DiffPano: Scalable and Consistent Text to Panorama Generation with Spherical Epipolar-Aware Diffusion

Weicai Ye, Chenhao Ji, Zheng Chen, Junyao Gao, Xiaoshui Huang,

Song-Hai Zhang, Wanli Ouyang, Tong He, Cairong Zhao, Guofeng Zhang

Problem Definition

- Input:
  - Text Prompt

  A well-furnished bedroom with beige walls, a white bed, and a white wardrobe. The floor is carpeted. The room also features a painting on the wall and a white door. The ceiling is tan, and the carpet is soft and comfortable.

  - Camera Trajectory
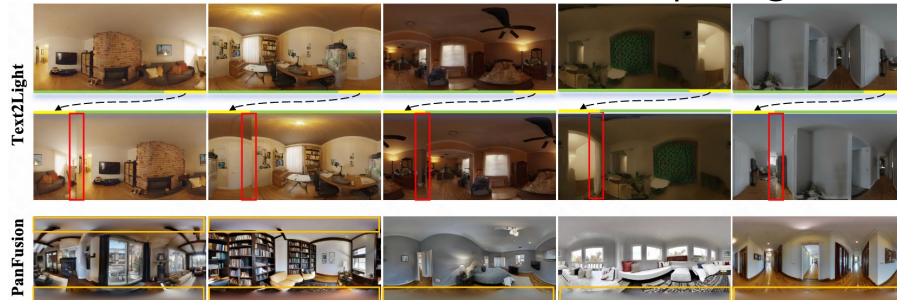

  Camera Trajectory

- Output: Consistent Multi-view Panoramas

Motivation

- Single-View Panorama Generation
  - Text2Light: poor left-right consistency
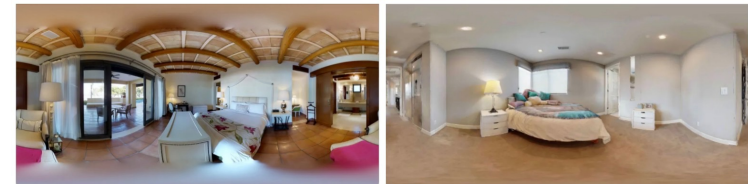 - PanFusion: bottom and top region blur

- Multi-View Panorama Generation
  - Inpainting solution lacks 3D awareness.
  - Extending MVDream to panorama: lack of multi-view consistency

- Panorama Dataset
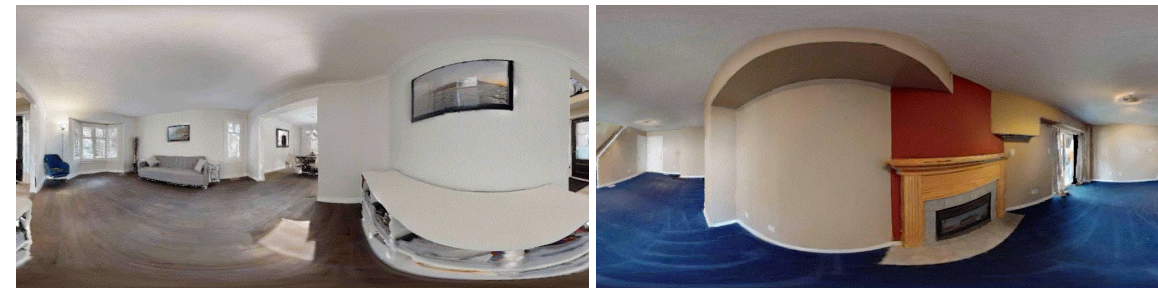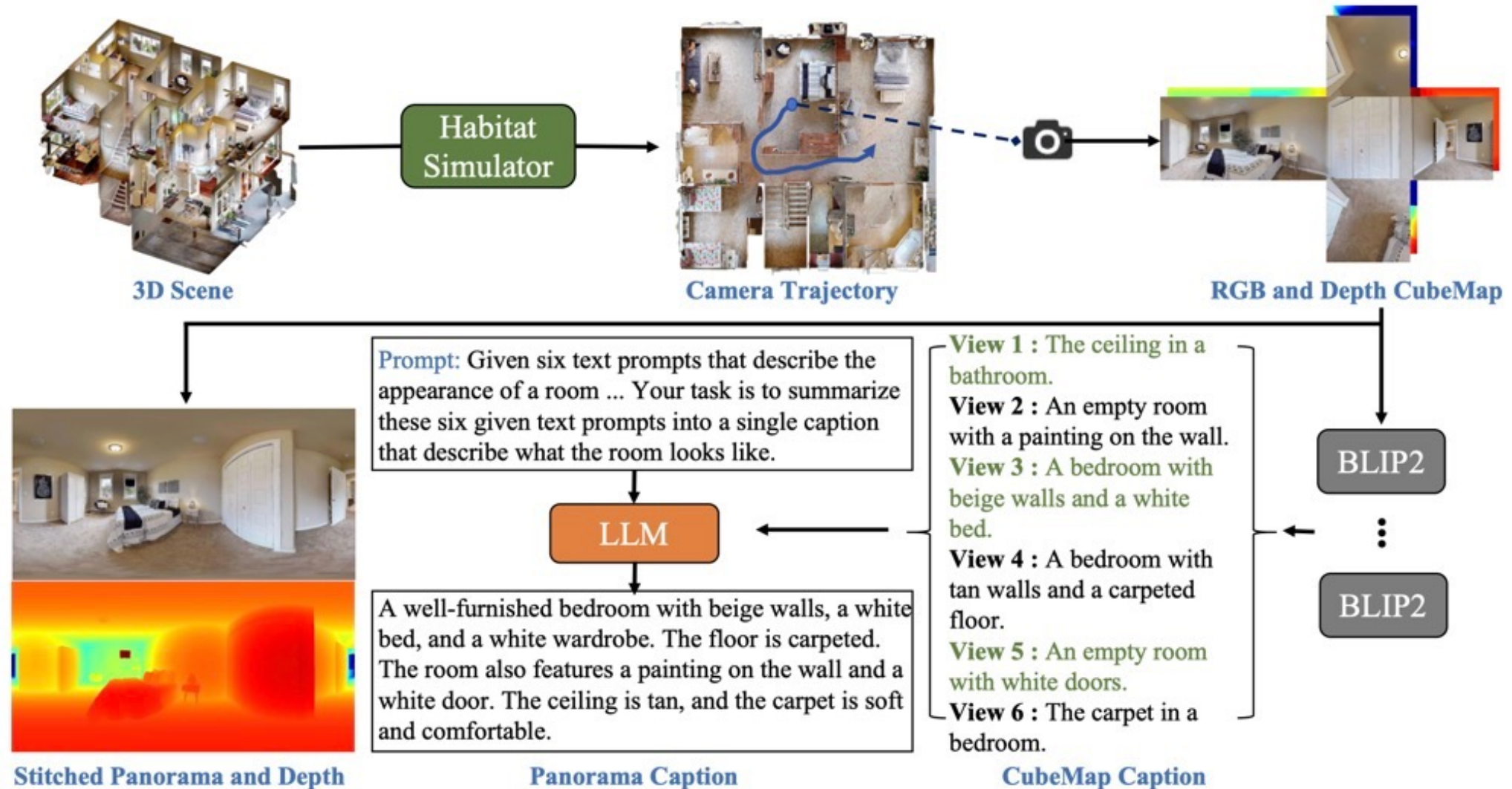 - PanFusion's dataset: blur, single-view; concise text, only four or five words.

PanFusion dataset: "a bedroom in a hotel" Our datset: "A well-lit bedroom with hardwood floors, a large bed, and a white dresser, surrounded by a tan carpet. The room is spacious and has a hallway leading to other parts of the home."

  - Lack of multi-view panorama dataset.
    - Our Panoramic Video-Text dataset

Panoramic Video-Text Dataset Pipeline

3D Scene → Habitat Simulator → **Camera Trajectory** → 📷 → **RGB and Depth CubeMap**

Prompt: Given six text prompts that describe the appearance of a room ... Your task is to summarize these six given text prompts into a single caption that describe what the room looks like.

LLM

A well-furnished bedroom with beige walls, a white bed, and a white wardrobe. The floor is carpeted. The room also features a painting on the wall and a white door. The ceiling is tan, and the carpet is soft and comfortable.

View 1 : The ceiling in a bathroom.
View 2 : An empty room with a painting on the wall.
View 3 : A bedroom with beige walls and a white bed.
View 4 : A bedroom with tan walls and a carpeted floor.
View 5 : An empty room with white doors.
View 6 : The carpet in a bedroom.

BLIP2
⋮
BLIP2

Stitched Panorama and Depth   **Panorama Caption**   **CubeMap Caption**

Method

Text Prompts $P_{1:N}$

Camera Poses $R, T_{1:N}$

Text Prompts: A hallway with hardwood floors, leading to a home office with a large desk and bookshelves. A home office with a large desk, bookshelves, hardwood floors and a comfortable chair.

Noise

Text Prompt

Noise $X_{1:N}$

E

D

$X_{1:N}$

Denoising UNet

Stable Diffusion

LoRA

Generated Panorama

Epipolar Line Computation

Target View

Source View

Spherical Epipolar-Aware Module

Text to Panorama

Text to MV Panoramas

Single-View Pano-based SD

Spherical Epipolar-Aware Multi-View Diffusion

Applications

Text to Single-View Panorama

"A living room with hardwood floors, a fireplace, and large windows."

"A cozy study with built-in bookshelves and a leather armchair."

"A bedroom with a ceiling fan, gray walls, hardwood floors, a bed, and a TV on the wall."

"A cozy built-in daybed with throw pillows and blankets."

"A spacious corridor with hardwood floors."

Text to Multi-View Panorama

"A peaceful sunroom with a daybed, plants, and floor-to-ceiling windows."

Qualitative Results

"A remodeled garage with a door open, displaying a cluttered space filled with various items on the walls. A well-lit hallway with a door and various furniture."

"A living room with white walls and a white door. The room is furnished with red leather couches and a TV mounted on the wall. A spacious room with hardwood floors, a kitchen and dining area, and an old-fashioned touch."

"A cozy room with wooden floors and a sliding glass door. The room is well-lit with a sink and wooden furniture. A cozy bedroom with wooden floors and a large bed. The room is furnished with a rug."

"A living area with a high ceiling, a light fixture, and a kitchenette. The room is furnished with a dining table and chairs and a painting on the wall, featuring a living room and dining room combination, and a white couch."

Qualitative Results

"A spacious house with wooden paneling features a stairway, and multiple doors leading to various rooms. A cozy wooden house with a stairway leading up to the second floor, surrounded by wood floors and adorned with a picture on the wall."

"A kitchen with white ceiling, hardwood floors, black cabinets and stainless steel appliances. A kitchen with black cabinets, hardwood floors, and a stove, connected to a another room with wood floors."

"A spacious living room with a brown couch, a TV, and a desk. A well-furnished room with a painting, and an open door, featuring a couch, desk, and TV on the wall, with a carpet."

"A bedroom with hardwood floors, a ceiling fan, a bed, and a TV on the wall, featuring white wall and brown floor. A two-story house with a small stairway, hardwood floors, and an open doorway leading to a bedroom."

Qualitative Results

"A home office with a large desk, bookshelves, hardwood floors and a comfortable chair. A hallway with hardwood floors, leading to a home office with a large desk and bookshelves."

"A bedroom with a bed, and a dresser, featuring a window with curtains and blinds and a TV on top of the dresser. A cozy bedroom with a bed, dresser, and TV, surrounded by a brown carpet."

"A cozy living room with a purple couch, a fireplace, and a rug. The ceiling has a light on it. A living room with a white staircase, wooden floors, a couch, and an open door that leads to another room."

"A small room with a wooden door and a bed, featuring a white wall with a picture. A spacious room with a carpeted floor, a hallway leading to a bedroom with wooden doors and a comfortable bed."

Text to Panoramic Video

Summary

- Propose large-scale panoramic video-text dataset
- Propose panorama generation framework with spherical epipolar-aware attention for text-to-single-view or multi-view panorama generation.
- Achieve scalable, consistent, and diverse multi-view panorama generation