Thus, participants spent more time filling a part of a 2D plane than filling 2D surfaces immersed in 3D coordinates. In our system, we did not add complex functions, limiting the range of action of each user. Again, the motivation for these differences is the unbalanced results between 3D and 2D sketch inputs when processed by the system. The users tried to compensate for the lack of information in 2D sketch compared with the 3D counterpart in two ways: firstly, by adding more details in the 2D drawing, or secondly, by redrawing partially or entirely the sketch to achieve better results. Another strategy emerged naturally from some users making multiple drawings of the chair from different points of view. Such a multiple drawing strategy, in some cases, achieves the target chair successfully because of the multiple-view nature of the backend. This technique is capable of inserting additional information in the canvas that transit towards the neural network but simultaneously demands more sketching time. These 2D sketches' lack of efficiency deteriorates the user experience. The 2D drawing is affected by a continuous exploration of the feature space that can trigger the right system's response. This attitude denatures the process of intuitive and quick representation of a model. In terms of user interaction, 2D sketch modalities require from the user more query submission to the system, more attention, a firmer hand, and in particular, with virtual tablet, two hands working at the same time. This could have caused discomfort after a few minutes of sketching and, eventually, a loss of accuracy.

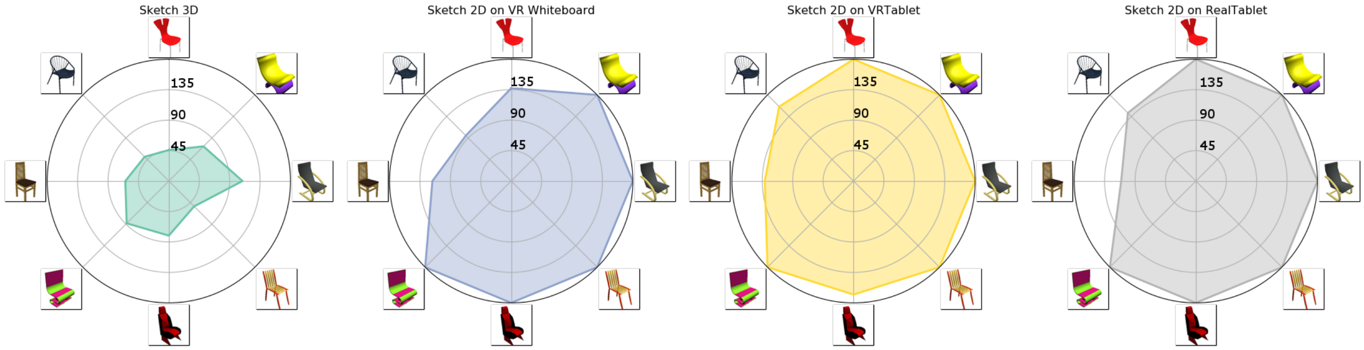


Fig. 3: The inner circles in each radar represent 45 seconds. The centre of each circle corresponds to time 0. Each radar shows the average time to complete the task for each chair, considering all the methods. The time is normalised to 3 minutes as the upper limit allowed for a search attempt.

Fig. 4: The four columns show representative images from the 3D sketch. The fifth column is the sketch from the virtual tablet method. The sixth column is the outcome of the whiteboard method, and the seventh column is the real tablet sketch. The last column is the image of the target chair.

A. 2D/3D Sketch Interaction and Shape Retrieval

During the experiment, we provide an application that takes sketches as inputs and provides a selected set of similar 3D chairs as output. We use the architecture described by Giunchi *et al*. [37] with the difference that our implementation uses only sketch without the possibility of combining it with the model. Such a difference allows us to compare 3D sketches and 2D sketches directly without the mediation of the 3D model in the scene. We identify two different stages for the 2D/3D sketching process: an interaction phase and a retrieval phase. We evaluate each phase by highlighting the pro and cons. In the 3D sketching interaction phase, users have the possibility of passing richer information to the system. Nevertheless, such a phase also presents the challenge of a novel sketching paradigm, as users are not used to sketching shapes in 3D dimensions. In particular, the third dimension requires the users to constantly change their point of view to control their drawing and avoid the depth compression effect [38] as well as the occlusion from previously designed sketch areas. Additionally, while the 2D canvas is a confined, limited space, the 3D space is not; such difference confers a sense of freedom during the 3D sketching but also a lack of spatial references. In addition, for a 2D sketch, it is simpler to achieve the desired result, both for the outlines and the filled areas. All these aspects impact the quality of the input sent to the system, undirectly impacting the output. The second phase is related to the implementation of the shape retrieval neural network and the input with which it is triggered. The 3D sketch is processed with a multiview projection that allows the processing in parallel of multiple points of view. The final pool layer store the most important values of the descriptors, merging them together. Such a mechanism permits saving the most important features in the descriptor, which are selected from different points of view. On the other hand, a 2D sketch has the disadvantage of a single point of view which may not contain the most relevant features that disambiguate the target chair from the others. While the multiview approach costs are higher as ten inputs will take ten times more to compute, users do not experience any perceivable latency because inference time is compatible with real-time interaction. Our exploratory study shows that some disadvantages of the 2D sketch can elicit discomfort or frustration if the system can not respond properly. Some mitigation strategies can be introduced. For example, multiple 2D canvas for sketching could be used to enrich the input offered by the 2D sketching technique. In this way, the user is focused on sketching, which is a pleasant activity, providing more relevant depictions from a different orientation. However, such a solution introduces a trade-off between time spent sketching and the accuracy of the final result. Secondly, a simple data augmentation method can be used to increase the number of points of view, giving the symmetry the chairs expose. A simple left-right flip operation can double the inputs and consider two different angles. A third possibility is to create a descriptors database that do not merge the features coming from the multiview process, allowing the algorithm to navigate through a larger descriptors space with 2D features instead of the merged (3D) version.

B. 2D Methods Comparison

Between 2D methods, the whiteboard shows better results than the virtual tablet for two main reasons: firstly, because of the fixed texture to draw on, and secondly, because of the larger canvas that can include more details and the interaction technique that allows higher precision. The fixed board contains the same number of pixels as the other 2D boards but with less pixel density. The fixed board provides a benefit in terms of less fatigue: only one hand was used to draw. The virtual tablet needs an effort from both the hands and arms, which increases the participant's fatigue. Moreover, this method requires arms-eye coordination effort. Although the distance between the surface and the drawing hand for the whiteboard is the largest among the 2D methods, this result shows that this aspect does not have a negative impact on the task. The real tablet is interesting because of the physical interaction that is completely absent in all the other methods. While this is a pleasant novelty at the beginning of the session, contrary to what was imagined, the sketch done directly with the finger was less precise if compared to the sketch generated by a remote controller. Moreover, mixing VR with a tracked canvas was not sufficient to guarantee a decent user experience. The absence of the embodiment (avatar's hand) lets the user, still immersed in the scene, become disoriented during the drawing process. It is unclear where the finger is regarding the canvas position. We managed this issue using a Leap Motion to track the finger continuously while displaying an avatar. However, due to the noise, the position of the user's finger was not correctly registered, deteriorating his/her ability to sketch. This could indicate that even synthetic haptic feedback could aid in the drawing for the other methods, even in 3D Sketching. combined with the success rate of the methods motivate the low value of triggering events for 3D sketch and real tablet techniques. 3D sketch had a high rate of success and a low number of queries indicating the method's high efficiency. On the contrary, even if the real tablet had a few triggering events, it does not mean that it is efficient. Its low success rate and problematic interaction lead the user to give up quickly, avoiding an extensive search as other 2D methods.

VI. Future Works and Conclusion

We investigate some limitations of this system to improve the search accuracy or experience, outlining the following:

Query Descriptors*:* One limitation is the number of views used when creating the descriptors. Generating and processing additional images could increase the accuracy but simultaneously needs more computational power. Moreover, the size of the textures used to infer the model shapes are fixed and limited to the input size 224×224 pixels. Such a limitation is connected to the fact that we used a VGG-M deep descriptor. However, more recent neural network models have outperformed this model both in accuracy and reduced number of parameters (i.e., less processing time). User experiences for both 3D and 2D sketches could benefit from these new versions because of latency reduction, better accuracy of chair predictions and higher resolution input images. We implemented our system as distinct modules that can be replaced easily. One possible improvement is extending the study to different neural networks to do a comprehensive survey. These solutions could be fine-tuned to VR or learnt through active learning to become bespoke to the users' style.

Expanded comparisons*:* Although our study aimed to compare four comprehensive methods for retrieval using 3D or 2D sketches, the study could be extended to other interaction methods. An additional method could be a tracked pen coupled with the tablet to increase accuracy. An interesting follow-up would be comparing the user experience and accuracy obtained in a virtual environment with a physical environment.

Sketch in Augmented Reality*:* Understanding how users sketch in VR allows us to perform controlled user studies, but many application fields require real-world interaction. By trivial adaptations to this method, we could see applications within Film & TV for set design or architecture for model creation. Such applications require their nuances to be considered and would need to take advantage of the extensions mentioned above. In scene modelling, different interaction strategies have been investigated within Virtual Reality. However, we specifically explore database navigation for scene modelling. Sketch interaction has been shown to be a reliable method when compared with text queries or even linear navigation of the collection, allowing the expressive visual description of the query. We extend on studies of 3D and 2D sketches which have been tested in different contexts related to collection navigation but a comparative study between them set within virtual reality is missing. We proposed a user study that fills this gap by comparing different sketch-based mechanisms, including 3D, 2D sketches in virtual reality with a tablet or a whiteboard and a method that considers the use of a physical tablet. In our pilot, we collected the time to perform, analyzed the different interactions and discovered that amongst the 2D Sketch methods, describing user strategies that arise from the asymmetric interaction and system response to the different visual queries. We believe these results suggest that 3D sketching is a promising general interface for sketch retrieval in virtual environments, but 2D sketching represents an alternative in terms of the possibility of quickly triggering a retrieval system in a more intuitive fashion and overcoming the time of becoming familiar with sketching with depth.

References

[1] A. del Bimbo and P. Pala, "Visual image retrieval by elastic matching of user sketches," *IEEE Trans. Pat. Ana. & Mach. Int.*, vol. 19, no. 2, pp. 121–132, February 1997.

[2] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, “A descriptor for large scale image retrieval based on sketched feature lines,” in *SBIM*, 2009, pp. 29–36.
[3] R. Hu and J. Collomosse, “A performance evaluation of gradient field hog descriptor for sketch based image retrieval,” *Comput. Vis. Image Underst.*, vol. 117, no. 7, pp. 790–806, Jul. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.cviu.2013.02.005
[4] D. Giunchi, S. James, and A. Steed, “Model retrieval by 3d sketching in immersive virtual reality,” in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 559–560.
[5] D. Giunchi, S. James, and A. Steed, “3d sketching for interactive model retrieval in virtual reality,” in *Proceedings of the Joint Symposium on Computational Aesthetics and Sketch-Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering*, ser. Expressive ’18. New York, NY, USA: ACM, 2018, pp. 1:1–1:12.
[6] M. F. Deering, “Holosketch: A virtual reality sketching/animation tool,” *ACM Trans. Comput.-Hum. Interact.*, vol. 2, no. 3, pp. 220–238, Sep. 1995.
[7] G. Wesche and H.-P. Seidel, “Freedrawer: A free-form sketching system on the responsive workbench,” in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, ser. VRST ’01. New York, NY, USA: ACM, 2001, pp. 167–174.
[8] M. J. Fonseca, A. Ferreira, and J. A. Jorge, “Towards 3d modeling using sketches and retrieval,” in *Proceedings of the First Eurographics Conference on Sketch-Based Interfaces and Modeling*, ser. SBM’04. Eurographics Association, 2004, pp. 127–136. [Online]. Available: http://dx.doi.org/10.2312/SBM/SBM04/127-136
[9] 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*, vol. 164, pp. 27–37, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314217301194
[10] T. Bui, L. Ribeiro, M. Ponti, and J. P. Collomosse, “Generalisation and sharing in triplet convnets for sketch based visual search,” *CoRR*, vol. abs/1611.05301, 2016.
[11] K. Pang, K. Li, Y. Yang, H. Zhang, T. M. Hospedales, T. Xiang, and Y.-Z. Song, “Generalising fine-grained sketch-based image retrieval,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 677–686.
[12] Y. Qi, Y. Song, H. Zhang, and J. Liu, “Sketch-based image retrieval via siamese convolutional neural network,” in *2016 IEEE International Conference on Image Processing, ICIP 2016, Phoenix, AZ, USA, September 25-28, 2016*, 2016, pp. 2460–2464. [Online]. Available: https://doi.org/10.1109/ICIP.2016.7532801
[13] B. Li, Y. Lu, F. Duan, S. Dong, Y. Fan, L. Qian, H. Laga, H. Li, Y. Li, P. Liu, M. Ovsjanikov, H. Tabia, Y. Ye, H. Yin, and Z. Xue, “3d sketch-based 3d shape retrieval,” in *Eurographics Workshop on 3D Object Retrieval*, A. Ferreira, A. Giachetti, and D. Giorgi, Eds. The Eurographics Association, 2016.
[14] J. Ashley, M. Flickner, J. L. Hafner, D. Lee, W. Niblack, and D. Petkovic, “The query by image content (QBIC) system,” in *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, San Jose, California, May 22-25, 1995.*, 1995, p. 475. [Online]. Available: http://doi.acm.org/10.1145/223784.223888
[15] P. M. A. Sousa and M. J. Fonseca, “Sketch-based retrieval of drawings using spatial proximity,” *J. Vis. Lang. Comput.*, vol. 21, no. 2, pp. 69–80, 2010. [Online]. Available: https://doi.org/10.1016/j.jvlc.2009.12.001
[16] A. D. Bimbo, P. Pala, and S. Santini, “Visual image retrieval by elastic deformation of object sketches,” in *Proceedings IEEE Symposium on Visual Languages, St. Louis, Missouri, USA, October 4-7, 1994*, 1994, pp. 216–223. [Online]. Available: https://doi.org/10.1109/VL.1994.363615
[17] A. Chalechale, G. Naghdy, and A. Mertins, “Sketch-based image matching using angular partitioning,” *IEEE Trans. Systems, Man, and Cybernetics, Part A*, vol. 35, no. 1, pp. 28–41, 2005. [Online]. Available: https://doi.org/10.1109/TSMCA.2004.838464
[18] E. D. Sciascio, G. Mingolla, and M. Mongiello, “Content-based image retrieval over the web using query by sketch and relevance feedback,” in *VISUAL*, ser. Lecture Notes in Computer Science, vol. 1614. Springer, 1999, pp. 123–130.
[19] D. G. Ruiz, S. Saito, and M. Nakajima, “Sketch-to-collage,” in *SIGGRAPH Posters*. ACM, 2007, p. 35.
[20] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, “3d shapenets: A deep representation for volumetric shapes,” in *CVPR*. IEEE Computer Society, 2015, pp. 1912–1920.
[21] K. V. Vishwanath, D. Gupta, A. Vahdat, and K. Yocum, “Modelnet: Towards a datacenter emulation environment,” in *2009 IEEE Ninth International Conference on Peer-to-Peer Computing*. IEEE, 2009, pp. 81–82.
[22] Q. Zhou and A. Jacobson, “Thingi10k: A dataset of 10, 000 3d-printing models,” *CoRR*, vol. abs/1605.04797, 2016. [Online]. Available: http://arxiv.org/abs/1605.04797
[23] H. Fu, R. Jia, L. Gao, M. Gong, B. Zhao, S. Maybank, and D. Tao, “3d-future: 3d furniture shape with texture,” *International Journal of Computer Vision*, vol. 129, no. 12, pp. 3313–3337, 2021.
[24] K. R. Bronstein M. A., Bronstein M. M., “Topology-invariant similarity of nonrigid shapes,” *Int. J. Comput. Vis.*, vol. 81, pp. 281 —- 301, 2009.
[25] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, “Shape distributions,” *ACM Transactions on Graphics*, vol. 21, pp. 807–832, 2002.
[26] R. M. Rustamov, “Robust volumetric shape descriptor,” in *3DOR*. Eurographics Association, 2010, pp. 1–5.
[27] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, “Multi-view convolutional neural networks for 3d shape recognition,” in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ser. International Conference on Computer Vision ’15. IEEE Computer Society, 2015, pp. 945–953. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2015.114
[28] R. Schmidt, A. Khan, G. Kurtenbach, and K. Singh, “On expert performance in 3d curve-drawing tasks,” in *Proceedings of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling*, ser. SBIM ’09. New York, NY, USA: ACM, 2009, pp. 133–140.
[29] E. Wiese, J. H. Israel, A. Meyer, and S. Bongartz, “Investigating the learnability of immersive free-hand sketching,” in *Proceedings of the Seventh Sketch-Based Interfaces and Modeling Symposium*, ser. SBIM ’10. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2010, pp. 135–142.
[30] P. Wacker, O. Nowak, S. Voelker, and J. Borchers, “Arpen: Mid-air object manipulation techniques for a bimanual ar system with pen and smartphone,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’19. New York, NY, USA: ACM, May 2019, pp. 619:1–619:10.
[31] R. Arora, R. Habib Kazi, T. Grossman, G. Fitzmaurice, and K. Singh, “Symbiosissketch: Combining 2d & 3d sketching for designing detailed 3d objects in situ,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’18. New York, NY, USA: ACM, 2018, pp. 185:1–185:15.
[32] R. Arora, R. H. Kazi, F. Anderson, T. Grossman, K. Singh, and G. Fitzmaurice, “Experimental evaluation of sketching on surfaces in vr,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’17. New York, NY, USA: ACM, 2017, pp. 5643–5654.
[33] P. Wacker, A. Wagner, S. Voelker, and J. Borchers, “Physical guides: An analysis of 3d sketching performance on physical objects in augmented reality,” in *Proceedings of the Symposium on Spatial User Interaction*, ser. SUI ’18. New York, NY, USA: ACM, 2018, pp. 25–35.
[34] T. Dorta, G. Kinayoglu, and M. Hoffmann, “Hyve-3d and the 3d cursor: Architectural co-design with freedom in virtual reality,” *International Journal of Architectural Computing*, vol. 14, no. 2, pp. 87–102, 2016.
[35] V. Biener, T. Gesslein, D. Schneider, F. Kawala, A. Otte, P. Kristensson, . Pahud, E. Ofek, C. Campos, M. Kljun, K. u. Pucihar, and J. Grubert, “Povrpoint: Authoring presentations in mobile virtual reality,” 2022.
[36] K. D. Jackson B., “Lift-off: Using reference imagery and freehand sketching to create 3d models in vr,” *IEEE Transactions on visualization and Computer Graphics*, vol. 22, no. 4, 2016.
[37] D. Giunchi, S. James, and A. Steed, “3d sketching for interactive model retrieval in virtual reality,” in *Proceedings of the Joint Symposium on Computational Aesthetics and Sketch-Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering*. ACM, 2018, p. 1.
[38] D. J. Finnegan, E. O’Neill, and M. J. Proulx, “Compensating for distance compression in audiovisual virtual environments using incongruence,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’16. Association for Computing Machinery, 2016, p. 200–212. [Online]. Available: https://doi.org/10.1145/2858036.2858065