Media to 3D

Работу выполнил: Кочетков Николай

Описание задачи

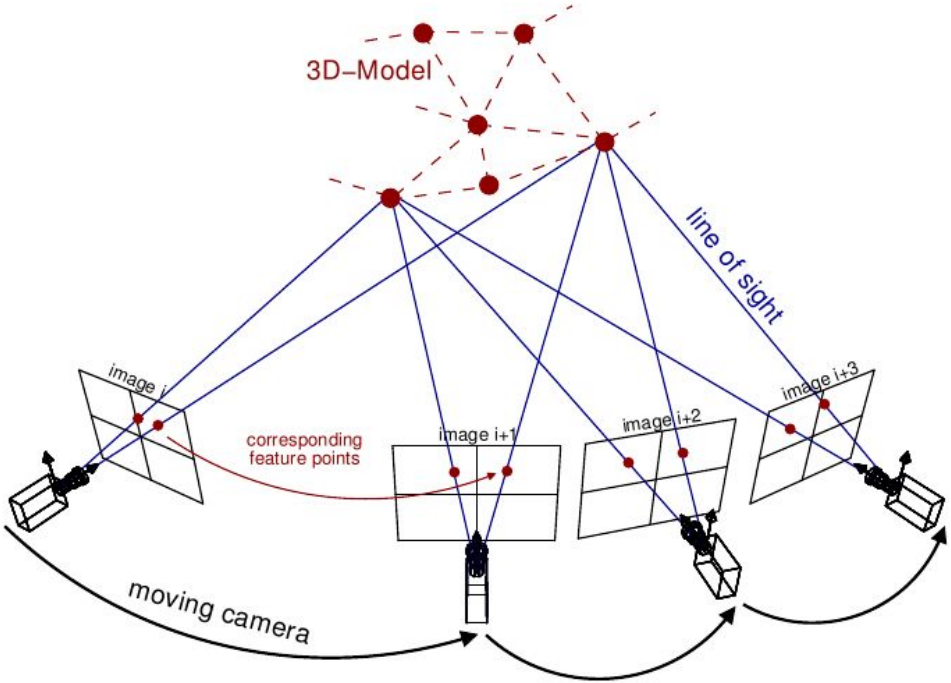
- ~~Реализовать сервис, который из текстового описания объекта генерирует его 3D представление (mesh)~~
- Реализовать сервис, которые способен восстановить 3D представление объекта из видео или фото.
- Результатом работы сервиса сделать видео или 3D Mesh (по выбору пользователя)

Краткий результат EDA

Восстановление 3D сцен по фотографиям или видео далеко не новая тема. Существует много классических подходов среди которых:

- Structure-from-motion
- Structured light
- Lidars, Radars
- И менее классческий NERF

Structure-from-motion (Sfm)



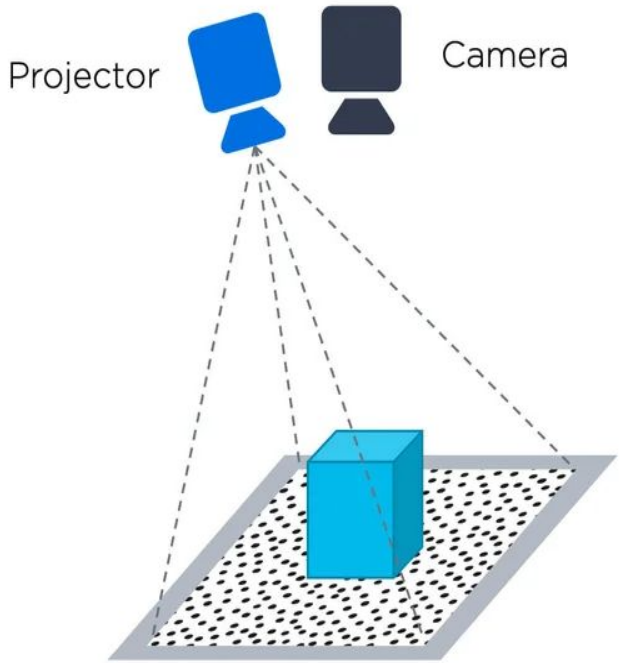
Приемущества

- Точная реконструкция

Недостатки:

- Качество 3D-модели сильно зависит от количества и качества исходных изображений

Structured light



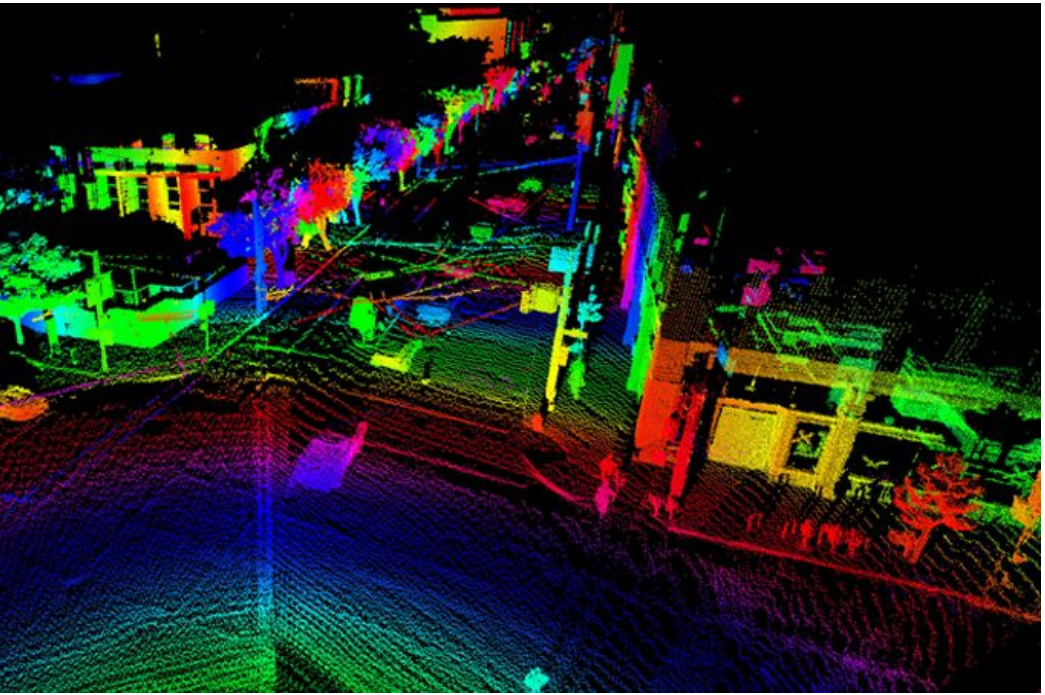
Приемущества:

- Высокая точность и детализация для фиксированных и контролируемых сцен.

Недостатки:

- Сложное оборудование
- Требует точной калибровки
- Нельзя попросить пользователей сделать подобные кадры

Lidars and Radars



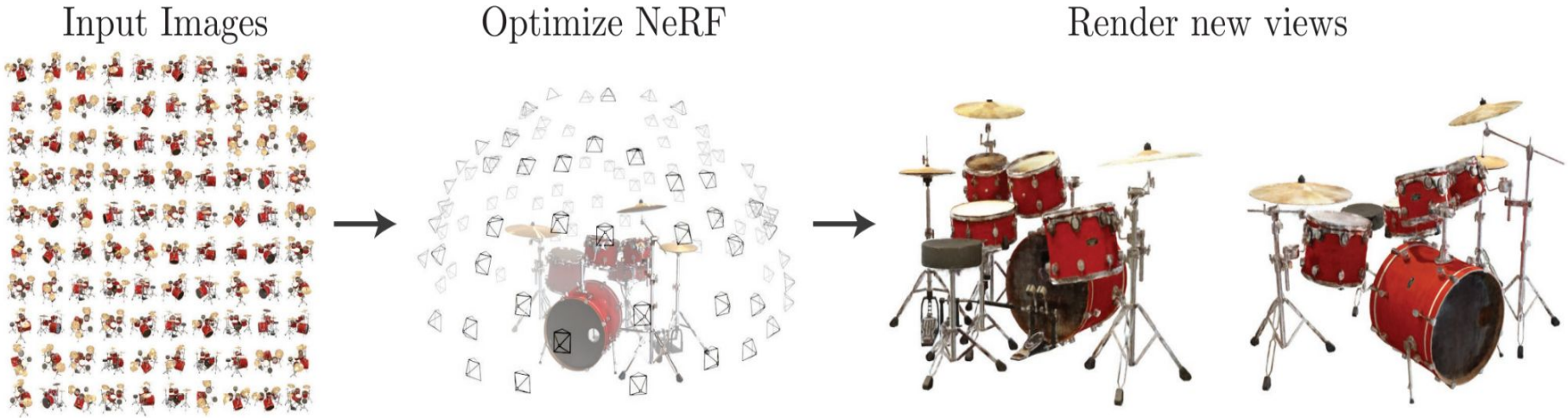
Приемущества:

- Точная реконструкция

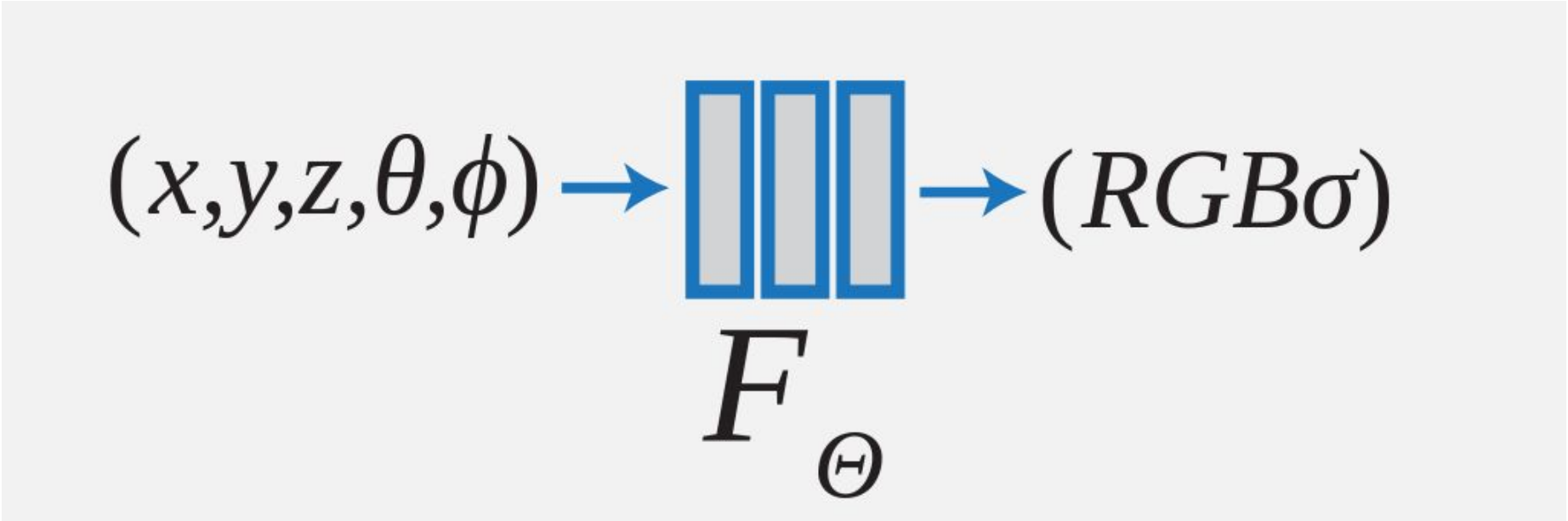
Недостатки:

- Дорогое оборудование
- Отклонение от задачи, так как мы планируем работать только с фотографиями или видео

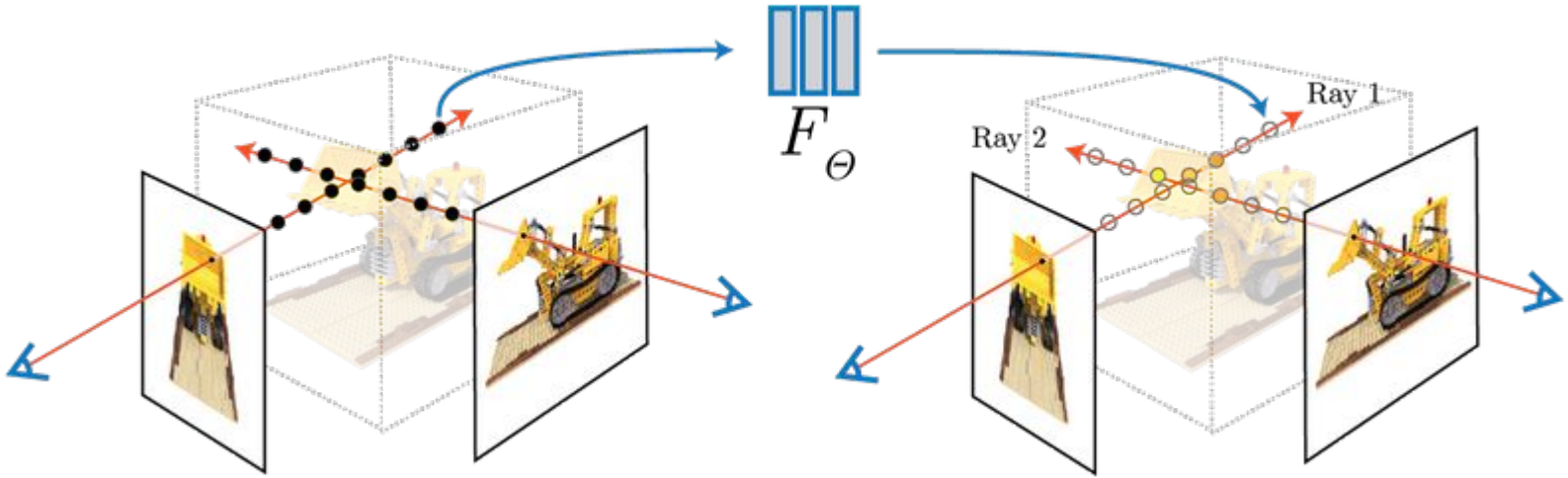
Neural Radiance Fields (NERF)



Neural Radiance Fields (NERF)



Neural Radiance Fields (NERF)



$$w_i = T_i(1 - \exp(-\sigma_i \delta_i)) \,. \qquad \hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i(1 - \exp(-\sigma_i \delta_i))\mathbf{c}_i \,, \text{ where } T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right)$$

Blend of Sfm and NeRF (From NeRF paper)

В результате было решено двигаться с Sfm и Nerf-ом. Sfm будет служить для оценки позиций камеры а Nerf будет учиться рендерить объект с вычисленных позиций

- В качестве Sfm был использован colmap: https://colmap.github.io/install.html
- NeRF был реализован мной (подглядывая куда только можно)

Сбор данных

- Данные будут приходить в виде фотографий или видео от пользователя.
- Из видео будет извлечено N кадров.
- Полученные фотографии пройдут sfm процессинг для оценки позиций камер
- Полученный позиции камер вместе с фотографиями будут использованы как входные данные в Nerf
- Также для валидации работы сети, был взят общедоступный датасет с синтетическими данными из статьи: https://www.matthewtancik.com/nerf

Технические детали

- Модель реализована с помощью pytorch
- Метрика качества - psnr (Peak Signal To Noise Ratio)
- Обучение модели производилось с помощью lightning
- Оптимизатор Adam с lr=5e-3, betas=(0.9, 0.999)

Что получилось

- Получились генерации на синтетических данных
- Получилось организовать полный конвейер обучения на реальных данных и возвращение видео пользователю.
- Реализованы чекпоинты, обучение производится на видеокарте
- Создан телеграм бот для работы с моделью
- Бот контейнизирован в docker container
- pre-commit, argparse

Что не получилось, что не доделал

- Сгенерировать качественное видео из реальных данных
- Сгенерировать Mesh
- Реализовать Fine model в Nerf

Reference

- https://arxiv.org/pdf/2003.08934
- https://github.com/google-research/multinerf
- https://github.com/bmild/nerf
- https://github.com/yenchenlin/nerf-pytorch
- https://docs.nerf.studio/
- https://github.com/Professor322/media-to-3d (Эта курсовая)