Eurographics Workshop on 3D Object Retrieval (2019)
SHREC - 3D Shape Retrieval Contest (2019)
S. Biasotti, G. Lavoué, B. Falcidieno, and I. Pratikakis (Editors)
R. C. Veltkamp (SHREC organizer)

SHREC'19 Track: Extended 2D Scene Sketch-Based 3D Scene Retrieval

Juefei Yuan[†‡1], Hameed Abdul-Rashid[†‡1], Bo Li[†‡*1], Yijuan Lu[†‡2], Tobias Schreck[†3], Ngoc-Minh Bui[‡4,5], Trong-Le Do[‡4,5], Khac-Tuan Nguyen[‡4], Thanh-An Nguyen[‡5], Vinh-Tiep Nguyen[‡6], Minh-Triet Tran[‡4,5], Tianyang Wang[‡7]

[1] School of Computing Sciences and Computer Engineering, University of Southern Mississippi, USA
[2] Department of Computer Science, Texas State University, USA
[3] Institute of Computer Graphics and Knowledge Visualization, Graz University of Technology, Austria
[4] Faculty of Information Technology, Vietnam National University - Ho Chi Minh City, Vietnam
[5] Software Engineering Lab, Vietnam National University - Ho Chi Minh City, Vietnam
[6] University of Information Technology, Vietnam National University - Ho Chi Minh City, Vietnam
[7] Department of Computer Science & Information Technology, Austin Peay State University, USA

Abstract

Sketch-based 3D scene retrieval is to retrieve 3D scene models given a user's hand-drawn 2D scene sketch. It is a brand new but also very challenging research topic in the field of 3D object retrieval due to the semantic gap in their representations: 3D scene models or views differ from non-realistic 2D scene sketches. To boost this interesting research, we organized a 2D Scene Sketch-Based 3D Scene Retrieval track in SHREC'18, resulting a ***SceneSBR18*** *benchmark which contains 10 scene classes. In order to make it more comprehensive, we have extended the number of the scene categories from the initial 10 classes in the* ***SceneSBR2018*** *benchmark to 30 classes, resulting in a new and more challenging benchmark* ***SceneSBR2019*** *which has 750 2D scene sketches and 3,000 3D scene models. Therefore, the objective of this track is to further evaluate the performance and scalability of different 2D scene sketch-based 3D scene model retrieval algorithms using this extended and more comprehensive new benchmark.*

In this track, two groups from USA and Vietnam have successfully submitted 4 runs. Based on 7 commonly used retrieval metrics, we evaluate their retrieval performance. We have also conducted a comprehensive analysis and discussion of these methods and proposed several future research directions to deal with this challenging research topic. Deep learning techniques have been proved their great potentials again in dealing with this challenging retrieval task, in terms of both retrieval accuracy and scalability to a larger dataset. We hope this publicly available benchmark, together with its evaluation results and source code, will further enrich and promote 2D scene sketch-based 3D scene retrieval research area and its corresponding applications.

Categories and Subject Descriptors (according to ACM CCS): H.3.3 [Computer Graphics]: Information Systems—Information Search and Retrieval

1. Introduction

2D scene sketch-based 3D scene model retrieval is to retrieve 3D scene models given a user's hand-drawn 2D scene sketch. Due to the intuitiveness in sketching, 2D scene sketch-based 3D scene model retrieval research topic has many important applications such as 3D scene reconstruction, 3D geometry video retrieval, and 3D AR/VR entertainment. It is a challenging research topic in the field of 3D scene model retrieval due to the semantic gap in their representations: non-realistic 2D scene sketches significantly differ from 3D scene models or their views. While, existing 3D model retrieval algorithms have mainly focused on single object retrieval and have not handled retrieving such 3D scene content, which involves a lot of new research questions and challenges. This situation is due to two major reasons: 1) there exist a very limited number of available 3D scene shape benchmarks, thus it is challenging to collect a large-scale 3D scene dataset; 2) a big semantic gap exists between the iconic representations of hand-drawn 2D scene sketches and the accurate 3D coordinate representations of 3D scenes. Therefore, retrieving 3D scene models using 2D scene

† Track organizers.
‡ Track participants.
*Corresponding author. For any question related to the track, please contact Bo Li. E-mail: bo.li@usm.edu or li.bo.ntu0@gmail.com.

DOI: 10.2312/3dor.20191059

sketch queries makes this research direction meaningful, interesting and promising, but challenging as well.

To promote this research direction, we built a **SceneSBR2018** benchmark [YLL18a, YLL*18b] and organized a SHREC'18 track on 2D scene sketch-based 3D scene retrieval. However, as can be seen, **SceneSBR2018** contains only 10 distinct scene classes, and this is one of the reasons that all the three deep learning-based participating methods have achieved excellent performance on it. Considering this, after the track we have tripled [YARLL19b] the size of **SceneSBR2018**, resulting in an extended benchmark **SceneSBR2019**, which has 750 2D scene sketches and 3,000 3D scene models. Similarly, all the 2D scene sketches and 3D scene models are equally classified into 30 classes. We have kept the same set of 2D scene sketches and 3D scene models belonging to the initial 10 classes of **SceneSBR2018**.

Hence, this track seeks participants who will provide new contributions to further advance 2D scene sketch-based 3D scene retrieval for evaluation and comparison, especially in terms of scalability to a larger number of scene categories, based on the new benchmark **SceneSBR2019**. Similarly, we also provide corresponding evaluation code for computing a set of performance metrics similar to those used in the Query-by-Model retrieval technique.

2. SceneSBR Benchmark

2.1. Overview

Building process. By referring to several of the most popular 2D/3D scene datasets, such as Places [ZLK*17] and SUN [XEH*16], we finally selected 30 scene classes (including the initial 10 classes in **SceneSBR2018**) based on the criteria of *popularity* (most commonly seen categories). Based on a voting mechanism among three people (two graduate student voters and a faculty moderator), the most popular 30 scene classes were selected from the 88 common scene labels in the Places88 dataset [ZLK*17]. It is worth noting that the 88 scene categories are already shared by ImageNet [DDS*09], SUN, and Places. For the additional 20 classes' (sketches and models) data collection, we gathered their sketches from Flickr and Google Image, and downloaded their SketchUp 3D scene models (in both the original ".SKP" format and transformed ".OBJ" format) from 3D Warehouse [Tri18].

Benchmark Details. Our extended 2D scene sketch-based 3D scene retrieval benchmark **SceneSBR2019** expands the initial 10 classes of **SceneSBR2018** by adding 20 new classes totaling a more comprehensive dataset of 30 classes. 500 more 2D scene sketches have been added to its 2D scene sketch dataset and 2,000 more SketchUp 3D scene models (".SKP" and ".OBJ" formats) to its 3D scene dataset. Each of the additional 20 classes has the same number of 2D scene sketches (25) and 3D scene models (100), as well. Therefore, **SceneSBR2019** contains a complete dataset of 750 2D scene sketches (25 per class) and 3,000 3D scene models (100 per class) across 30 scene categories. Examples for each class are demonstrated in both **Fig.** 1 and **Fig.** 2.

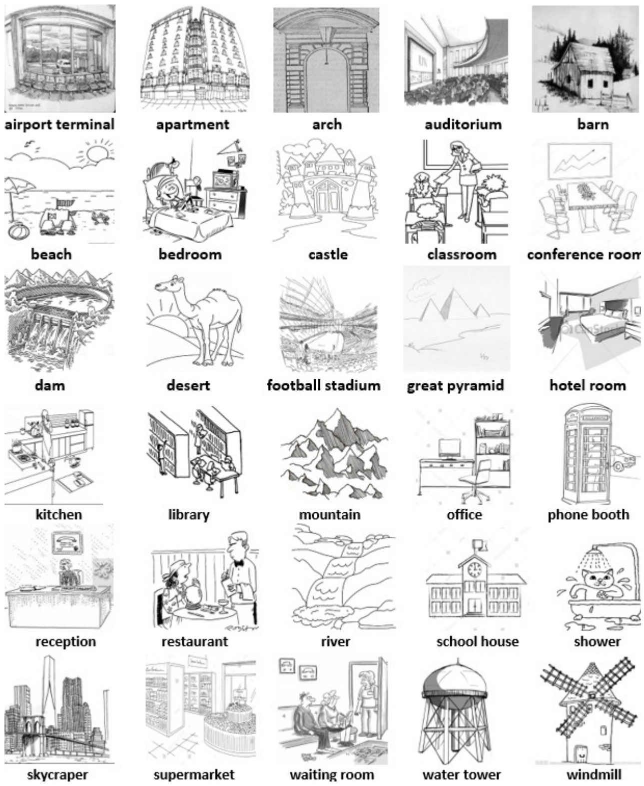

Figure 1: *Example 2D scene sketches (one example per class) in our* ***SceneSBR2019*** *benchmark.*

Similar to the SHREC'18 sketch track, we randomly select 18 sketches and 70 models from each class for training and the remaining 7 sketches and 30 models are used for testing, as shown in **Table** 1. Participants need to submit results on the training and testing datasets, respectively, if they use a learning-based approach. Otherwise, the retrieval results based on the complete (750 query sketches, 3,000 target scene models) datasets are needed.

Table 1: *Training and testing datasets (per class) of our* ***Extended SceneSBR*** *benchmark.*

SceneSBR Benchmark	Sketch	Model
Training	18	70
Testing	7	30
Total (per class)	25	100

2.2. 2D Scene Sketch Dataset

The 2D scene sketch dataset comprises 750 2D scene sketches (10 classes, each with 25 sketches). One example per class is demonstrated in **Fig.** 1.

Figure 2: *Example 3D scene models (one example per class, shown in one view) in our* ***SceneSBR2019*** *benchmark.*

2.3. 3D Scene Model Dataset

The 3D scene model dataset is built on the selected 3,000 3D scene models downloaded from 3D Warehouse. Each class has 100 3D scene models. One example per class is shown in **Fig.** 2.

2.4. Evaluation Method

The objective of this track is to evaluate the performance of different 2D scene sketch-based 3D scene retrieval algorithms using a 2D sketch query dataset and a 3D scene model dataset. While, each algorithm targets retrieving 3D scene models that belong to the same class as that of each query 2D scene sketch. To have a comprehensive evaluation of the retrieval algorithm, we employ seven commonly adopted performance metrics in the 3D model retrieval community [LLL*15, LLG*14]. They are Precision-Recall (PR) diagram, Nearest Neighbor (NN), First Tier (FT), Second Tier (ST), E-Measure (E), Discounted Cumulated Gain (DCG) and Average Precision (AP). We also have developed the code to compute them[1].

3. Participants

There were two groups who registered for the track. One group comes from USA, and one group comes from Vietnam. Each group was given one month to complete the competition. They were requested to submit both their results and methods' description.

Both two groups have successfully participated in the SHREC'19 track on Extended 2D Scene Sketch-Based 3D Scene Retrieval. Four (4) rank list results (runs) for two (2) different methods developed by two (2) groups have been submitted. The participants and their runs are listed as follows:

- *RNSRAP1* and *RNSRAP2* submitted by Ngoc-Minh Bui, Trong-Le Do, Khac-Tuan Nguyen, Thanh-An Nguyen, Vinh-Tiep Nguyen, Minh-Triet Tran from Vietnam National University - Ho Chi Minh City, Vietnam (Section 4.1);
- *VMV-AlexNet* and *VMV-VGG* submitted by Juefei Yuan, Hameed Abdul-Rashid, Bo Li, Tianyang Wang, and Yijuan Lu from the University of Southern Mississippi, USA, Austin Peay State University, USA and Texas State University, USA (Section 4.2).

4. Methods

4.1. RNSRAP: ResNet50-Based Sketch Recognition and Adapting Place Classification for 3D Models Using Adversarial Training, by N. Bui, T. Do, K. Nguyen, T. Nguyen, V. Nguyen, and M. Tran

4.1.1. Sketch Image Classification with Data Augmentation

They use data augmentation to enrich the training data for sketch recognition. They first collect a dataset of natural scene images from Google. They do not only crawl images with exactly 30 concepts in this track but also extend the list of concepts with semantically related concepts. For example, instead of searching only "desert" images, they expand the query terms into "desert", "camel", "cactus", etc. By using this query expansion strategy, they expect that their natural scene corpus can be used to link the gap of visual differences in the sketch image dataset.

The natural scene images are transformed into sketch-like images. For this track, they simply use automated tools for image transformation. However, they intend to use image translation to adapt images from the natural domain into the sketch-like domain.

For each image in the enriched dataset, they use ResNet-50 [HZRS16] to extract features and train a simple image classification network with 30 concepts.

4.1.2. 3D Scene Classification with Multiple Screenshots, Domain Adaptation, and Concept Augmentation

They perform a two-step process for 3D scene classification with multiple screenshots. The first step of their method is to use a number of classification models and domain adaptation to classify the 3D scene. The second step is to take advantage of visual concepts to refine the final result. The overview of the method is illustrated in **Fig.** 3. In the first step, they train multiple classification models and use the voting scheme to ensemble the classification results. Because there are fair resemblances between 3D scene models and scenery images, they perform transfer learning from models pretrained on two datasets: ImageNet [DDS*09] and Places365 [ZLK*17].

The first model is to extract feature vectors for each image using ResNet-50 [HZRS16] pretrained on ImageNet and Places365

[1] URL: http://orca.st.usm.edu/~bli/SceneSBR2019/Evaluation.html.

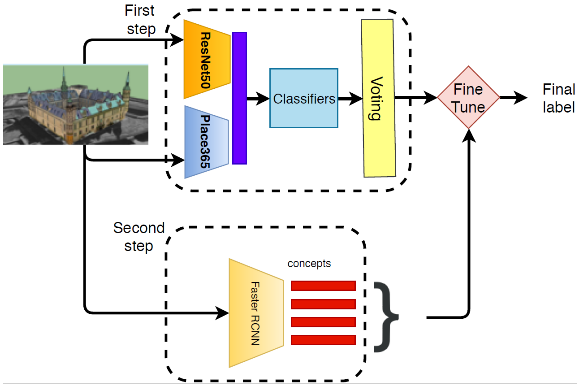

Figure 3: *Two-step process of the 3D scene classification method.*

dataset, respectively, then feed these feature vectors to a fully-connected neural network that has one to two hidden layers. The number of nodes in each hidden layers is set to 128, 192, 256, or 320 nodes and they choose the architecture that yields the highest classification accuracy to be the final result of this model.

They also extract 365-D scene attribute features for each image using Places365 and concatenate with the 2048-D feature vector of that image to form a 2413-D feature, which is later reduced to 512-D by PCA to train a third classification network. The extracted scene attributes may provide useful information, such as "outdoor", "natural light", "trees" for a screenshot from a model in the "mountain" category. Concatenating two vectors' results in a higher dimensional input may make the model prone to overfitting. Therefore, each feature is normalized to have zero mean and unit variance and then they use PCA to reduce the size of the input vectors to 512-D.

Their second model is to collect real images of the 30 different categories from Places365 dataset and the Internet (for the "great_pyramid" class). They collect 1,000 images per category. Then they use the weights of the last fully connected layer trained by this small-scale dataset to initialize the weights when training on the screenshots dataset.

Next, they apply their proposal of domain adaptation (used in SHREC 2018) [YLL18a, YLL*18b] to classify a 2D screenshot of a 3D scene. Concretely, they first train an adversarial network to learn the representation of a 3D model to be close to the representation of a corresponding natural image. They treat the natural image domain as the source domain and the screenshots of the 3D model as the target domain. A discriminator is used to distinguish between the representations of the two domains. They train the target representation via an adversarial loss so that the two representations are indistinguishable to the discriminator. Then, using the adaptive representation of a 3D model, they train a number of simple networks. The predicted labels from the networks are assembled via voting to select the final label for the 3D model.

Because of the wide variation in the design of a 3D scene, it is not enough to classify the category of a scene simply by extracting the feature (from ResNet-50) or from the features of scene attributes (from Places365, even after domain adaptation). This motivates their proposal to employ object/entity detectors to identify entities related to certain concepts existing in a screenshot.

In the second step of the proposed method, they first collect a dataset of natural images from the Internet corresponding to the concepts that are related to the 30 scene categories. For example, they use the query terms such as "cactus", "camel", etc to serve the scene classification for "desert". They train their set of object detectors from this dataset of natural images with Faster RCNN [RHGS15]. Then they apply their detectors to identify entities that might appear in a scene, such as "book" (in a library), "umbrella" (in a beach), etc. By this way, they further refine their retrieval results.

4.2. VMV-AlexNet, VMV-VGG: View and Majority Vote Based 3D Scene Retrieval Algorithm, by J. Yuan, H. Abdul-Rashid, B. Li, T. Wang and Y. Lu

They proposed a View and Majority Vote based 3D scene retrieval algorithm (VMV) [YARLL19b] by either employing the AlexNet or the VGG-16 model. Its architecture is illustrated in **Fig.** 4.

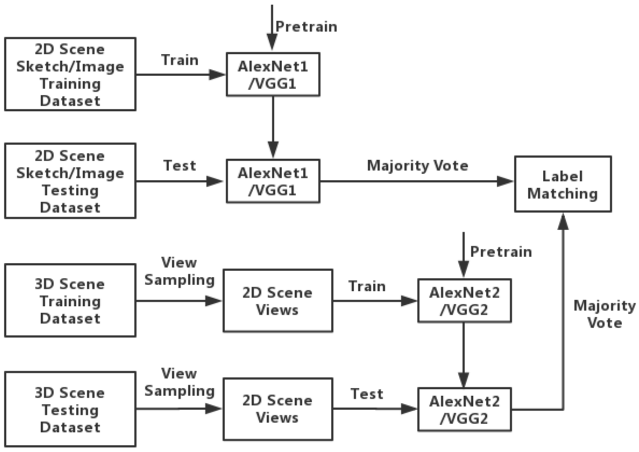

Figure 4: *VMV architecture [YARLL19b].*

4.2.1. 3D Scene View Sampling

For each 3D scene model, they centered each 3D scene model in a 3D sphere. They developed a QMacro script program to automate the operations of the SketchUp software to perform the view sampling, and sampled 13 scene view images automatically. They uniformly arrange 12 cameras on the equator of the bounding sphere of a 3D scene model, and one on the top of the sphere. One example is shown in **Fig.** 5.

4.2.2. Data Augmentation

To avoid overfitting issues, before each pre-training or training, they employed data augmentation technique (rotations, shifts and flips) [YLL16] to enlarge the related dataset's size by 500 times.

Figure 5: *A 13 sampled scene view images example of an apartment scene model [YARLL19b].*

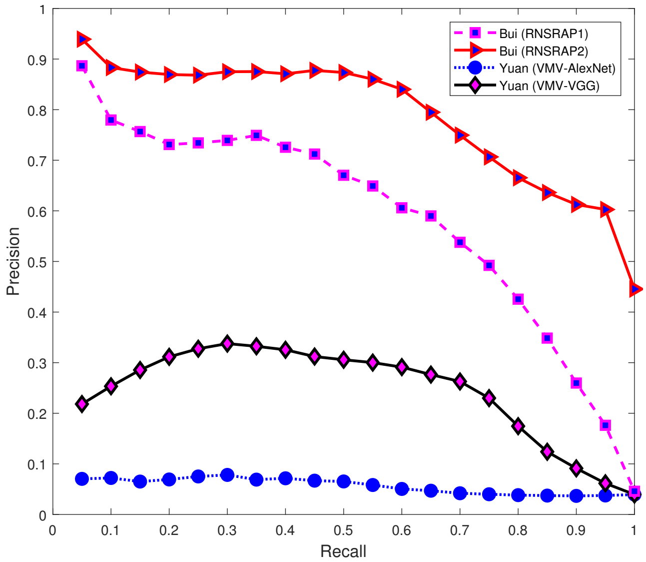

Figure 6: *Precision-Recall diagram performance comparisons on testing dataset of of our **SceneSBR19** benchmark for two learning-based participating methods.*

4.2.3. Pre-training and Fine-tuning

They pre-trained the AlexNet1/VGG1 model on the TU-Berlin sketch dataset [EHA12] for 500 epochs, and pre-trained AlexNet2/VGG2 on the Places scene image dataset [ZLK*17] for 100 epochs. After pre-training, they fine-tuned the AlexNet1/VGG1 on the 2D scene sketches training dataset, and fine-tuned the AlexNet2/VGG2 on the 2D scene views training dataset, respectively.

4.2.4. Sketch/View Classification and Majority Vote-based Label Matching

They obtained classification vectors by feeding well-trained AlexNet1/VGG1 with a 2D scene sketch query, or AlexNet2/VGG2 with the target 2D scene views testing dataset. Finally, based on the query's classification vector and a target 3D scene's 13 classification vectors, they generated the rank list for each sketch query by using a majority vote-based label matching method.

For more details, please refer to [YARLL19b], while the code is also publicly accessible[2].

5. Results

In this section, we comparatively evaluate the four runs of the two methods submitted by the two groups. The seven metrics aforementioned in Section 2.4 are adopted to measure the retrieval performance. **Fig.** 6 and **Table** 2 show the comparative results of the two learning-based participating methods on the target dataset.

All the four runs of the two methods submitted by both participating groups are learning-based methods. As shown in the **Fig.** 6 and **Table** 2, Bui's RNSRAP algorithm (run 2) performs the best. More details about the retrieval performance of each individual query of every participating method are available on the **SceneSBR2019** track homepage [YARLL19a].

Both of the two submitted approaches utilized CNN models, which contribute a lot to the achieved performance of those two learning-based approaches. Since deep learning techniques are widely utilized in many latest sketch-based 3D model retrieval methods, it can be regarded as the currently most popular and promising machine learning technique for 2D/3D feature learning and related retrieval. In fact, we can see that the deep learning models which are adopted in these two methods, especially Bui's method, perform well in dealing with this challenging retrieval task. They improved their method used in the **SceneIBR2018** track by utilizing object-level semantic information for data augmentation and refining retrieval results, which helps to advance the retrieval performance further. Considering there is still much room for further improvement in the retrieval accuracy as well as the scalability issue, we believe it is very promising to further propose a practical retrieval algorithm for large-scale 2D sketch-based 3D scene retrieval by utilizing both deep learning and scene semantic information.

Using the same target 3D scene dataset of the (**SceneSBR2019**) benchmark, we also organized another SHREC'19 track titled "Extended 2D Image-Based 3D Scene Retrieval" (**SceneIBR2019**) [ARYLL19]. We replaced the query dataset with a 2D query image dataset which contains 1000 images for each of the 30 classes. The **SceneIBR2019** track has one more participating method. We can see that from the corresponding figures and tables, the overall performance achieved on the **SceneIBR2019** track is better than that on the **SceneSBR2019** track. We believe

[2] URL: http://orca.st.usm.edu/~bli/Scene_SBR_IBR/index.html.

Table 2: *Performance metrics comparison on two different datasets of of our* ***SceneSBR2019*** *benchmark for two learning-based participating method.*

Participant	Method	NN	FT	ST	E	DCG	AP
Testing dataset							
Bui	RNSRAP1	0.914	0.668	0.728	0.665	0.825	0.581
	RNSRAP2	**0.943**	**0.818**	**0.870**	**0.814**	**0.913**	**0.786**
Yuan	VMV-AlexNet	0.024	0.046	0.084	0.047	0.386	0.057
	VMV-VGG	0.081	0.281	0.369	0.280	0.533	0.243

that there are at least three possible reasons leading to this better performance: (1) **SceneIBR2019** has a much larger 2D image query dataset which contributes a lot to the deep neural networks' training; (2) **SceneIBR2019**'s query images have more accurate 3D shape information than **SceneSBR2019**'s query sketches; and (3) the semantic gap between **SceneIBR2019**'s query images and target datasets is much smaller because each **SceneIBR2019** query image has additional color information which is directly related to the texture and color information existing in the 3D scene models.

Finally, we classify all the participating methods based on the techniques adopted: both two participating groups (Bui, Yuan) utilize local features, employ a deep learning framework to automatically learn the features, and apply regular transformations (e.g., flipping, translation, rotation). While, Bui further applies adversarial training as well. On the other hand, Yuan mainly adopts an image/sketch classification framework and then uses majority vote-based label matching to generate the retrieved result, while Bui conducts the retrieval based on both 2D sketch recognition, 3D model classification, and object detection and recognition as well.

6. Conclusions and Future Work

6.1. Conclusions

Since the semantic gap between the abstract and simplified 2D scene sketches queries and much more informative 3D scene model representations is huge, deep learning techniques have been proved their potentials in bridging the gap. Compared with SHREC'18 [YLL18a, YLL*18b], SHREC'19 contains 20 more categories such that we can evaluate the scalability of a 2D scene sketch-based 3D scene retrieval algorithm as well. Based on the experience and success of SHREC'12 [LSG*12, LLG*14], SHREC'13 [LLG*13, LLG*14], SHREC'14 [LLL*14, LLL*15], SHREC'16 [LLD*16] and SHREC'18 [YLL18a, YLL*18b] sketch-based 3D shape retrieval tracks that we organized in the past years, we believe this extended 2D scene sketch-based 3D scene model retrieval track can further promote this challenging and interesting research direction in the field of sketch-based 3D model retrieval.

Compared to 2D sketch-based 3D model retrieval, 2D scene sketch-based 3D scene model retrieval is much more challenging. However, for a even more challenging track this year, we still have two groups who have successfully participated in this track and contributed four runs of two methods. Through the platform provided by this track, it solicits current 2D scene sketch-based 3D scene retrieval approaches. We also wish that the **SceneSBR2019** benchmark and the results we obtained in this track will be a useful reference for researchers interested in this research area.

6.2. Future Work

For this interesting, challenging, and promising research topic, this track not only provides a platform for soliciting state-of-the-art methods, but also helps us identify the current challenges and future research directions.

- **Building a large 2D scene-based 3D scene retrieval benchmark in terms of number of categories and variations within each category.** Although our proposed **SceneSBR2019** has extended from ten scene classes in SHREC'18 to thirty classes, it is still far from a large-scale benchmark. This also explains why the top deep learning-based participating method has achieved excellent results. Due to the importance of the scalability to large-scale 2D scene sketch-based 3D scene retrieval and its corresponding applications, we are on the way to significantly enlarge the **SceneSBR2019** benchmark to build a large-scale benchmark for the community.
- **Build/search other more realistic 3D scenes models.** Some of the SketchUp 3D scene models that we downloaded from 3D Warehouse [Tri18] are not as realistic as relevant 2D scene images. For example, in the "mountain" category, the ratio between trees and mountains is not real, which could reduce the 3D scene retrieval accuracy. Due to this reason, a more realistic 3D scene dataset is also necessary.
- **2D scene sketch-based 3D scene retrieval by incorporating semantic information.** Since a scene is composed of one or more objects, the semantic information existing in 2D scene sketches and 3D scene models and the relationships between objects or between objects and related scenes are very useful for 3D scene retrieval. For instance, Bui's team utilized the known semantic information for data augmentation, e.g., they manually collected and added "camel" and "cactus" images to the "desert" category during training. They also employed object detection and recognition to refine their retrieval results. We believe that both the efficiency and accuracy will be further improved if the semantic information in both the 2D sketch queries and target 3D scene models is appropriately utilized.
- **Extend the feature vectors by incorporating the geolocation estimation features.** Photo geolocation estimation is to predict the GPS coordinates for a photo image. This information is helpful in classifying certain scene images. By classifying the earth's geographical cells based on deep learning, a recent work [MBPIE18] has shown that photo geolocalization

without any limitations can work to some extend reliably, even though with a small training dataset. Therefore, it is promising to achieve even better results by taking the scene's geographical information into account when forming a feature representation for the retrieval.

- **2D scene-based 3D scene retrieval related applications.** For instance, Disney World's Avatar Flight of Passage Ride [Wik18, Att18, tM18] is a 3D immersive program, which involves a lot of 3D scene contents. Other applications include retrieving indoor/outdoor scene candidates for cartoon or movie productions, such as automatically retrieving scenes from movies, computer games, and educational content by utilizing text and speech recognition to extract semantic scene information. This will help us build much larger benchmarks as well.
- **Deep learning models specifically designed for 3D scene retrieval.** From the submitted results' evaluation, we can find that deep learning techniques have great potential in achieving promising retrieved performance. However, both submitted methods adapted the existing neural network models designed for other purposes (e.g., objects classification), thus lacking considerations of the characteristics of this 2D sketche-based 3D scene retrieval problem. Therefore, it is promising to achieve even better retrieval result if we develop new deep learning models which fit this scenario well.

Acknowledgments

This project is supported by the University of Southern Mississippi Faculty Startup Funds Award to Dr. Bo Li and the Texas State Research Enhancement Program and NSF CR1-1305302 Awards to Dr. Yijuan Lu. We gratefully acknowledge the support from NVIDIA Corporation for the donation of the Titan X/Xp GPUs used in this research and anonymous content creators from the Internet.

References

[ARYLL19] ABDUL-RASHID H., YUAN J., LI B., LU Y.: SHREC'19 Extended 2D Scene Image-Based 3D Scene Retrieval Track Website. http://orca.st.usm.edu/~bli/SceneIBR2019/, 2019. 5

[Att18] ATTRACTIONS W.: New ride!!!! disney world animal kingdom: Avatar flight of passage ride video 4k hd video (pov). http://www.youtube.com/watch?v=f-cw7iCUY3c, 2018. 7

[DDS*09] DENG J., DONG W., SOCHER R., LI L.-J., LI K., FEI-FEI L.: ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09* (2009). 2, 3

[EHA12] EITZ M., HAYS J., ALEXA M.: How do humans sketch objects? *ACM Trans. Graph. (Proc. SIGGRAPH) 31*, 4 (2012), 44:1–44:10. 5

[HZRS16] HE K., ZHANG X., REN S., SUN J.: Deep residual learning for image recognition. In *CVPR* (2016), pp. 770–778. 3

[LLD*16] LI B., LU Y., DUAN F., DONG S., FAN Y., QIAN L., LAGA H., LI H., LI Y., LIU P., OVSJANIKOV M., TABIA H., YE Y., YIN H., XUE Z.: SHREC'16: 3D sketch-based 3D shape retrieval. In *3DOR 2016* (2016). 6

[LLG*13] LI B., LU Y., GODIL A., SCHRECK T., AONO M., JOHAN H., SAAVEDRA J. M., TASHIRO S.: SHREC'13 track: Large scale sketch-based 3D shape retrieval. In *3DOR* (2013), pp. 89–96. 6

[LLG*14] LI B., LU Y., GODIL A., SCHRECK T., BUSTOS B., FERREIRA A., FURUYA T., FONSECA M. J., JOHAN H., MATSUDA T., OHBUCHI R., PASCOAL P. B., SAAVEDRA J. M.: A comparison of methods for sketch-based 3D shape retrieval. *CVIU 119* (2014), 57–80. 3, 6

[LLL*14] LI B., LU Y., LI C., GODIL A., SCHRECK T., AONO M., BURTSCHER M., FU H., FURUYA T., JOHAN H., LIU J., OHBUCHI R., TATSUMA A., ZOU C.: SHREC'14 Track: extended large scale sketch-based 3D shape retrieval. In *3DOR* (2014), pp. 121–130. 6

[LLL*15] LI B., LU Y., LI C., GODIL A., SCHRECK T., AONO M., BURTSCHER M., CHEN Q., CHOWDHURY N. K., FANG B., FU H., FURUYA T., LI H., LIU J., JOHAN H., KOSAKA R., KOYANAGI H., OHBUCHI R., TATSUMA A., WAN Y., ZHANG C., ZOU C.: A comparison of 3D shape retrieval methods based on a large-scale benchmark supporting multimodal queries. *CVIU 131* (2015), 1–27. 3, 6

[LSG*12] LI B., SCHRECK T., GODIL A., ALEXA M., BOUBEKEUR T., BUSTOS B., CHEN J., EITZ M., FURUYA T., HILDEBRAND K., HUANG S., JOHAN H., KUIJPER A., OHBUCHI R., RICHTER R., SAAVEDRA J. M., SCHERER M., YANAGIMACHI T., YOON G.-J., YOON S. M.: SHREC'12 track: Sketch-based 3D shape retrieval. In *3DOR* (2012), pp. 109–118. 6

[MBPIE18] MÜLLER-BUDACK E., PUSTU-IREN K., EWERTH R.: Geolocation estimation of photos using a hierarchical model and scene classification. In *Computer Vision – ECCV 2018* (Cham, 2018), Springer International Publishing, pp. 575–592. 6

[RHGS15] REN S., HE K., GIRSHICK R. B., SUN J.: Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR abs/1506.01497* (2015). 4

[tM18] THE MAGIC I.: New flight of passage ride queue, preshow in pandora - the world of avatar at walt disney world. http://www.youtube.com/watch?v=eM8f47Igtu8, 2018. 7

[Tri18] TRIMBLE I.: 3D Warehouse. http://3dwarehouse.sketchup.com/?hl=en, 2018. 2, 6

[Wik18] WIKIPEDIA: Avatar flight of passage. http://en.wikipedia.org/wiki/Avatar_Flight_of_Passage, 2018. [Online; accessed 1-March-2018]. 7

[XEH*16] XIAO J., EHINGER K. A., HAYS J., TORRALBA A., OLIVA A.: Sun database: Exploring a large collection of scene categories. *Int. J. Comput. Vision 119*, 1 (Aug. 2016), 3–22. 2

[YARLL19a] YUAN J., ABDUL-RASHID H., LI B., LU Y.: SHREC'19 Extended 2D Scene Sketch-Based 3D Scene Retrieval Track Website. http://orca.st.usm.edu/~bli/SceneSBR2019/, 2019. 5

[YARLL19b] YUAN J., ABDUL-RASHID H., LI B., LU Y.: Sketch/image-based 3D scene retrieval: Benchmark, algorithm, evaluation. In *MIPR* (2019), IEEE. 2, 4, 5

[YLL16] YE Y., LI B., LU Y.: 3D sketch-based 3D model retrieval with convolutional neural network. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (Dec 2016), pp. 2936–2941. 4

[YLL18a] YUAN J., LI B., LU Y.: SHREC'18 2D Scene Sketch-Based 3D Scene Retrieval Track Website. http://orca.st.usm.edu/~bli/SceneSBR2018/, 2018. 2, 4, 6

[YLL*18b] YUAN J., LI B., LU Y., BAI S., BAI X., BUI N., DO M. N., DO T., DUONG A. D., HE X., LE T., LI W., LIU A., LIU X., NGUYEN K., NGUYEN V., NIE W., NINH V., SU Y., TON-THAT V., TRAN M., XIANG S., ZHOU H., ZHOU Y., ZHOU Z.: 2D scene sketch-based 3D scene retrieval. In *Eurographics Workshop on 3D Object Retrieval, 3DOR 2018, 16 April 2018, Delft, The Netherlands.* (2018), pp. 29–36. 2, 4, 6

[ZLK*17] ZHOU B., LAPEDRIZA A., KHOSLA A., OLIVA A., TORRALBA A.: Places: a 10 million image database for scene recognition. *IEEE Trans. on PAMI* (2017). 2, 3, 5