Sketch/Image-Based 3D Scene Retrieval: Benchmark, Algorithm, Evaluation

Hameed Abdul-Rashid[1], Juefei Yuan[1], Bo Li[1], Yijuan Lu[2]
[1]University of Southern Mississippi, [2]Texas State University
Presenter: Yijuan Lu

Outline

- Introduction → besides colored, also using bold font. so sf.
- Related Work
- Benchmark
- Method
- Evaluation
- Conclusions and Future Work

Insert slide number

Introduction

- **2D Scene Sketch/Image-Based 3D Scene Retrieval (Scene_SBR_IBR)** focuses on retrieving relevant 3D scene models using scene sketches/image(s) as input
- **Motivation:**
  - Vast applications: 3D scene reconstruction, autonomous driving cars, 3D geometry video retrieval, and 3D AR/VR Entertainment
- **Challenges**
  - 2D sketches/images lack 3D scene information they are supposed to present
  - Semantic gap between 2D scene iconic sketches or realistic images and accurate 3D scene models

2D Scene Sketch/Image-based 3D Scene retrieval (Scene_SBR_IBR) focuses on retrieving relevant 3D scene models using scene sketch(es)/image(s) as input.

The Motivation of the Scene_SBR_IBR is that:

- It has vast applications such as 3D scene reconstruction, autonomous driving cars, 3D geometry video retrieval, and 3D AR/VR Entertainment

Introduction

- **Challenges contd.**
  - Brand new research topic in the field of sketch/image-based 3D object retrieval (Scene_SBR_IBR)
    - ✓ A query sketch/image contains several objects
    - ✓ Objects may overlap with each other
    - ✓ Relative context configurations among the objects
- To **promote** this challenging research direction, **we** built the most comprehensive and largest 2D scene sketch/image-based benchmark 3D scene retrieval benchmark, **Scene_SBR_IBR.**

But there are some existing challenges with 2D sketch/image-based 3D Scene Retrieval, which are:

- Firstly, 2D sketches/images lack 3D scene information they are supposed represent
- Secondly, there is still a semantic gap between 2D scene sketches/images and accurate 3D scene models
- Finally, it is a brand new research topic in the field of sketch/image-based 3D object retrieval:
  - ✓ A query image contains several objects
  - ✓ Objects may overlap with each other
  - ✓ There existing relative context configurations among the objects in a scene image/model

Considering the above, we built the most comprehensive and largest 2D scene sketch/image-based benchmark 3D scene retrieval benchmark, **Scene_SBR_IBR.**

Outline

- Introduction
- Related Work
- Benchmark
- Method
- Evaluation
- Conclusions and Future Work

Related Work

- 3D Scene Retrieval
  - Fisher and Hanrahan proposed context-based 3D model retrieval [1]
  - Xu et. al proposed Sketch2Scene, a system for automatic 2D sketch-based 3D scene composition [2]

delete a space

insert → two sub items

✓ 3D box

✓ dimensionality & context information

insert → two sub items

✓ functional & spatial relationships

✓ using structural groups

[1] M. Fisher and P. Hanrahan. Context-based search for 3D models. ACM Trans. Graph., 29:182:1–182:10, 2011.

[2] B. Li and et al. A comparison of 3D shape retrieval methods based on a large-scale benchmark supporting multimodal queries. Computer Vision and Image Understanding, 131:1–27, 2015.

Fisher and Hanrahan proposed context-based 3D model retrieval, which retrieves models according to their spatial context in a 3D scene. They first locate the position of the model by drawing a 3D box and then searching relevant 3D models based on the dimensionality and context information.

Xu et. al proposed Sketch2Scene for automatic 2D sketch-based 3D scene composition by representing 3D scene objects' functional and spatial relationships based on structural groups.

Related Work

- 2D/3D Scene Datasets
  - Xiao et. al built Scene UNderstanding (SUN) datasets
    - 130,519 images across 899 scene categories [3]
    - Expanded to 908 classes [4]
  - Xiao et. al created SUN3D [5]
    - RGB-D video database with camera pose and object labels

delete the bars

[3] - J. Xiao and et al. SUN database: Large-scale scene recognition from abbey to zoo. In CVPR, pages 3485–3492. IEEE Computer Society, 2010.

[4] - J. Xiao and et al. SUN database: Exploring a large collection of scene categories. International Journal of Computer Vision, 119(1):3–22, 2016.

[5] - J. Xiao and et al. SUN3D: A database of big spaces reconstructed using SfM and object labels. In ICCV, pages 1625–1632, 2013.

Xiao et. al built SUN and introduced 130,519 images across 899 scene categories. This was later extended to include 908 scenes classes.

Xiao et. al also built SUN3D, ~~is~~ a RGB-D video database that captures the full extent of 3D scenes with camera pose and object labels. ~~Videos were used for partial 3D reconstruction and propagated labels between frames. Labels were then used to refine the final partial reconstruction.~~

Related Work

- 2D/3D Scene Datasets contd.
  - Song et. al ~~constructed~~ SUNCG [6]   synthetic
    - 46,622 Synthetic 3D scenes with 2,644 objects
    - ~~across~~ 84 scene categories
  - Zhou et. al compiled Places   [7]
    - 10,624,928 images ~~across~~ 434 scene categories

delete the box

[6] - S. Song and et al. Semantic scene completion from a single depth image. In CVPR, pages 190–198. IEEE Computer Society, 2017.

[7] → ?

Song et. al constructed SUNCG, a database of synthetic 3D scenes with ~~manually labelled voxel occupancy and semantic labels~~. SUNCG is comprised of 45,622 scenes and 2,644 objects across 84 scene categories.

Zhou et. al compiled Places, a database of 10,624,928 scene images across 434 scene categories. While Places does not provide annotations at the object level, it provides the most diverse scene composition.

Outline

- Introduction
- Related Work
- Benchmark
- Method
- Evaluation
- Conclusions and Future Work

① Insert an overview slide first. (first 3 paragr of Section 3) See my previous comments to your version 1 on the bottom right of page 1.

② also mention it supports both sketch-based and image-based 3D scene retrieval

Scene_SBR_IBR Benchmark (1/3)

- 2D Scene Sketch Query Dataset
  - SceneSBR ~~SBR benchmark [7], augmented further with 500 sketches from Flickr and Google Images~~ put here
  - 750 2D scene sketches
  - 30 classes ~~(25 sketches per class)~~ delete

[7] J. Yuan and et al. SHREC'18 track: 2D scene sketch-based 3D scene retrieval. In 3DOR, pages 1–8, 2018

The 2D Scene Sketch Query Dataset ~~utilizes the 2D scene images in SBR benchmark further extends with sketches from Flickr and Google images.~~ ed

- ~~**The 2D Scene Sketch Query Dataset**~~ contains **750** 2D scene images categorized into 30 classes, each with 25 sketches each

Scene_SBR_IBR Benchmark (1/3)

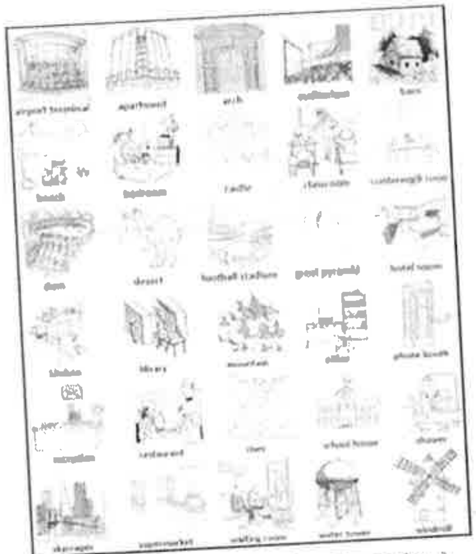


Fig. 1 Example 2D sketches (1 per class)

Scene_SBR_IBR Benchmark (2/3)

- 2D Scene Image Query Dataset
  - ~~Places88 [8]~~ put in the previous overview slide
  - 30,000 images
  - 30 classes (1,000 images per class)

[8] - B. Zhou and et al. Places: A 10 million image database for scene recognition. IEEE Trans. Pattern Anal. Mach. Intell., 40(6):1452–1464, 2018

The 2D Scene Image Query Dataset util~~izes the 2D scene images in Places88 as its 2D scene image dataset~~

- **The Places88 scene images contain** 30,000 2D scene images categorized into 30 classes, each with 1,000 images each

[some are not coming from places.
we only use categories information of places]

Scene_SBR_IBR Benchmark (2/3)

Fig. 1 Example 2D scenes (1 per class)

Scene_SBR_IBR Benchmark (3/3)

- 3D Scene Model Target Dataset
  - ~~3D Warehouse~~ [9] Same (overview) slide
  - 3000 scene models
  - 30 classes (100 models per class)

[9] - 3D Warehouse. http://3dwarehouse.sketchup.com/?hl=en. 2018.

○**The 3D scene dataset contains** 3000 3D scene models ~~collected~~ from the 3D Warehouse. ~~Similarly,~~ they are categorized into the same 30 classes, ~~but~~ with each having 100 models.

Scene_SBR_IBR Benchmark (3/3)

Fig. 1 Example 3D scenes (1 per class)

+ Scene-SBR-IBR
- Scene-SBR: Subsets 1 & 3 sketch-based 3D scene retrieval
- Scene-IBR: Subsets 2 & 3 image-based 3D scene retrieval

+ Table 1.

Testing ← for learning based
Training

Evaluation

- **Seven** commonly adopted performance metrics in 3D model retrieval techniques [10, 11]:
  - Precision-Recall plot (PR)
  - Nearest Neighbor (NN)
  - First Tier (FT)
  - Second Tier (ST)
  - E-Measures (E)
  - Discounted Cumulated Gain (DCG)
  - Average Precision (AP)
- We also have developed the code to compute them
  - http://orca.st.usm.edu/~bli/Scene_SBR_IBR/data.html

[10] - H. Abdul-Rashid and et al. SHREC'18 track: 2D scene image-based 3D scene retrieval. In 3DOR, pages 1–8, 2018.
[11] - J. Yuan and et al. SHREC'18 track: 2D scene sketch-based 3D scene retrieval. In 3DOR, pages 1–8, 2018.

There are seven commonly adopted performance metrics in 3D model retrieval techniques, which are PR, NN, FT, ST, E, DCG and AP.
We also have developed the code to compute them, and the code can be downloaded from the provided link.

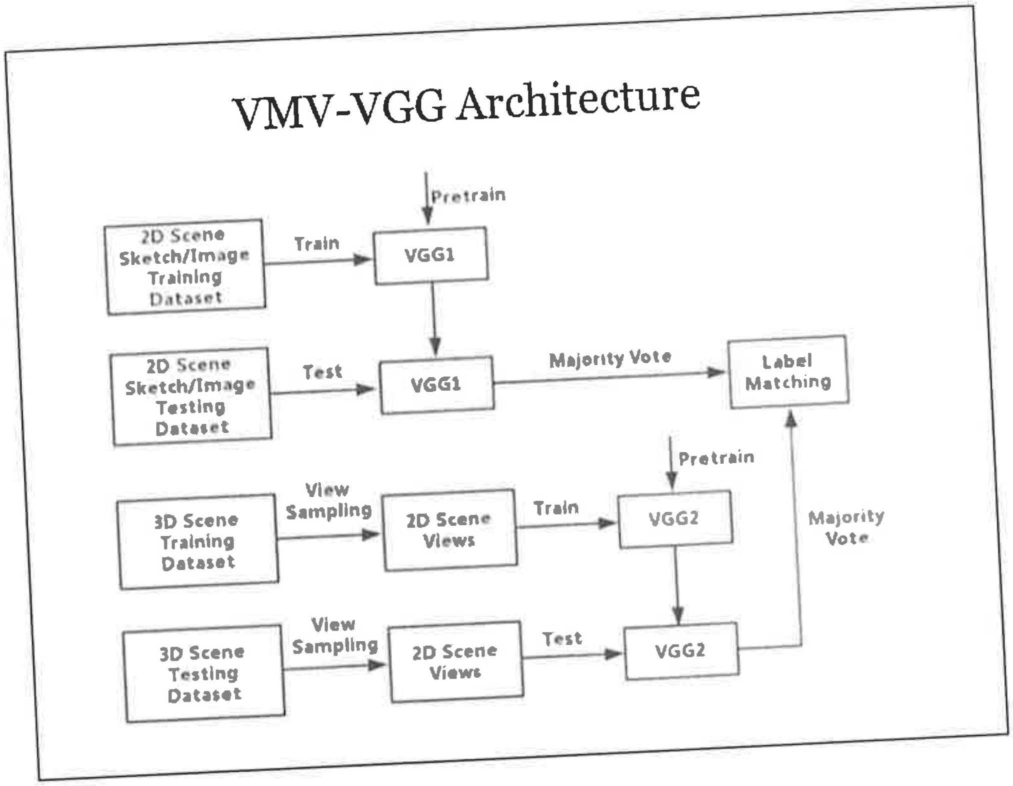
Outline

- Introduction
- Related Work
- Benchmark
- Method
- Evaluation
- Conclusions and Future Work

Our Retrieval Algorithm VMV-VGG

Create an overview slide based on the 1st paragraph of Section 4.

Revised by Jueifei since here



VMV-VGG Architecture
Pretrain
2D Scene Sketch/Image Training Dataset
Train
VGG1
2D Scene Sketch/Image Testing Dataset
Test
VGG1
Majority Vote
Label Matching
Pretrain
3D Scene Training Dataset
View Sampling
2D Scene Views
Train
VGG2
Majority Vote
3D Scene Testing Dataset
View Sampling
2D Scene Views
Test
VGG2

VMV-VGG

- Incorporates two different VGG-16 based models [12] (VGG1 and VGG2)
- (1) 3D Scene view sampling
- (2) Data Augmentation on each training Batch
- (3) Pre-Training and Training on VGG1 and VGG2
- (4) Fine-tuning
- (5) Sketch/Image/View Scene Classification
- (6) Majority vote-based label matching

font → revise

[12] - K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014.

- (1) Scene view sampling
  1. Automate sample through QMacro script
  2. Uniformly sample 12 views along the equator of the sphere and 1 top-down view, for 13 views in total.
- (2) Data Augmentation
  1. Perform random rotations, reflections or translations to augment each batch size per epoch.
- (3) Pre-Training and Training on VGG1 and VGG2

Simplify sentences

sosf.

i. For the sketch-based retrieval
   1. VGG1 is pre-trained only on the TU-Berlin dataset [7] for 500 epochs
   2. VGG2 is pre-trained on only the Places data set for just 100 epochs

ii. For the image-based retrieval
   1. VGG1 and VGG2 is pre-trained on only the Places data set for just 100 epochs

iii. For Training
   1. VGG1 is trained on the 2D Scene Sketch/Image Query Dataset for 100 epochs
   2. VGG2 is trained on 2D views of the 3D scene models for 50 epochs

- (4) Fine-tuning
   i. Fine-tune the pre-trained VGG1/VGG2 models each 100/50 epochs
- 5) Sketch/Image/View Classification
   i. We feed the well-trained model (VGG1/VGG2) alongside its corresponding testing query sketch/image or target scene view to obtain two classification vectors.

- (6) Majority vote-based label matching
    i. We generate a rank list for each query by using a majority vote-based label matching method based on the query's classification vector and the target 3D scene's 13 classification vectors.

Show figure 5. in a slide.

Outline

- Introduction
- Related Work
- Benchmark
- Method
- Evaluation
- Conclusions and Future Work

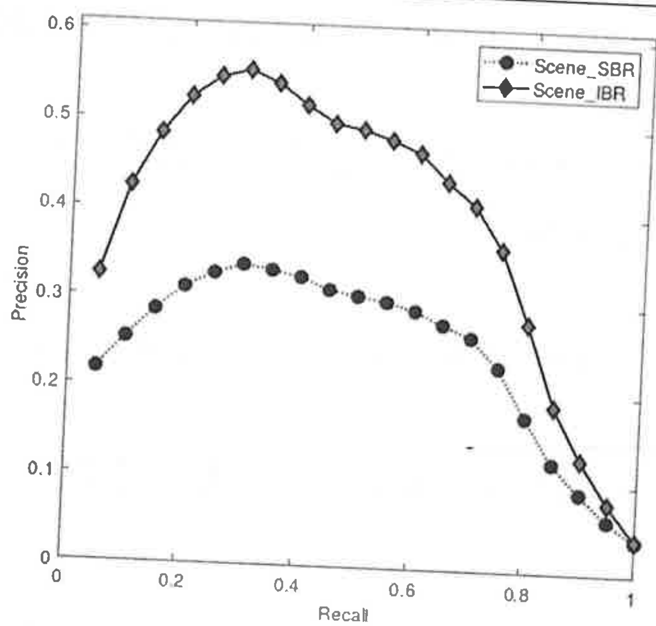


Figure 6. Precision-Recall diagram performance of our VMV-VGG on our Scene SBR IBR benchmark.

We have found that compared with the performance that has been achieved in our two SHREC'18 tracks [21, 2] which used a much smaller benchmark containing only 10 classes, in contrast to the current 30 classes available in our new benchmark, the overall performance dropped significantly for either type of retrieval. For example, Li's MMD-VGG method, which also utilizes VGG, has achieved an excellent overall performance in terms of DCG (0.856) or AP (0.685) on the

SceneSBR benchmark, while they drop to DCG (0.533) and AP (0.244) respectively based on our VMV-VGG.

Results: Performance Metrics

Benchmark	NN	FT	ST	E	DCG	AP
Scene_SBR	0.081	0.281	0.369	0.280	0.533	0.244
Scene_IBR	0.122	0.458	0.573	0.452	0.644	0.392

Performance metrics generated by running our VMV-VGG on our Scene SBR IBR benchmark.

This performance decrease should be a direct and natural result after a substantial increase in the comprehensiveness and challenge level that exist in Scene SBR IBR after we incorporate much more scene categories. The addition of more classes will cause more ambiguities during the retrieval process and the retrieval algorithm may fail to properly distinguish between classes that share certain similarities.

Add a slide based on Section 5, 2nd paragraph, to illustrate your findings.

Outline

- Introduction
- Related Work
- Benchmark
- Method
- Evaluation
- Conclusions and Future Work

Conclusions & Future Work

- **Objective:** To foster this challenging and interesting research direction: *Scene Sketch/Image-Based 3D Scene Retrieval*
- **Dataset:** Build ***the current largest*** 2D Scene sketch/image 3D scene retrieval benchmark
- **Method:** Baseline performance has been provided by VMV-VGG
- **Evaluation:** Performed a *comparative evaluation* on the accuracy

Sketch/Image-based 3D scene retrieval are research topics with a lot of ap-
plication potentials. There is extremely limited preliminary work in this field, which allows us to explore many promis-
ing ideas and interesting results. In this paper, the currently

largest 3D scene retrieval benchmark Scene SBR IBR is

proposed with the hope to advance this research direction. To assist other interested researchers, the baseline

performance on the benchmark has been provided by conducting evaluation based on a proposed CNN classifier-based 3D scene retrieval algorithm VMV-VGG. Our future goals include: (1) building a large-scale and/or multimodal 2D scene sketch/image-based 3D scene retrieval

benchmark; (2) semantics-driven 2D scene sketch/image-based 3D scene retrieval.

Conclusions & Future Work

- Build a large-scale and/or multimodal 2D scene-based 3D scene retrieval benchmark
- Semantics-driven 2D scene image-based 3D scene retrieval
- **Impact:** Provided *the largest and most comprehensive common platform* for evaluating 2D scene sketch/image-based 3D scene retrieval
- Build a large-scale and/or multimodal 2D scene-based 3D scene retrieval benchmark
- Semantics-driven 2D scene image-based 3D scene retrieval

Sketch/Image-based 3D scene retrieval are research topics with a lot of ap-
plication potentials. There is extremely limited preliminary work in this field, which allows us to explore many promis-
ing ideas and interesting results. In this paper, the currently

largest 3D scene retrieval benchmark Scene SBR IBR is

proposed with the hope to advance this research direction. To assist other interested researchers, the baseline

performance on the benchmark has been provided by conducting evaluation based on a proposed CNN classifier-based 3D scene retrieval algorithm VMV-VGG. Our future goals include: (1) building a large-scale and/or multimodal 2D scene sketch/image-based 3D scene retrieval

benchmark; (2) semantics-driven 2D scene sketch/image-based 3D scene retrieval.

References

- [1] - M. Fisher and P. Hanrahan. Context-based search for 3D models. ACM Trans. Graph., 29:182:1–182:10, 2011.
- [2] - B. Li and et al. A comparison of 3D shape retrieval methods based on a large-scale benchmark supporting multimodal queries. Computer Vision and Image Understanding, 131:1– 27, 2015.
- [3] - J. Xiao and et al. SUN database: Large-scale scene recognition from abbey to zoo. In CVPR, pages 3485–3492. IEEE Computer Society, 2010.
- [4] - J. Xiao and et al. SUN database: Exploring a large collection of scene categories. International Journal of Computer Vision, 119(1):3–22, 2016.
- [5] - J. Xiao and et al. SUN3D: A database of big spaces reconstructed using SfM and object labels. In ICCV, pages 1625–1632, 2013.
- [6] - S. Song and et al. Semantic scene completion from a single depth image. In CVPR, pages 190–198. IEEE Computer Society, 2017.
- [7] - J. Yuan and et al. SHREC'18 track: 2D scene sketch-based 3D scene retrieval. In 3DOR, pages 1–8, 2018
- [8] - B. Zhou and et al. Places: A 10 million image database for scene recognition. IEEE Trans. Pattern Anal. Mach. Intell., 40(6):1452–1464, 2018
- [9] - 3D Warehouse. http://3dwarehouse.sketchup. com/?hl=en. 2018.
- [10] - H. Abdul-Rashid and et al. SHREC'18 track: 2D scene image-based 3D scene retrieval. In 3DOR, pages 1–8, 2018.
- [11] - J. Yuan and et al. SHREC'18 track: 2D scene sketch-based 3D scene retrieval. In 3DOR, pages 1–8, 2018.
- [12] - K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014.

Thank you!

Q&A?

E-mail: bo.li@usm.edu