Novel View Synthesis

3D Graphics의 한 분야.
주어진 장면의 여러 시점 이미지를 활용
촬영되지 않은 새로운 시점의 이미지를 생성하는 기술

Before ML (~ 2010s)

SfM
(Structure from Motion)

MVS
(Multi-View Stereo)

- 3d model acquisition from extended image sequences (1996)
- OpenMVG: Open Multiple View Geometry (2016)

...

- Structure-from-Motion Revisited (CVPR, 2016)

...

COLMAP

NeRF & 3DGS

2020.03

NeRF
(Neural radiance field)

2023.08

3DGS
(3D Gaussian Splatting)

COLMAP-free

BEYOND

NeRF 계열

- Traning, Rendering Speed
- Memory
- Quality

- Pose Estimation
- Sparse Inputs

- 3D Geometry w/ Texturing
- Dynamic Scene, Multi-Scale

3DGS 계열

- Speed
- Quality
- Memory

- Pose Estimation
- Sparse Inputs

- Diffusion to 3D
(Image or Text to 3D)
- Segmentation

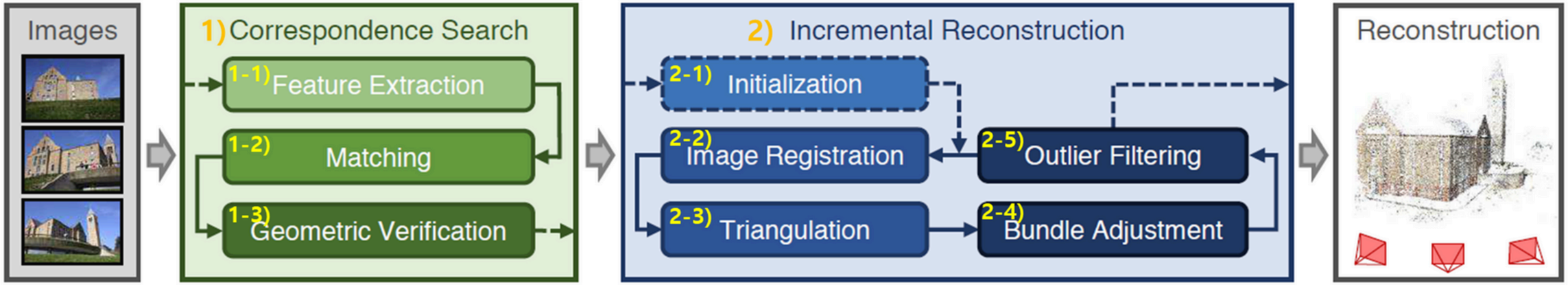
ETC

DUSt3R (2023.12)
InstantSplat(24.03)
MASt3R (2024.06)
MASt3R-SLAM
(2024.12)

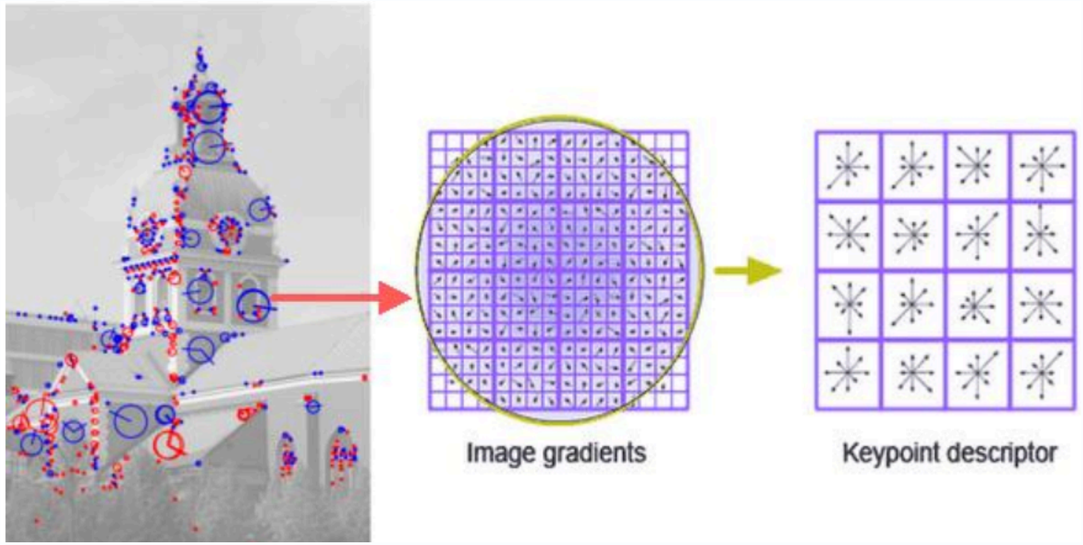
SVRaster (2025.05)

...

COLMAP-Incremental SfM

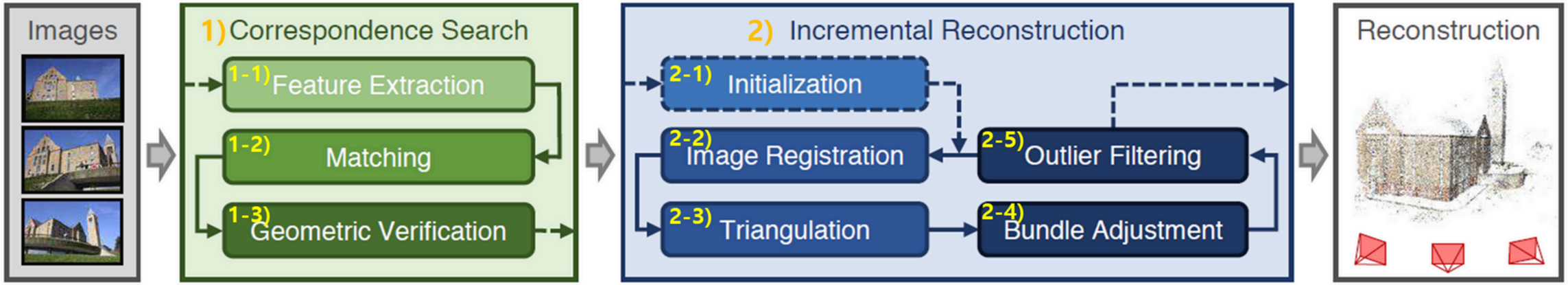


1-1) Feature Extraction

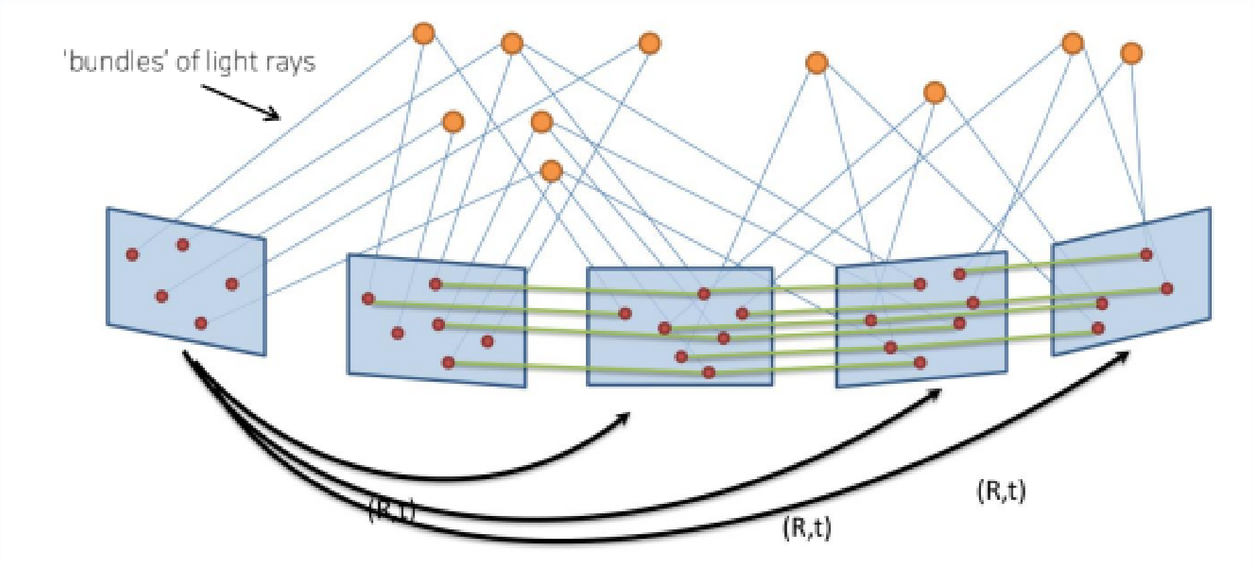


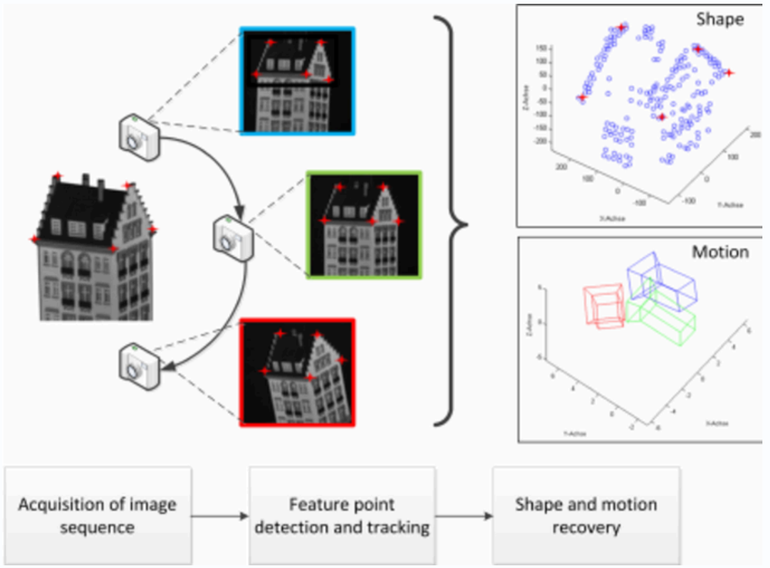
1-2) Matcing & 1-3) Geometric Verification

COLMAP-Incremental SfM

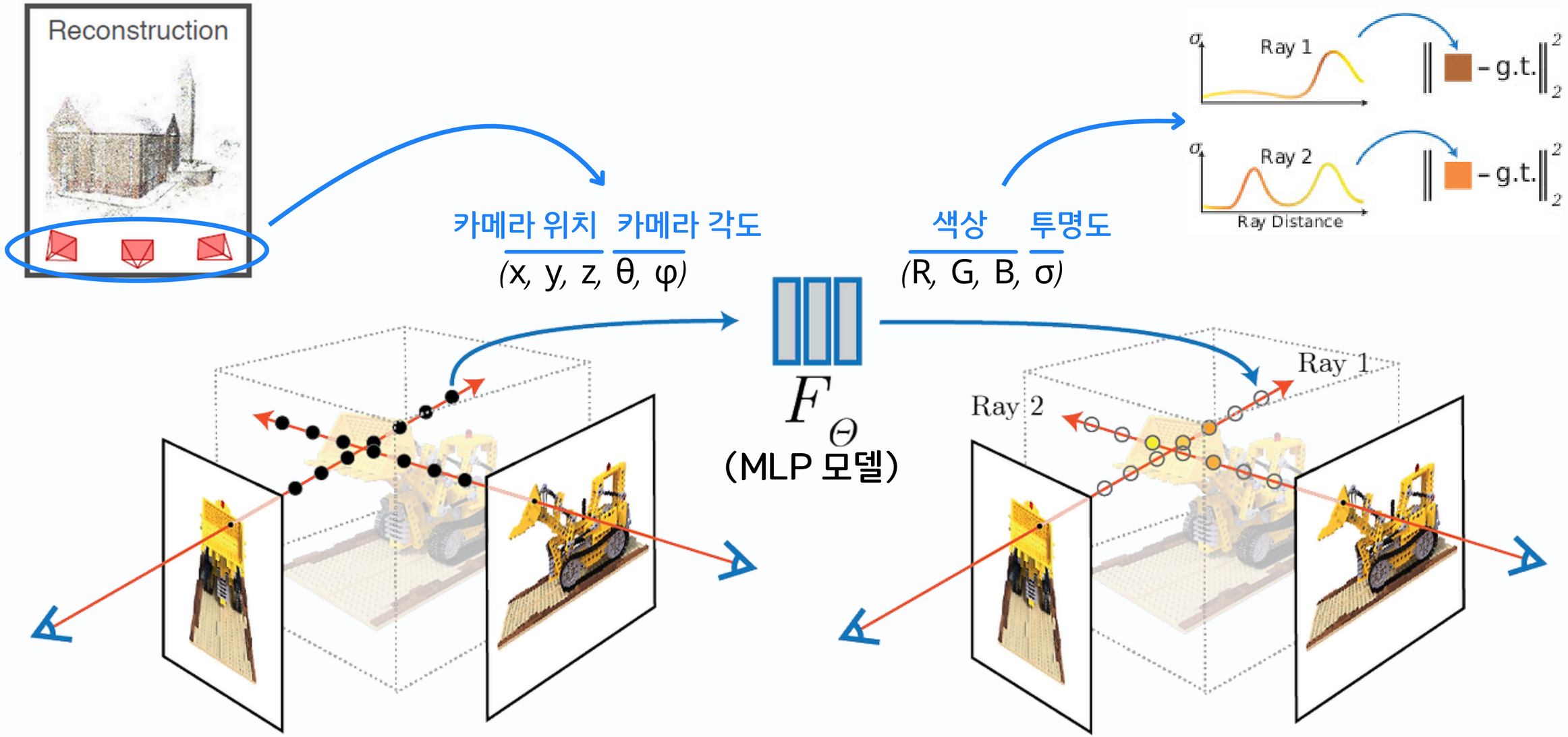


2) Incremental Reconstuction

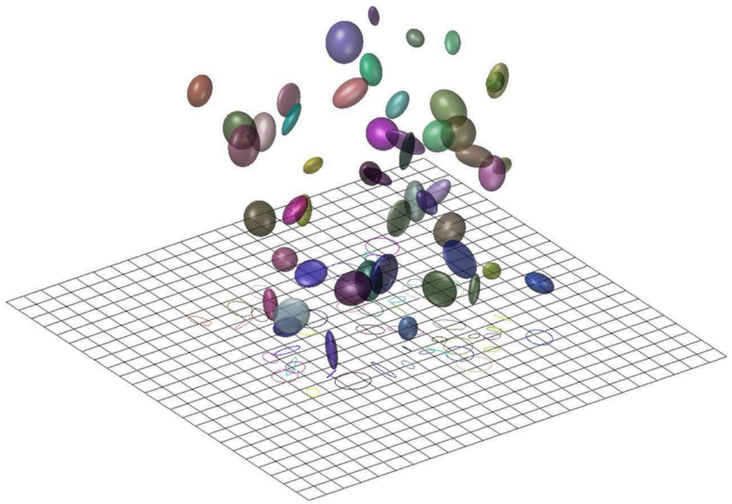




NeRF (Neural radiance field)



3DGS (3D Gaussian Splatting)



카메라 위치

point cloud

SfM Points

Initialization

3D Gaussians

Camera

Projection

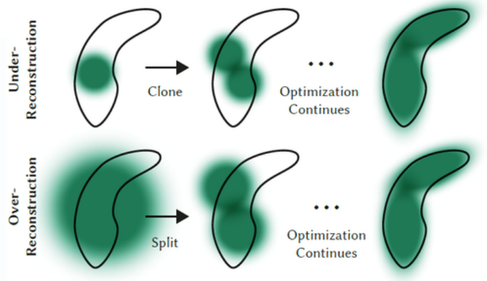
Differentiable Tile Rasterizer

Image

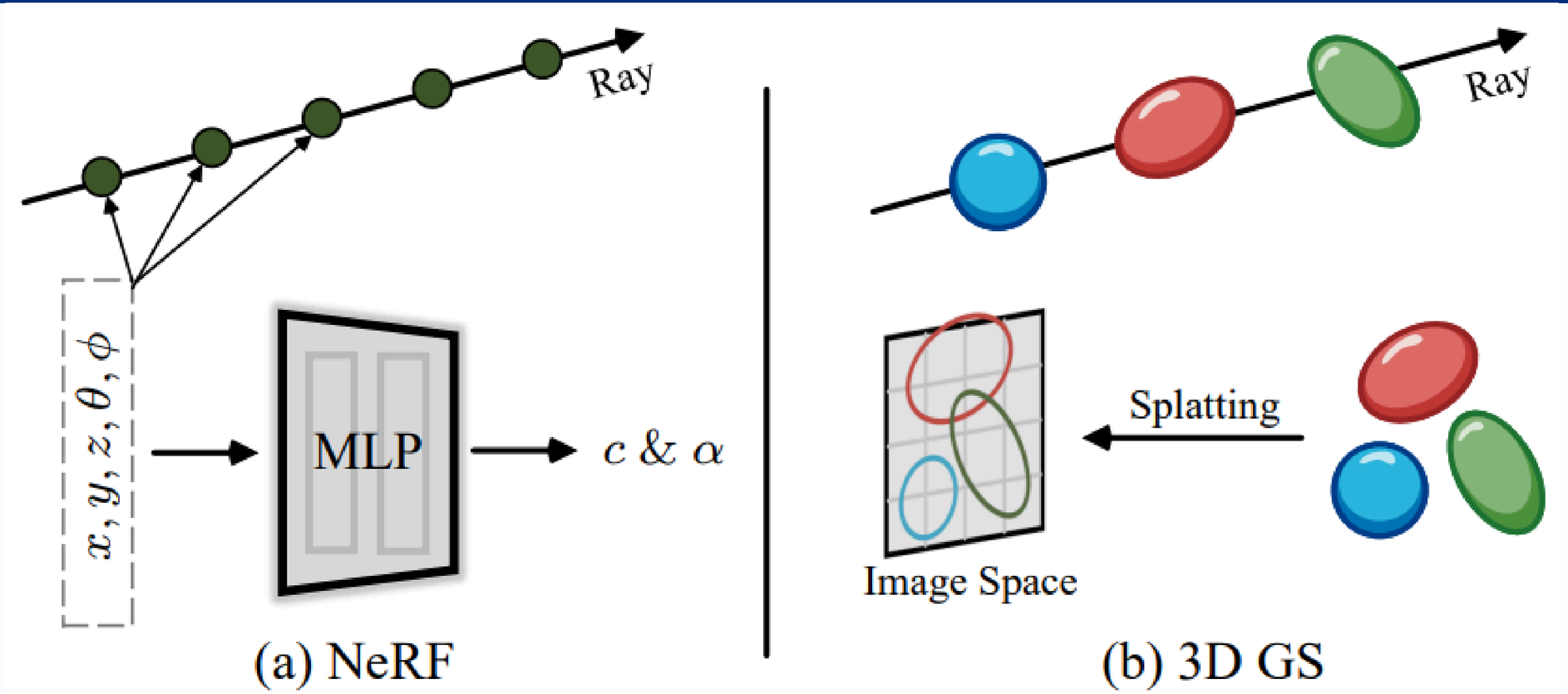
Adaptive Density Control

Operation Flow

Gradient Flow



NeRF & 3DGS



NeRF
3DGS

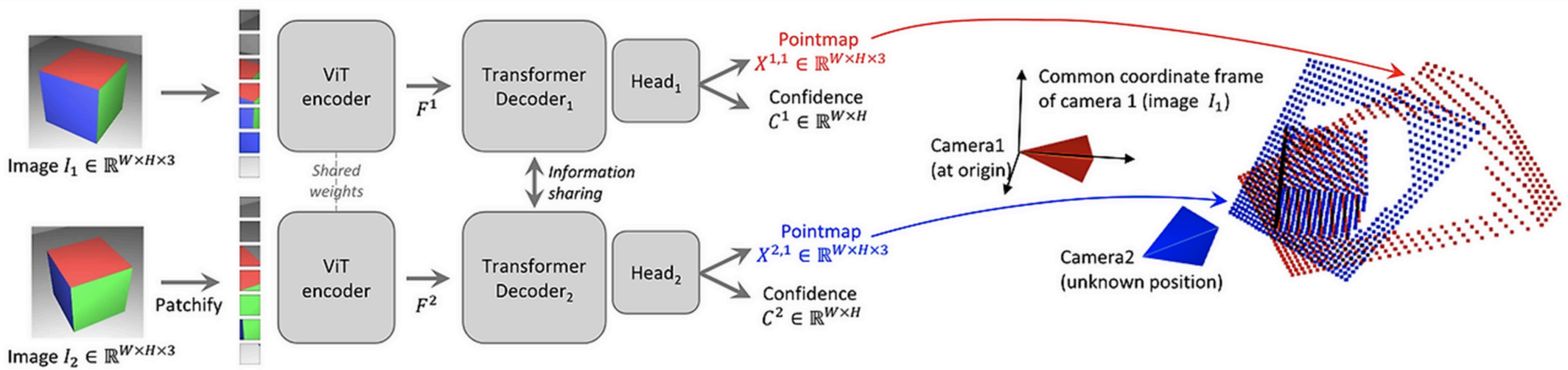
문제점 : COLMAP 의존적

1. 많은 input image 필요 (1000장 단위)
2. 복잡도 증가
3. 속도 저하
4. 퀄리티의 한계

해결책 : COLMAP – free

1. 머신러닝을 활용해(주로 transformer) 자체적으로 camera parameter, point cloud 예측
2. 적은 입력, 획기적으로 빠른 속도 가능
3. ex) DUSt3R / MASt3R ...

DUSt3R



1. Network architecture

$$F^1 = \text{Encoder}(I^1), F^2 = \text{Encoder}(I^2).$$

$$G_i^1 = \text{DecoderBlock}_i^1\left(G_{i-1}^1, G_{i-1}^2\right),$$
$$G_i^2 = \text{DecoderBlock}_i^2\left(G_{i-1}^2, G_{i-1}^1\right),$$

$$X^{1,1}, C^{1,1} = \text{Head}^1\left(G_0^1, \ldots, G_B^1\right),$$
$$X^{2,1}, C^{2,1} = \text{Head}^2\left(G_0^2, \ldots, G_B^2\right).$$

2. Training Objective

$$\ell_{\text{regr}}(v,i) = \left\| \frac{1}{z} X_i^{v,1} - \frac{1}{\bar{z}} \bar{X}_i^{v,1} \right\|.$$

$$z = \text{norm}(X^{1,1}, X^{2,1}), \bar{z} = \text{norm}(\bar{X}^{1,1}, \bar{X}^{2,1})$$

$$\mathcal{L}_{\text{conf}} = \sum_{v\in\{1,2\}} \sum_{i\in\mathcal{D}^v} C_i^{v,1} \ell_{\text{regr}}(v,i) - \alpha \log C_i^{v,1},$$

3. Downstream Applications

- Point Matching

$$\mathcal{M}_{1,2} = \{(i,j) \mid i = \text{NN}_1^{1,2}(j) \text{ and } j = \text{NN}_1^{2,1}(i)\}$$
$$\text{with } \text{NN}_k^{n,m}(i) = \underset{j\in\{0,\ldots,WH\}}{\arg\min} \left\| X_j^{n,k} - X_i^{m,k} \right\|.$$

- Recovering Intrinsics

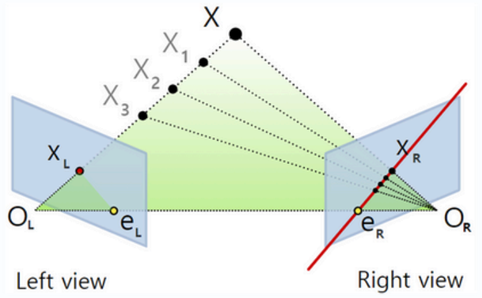
$$f_1^* = \underset{f_1}{\arg\min} \sum_{i=0}^{W} \sum_{j=0}^{H} C_{i,j}^{1,1} \left\| (i',j') - f_1 \frac{(X_{i,j,0}^{1,1}, X_{i,j,1}^{1,1})}{X_{i,j,2}^{1,1}} \right\|,$$

- Relative/Absolute Pose Estimation

$$R^*, t^* = \underset{\sigma,R,t}{\arg\min} \sum_i C_i^{1,1} C_i^{1,2} \left\| \sigma(R X_i^{1,1} + t) - X_i^{1,2} \right\|^2,$$

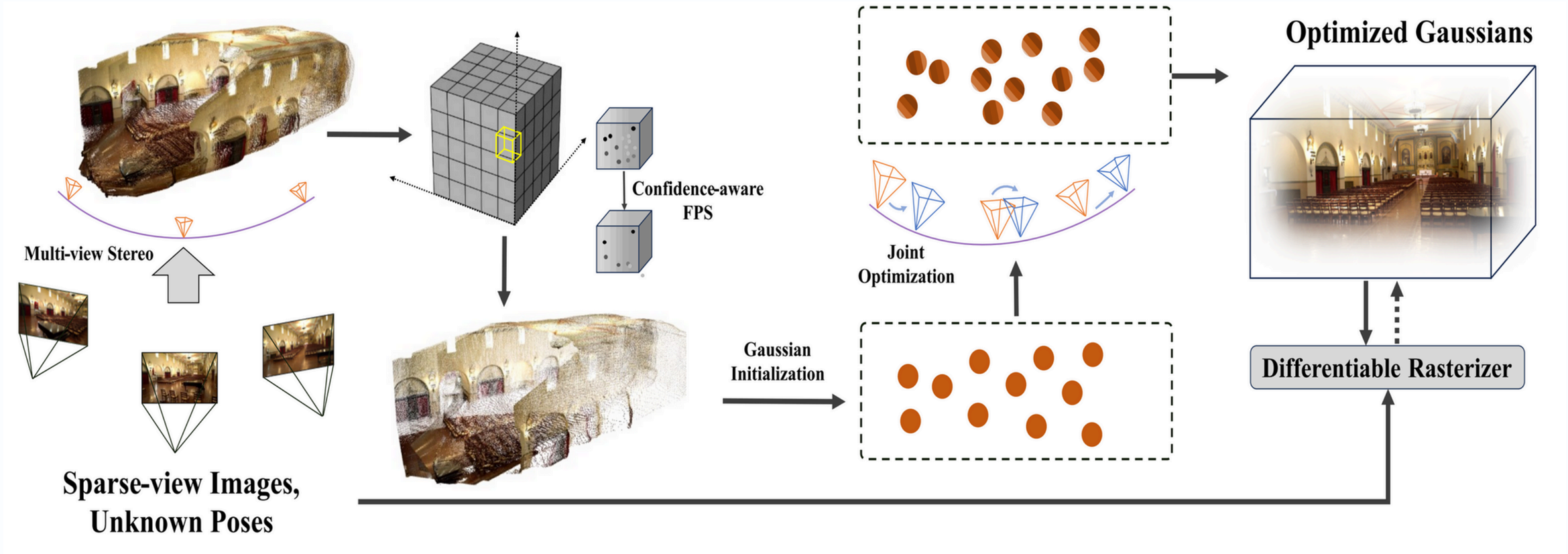
4. Global alignment (visual localization)

$$\chi^* = \underset{\chi,P,\sigma}{\arg\min} \sum_{e\in\mathcal{E}} \sum_{v\in e} \sum_{i=1}^{HW} C_i^{v,e} \left\| \chi_i^v - \sigma_e P_e X_i^{v,e} \right\|.$$



DUSt3R

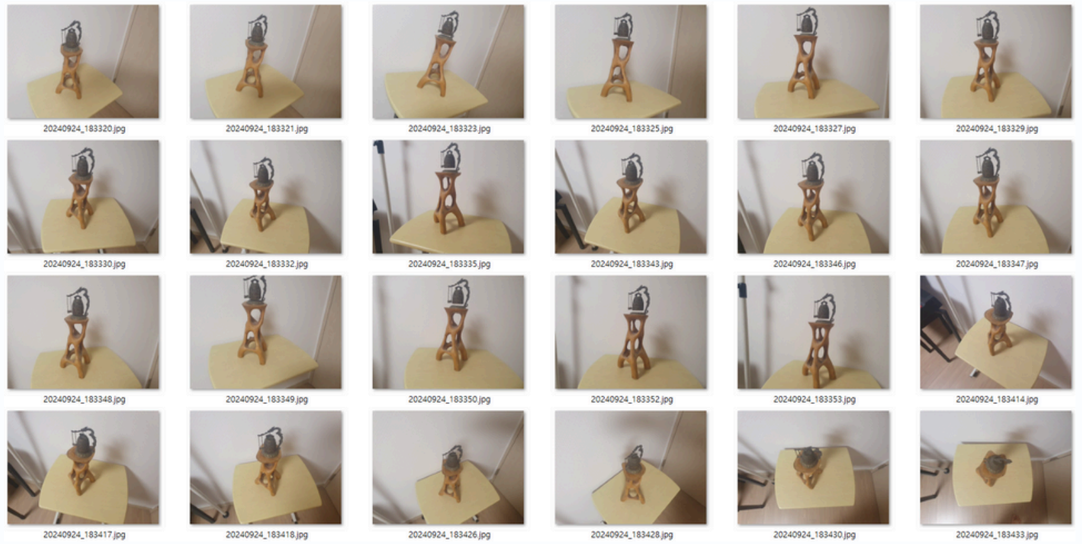
InstantSplat

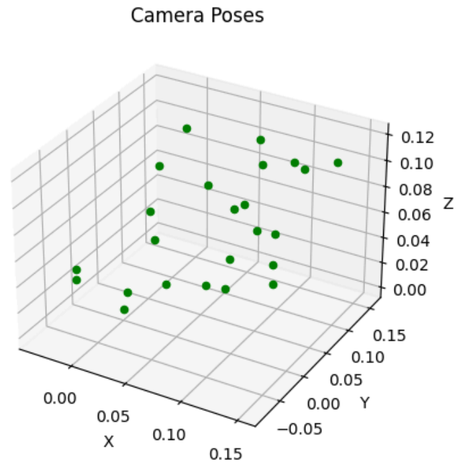


1. DUSt3R에 여러장을 입력으로 줌
   (모든 이미지들을 다 한번씩 붙여서 새로운 결과에 덧붙임)
2. Gaussian도 함께 학습함.

myDEMO
InstantSpat

InstantSplat - demo



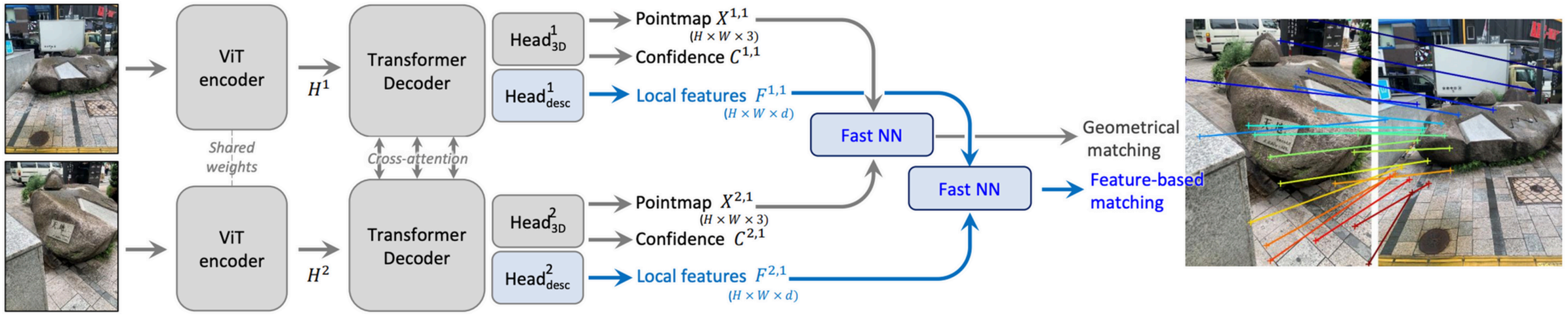


Pose Estimate Time for 24 Views
: 276.17 sec

[ITER 1000]
Time : 00:01:42
L 1 : 0.0189919345624124
PSNR : 30.71599189440409

MASt3R

0. framework



1. Matching prediction head and loss

$$\ell_{\text{regr}}(v,i) = \|X^i_{v,1} - \hat{X}^i_{v,1}\| / \hat{z}$$

$$\mathcal{L}_{\text{conf}} = \sum_{v\in\{1,2\}} \sum_{i\in\mathcal{D}^v} C^{v,1}_i \ell_{\text{regr}}(v,i) - \alpha \log C^{v,1}_i,$$

$$\mathcal{L}_{\text{match}} = -\sum_{(i,j)\in\hat{\mathcal{M}}} \log \frac{s_\tau(i,j)}{\sum_{k\in\mathcal{P}^1} s_\tau(k,j)} + \log \frac{s_\tau(i,j)}{\sum_{k\in\mathcal{P}^2} s_\tau(i,k)},$$

(from infoNCE)

$$\text{with } s_\tau(i,j) = \exp[-\tau D^{1\top}_i D^2_j].$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{conf}} + \beta \mathcal{L}_{\text{match}}$$

$$D^1 = \text{Head}^1_{\text{desc}}([H^1, H'^1])$$

$$D^2 = \text{Head}^2_{\text{desc}}([H^2, H'^2]).$$

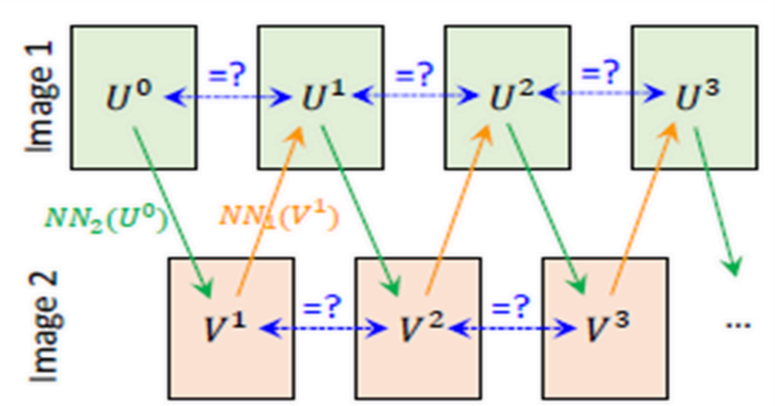
myDEMO

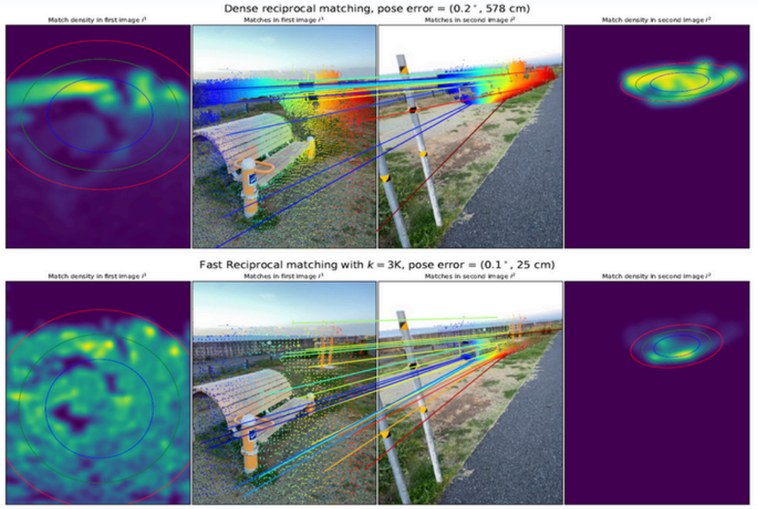
DEMO

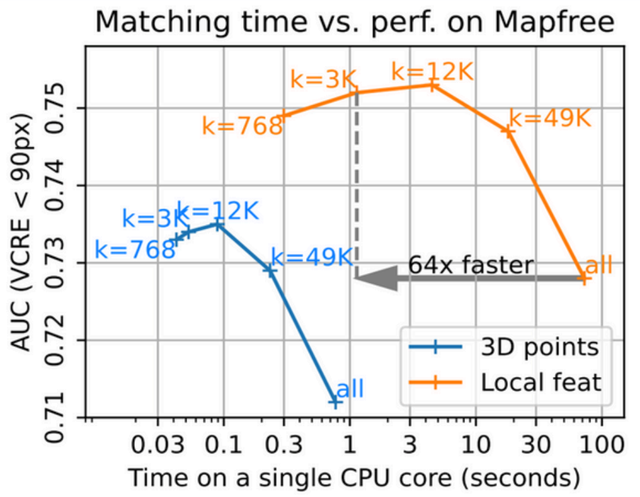
MASt3R

MASt3R

2. Fast reciprocal matching







$$U^t \longmapsto [\mathrm{NN}_2(D^1_u)]_{u \in U^t} \equiv V^t \longmapsto [\mathrm{NN}_1(D^2_v)]_{v \in V^t} \equiv U^{t+1}$$

3. Coarse-to-fine matching

1) 긴 변이 512 pixel이 되도록 downscale하여 matching 수행

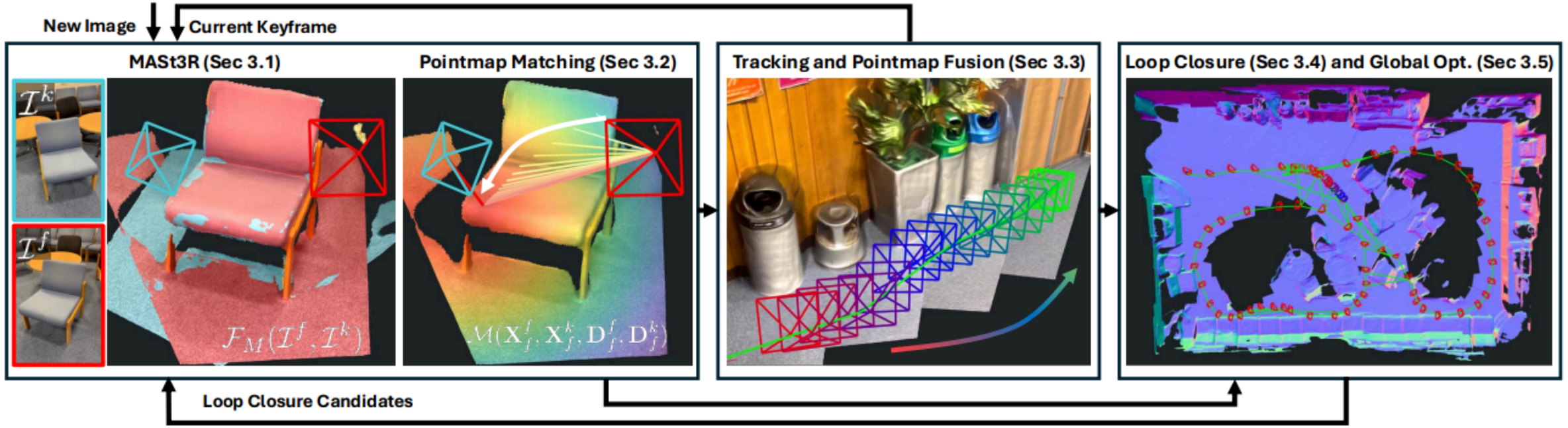
2) 원 이미지를 gird화 (긴 변 512 pixel, 50% overlap)
2-1) coarse matching 결과를 활용해
각 window pair들 끼리 matching 수행
2-2) coarse matching pair의 90% 이상이 커버되면 중지

$$D^{w_1}, D^{w_2} = \mathrm{MASt3R}(I^1_{w_1}, I^2_{w_2})$$

$$\mathcal{M}_k^{w_1,w_2} = \mathrm{fast\_reciprocal\_NN}(D^{w_1}, D^{w_2})$$

MASt3R-SLAM

SLAM : 로봇/자율주행 시스템이 실시간으로 주변 지도를 생성하며 동시에 본인 위치를 추정하는 기술



1. 기존 MASt3R은 입력받은 이미지의 쌍을 가능한 모든 조합으로 연결.
   but MASt3R-SLAM은 바로 직전 프레임에 대해서만 먼저 matching 수행
2. Keyframe(중요한 프레임 쌍) 후보 정보를 따로 저장해놨다가 최적화에 활용.

→ 실시간 처리 가능 & 정확도/일관성 확보

MASt3R-SLAM

More Fascinating 3DGS Applications

TransGS : 3DGS를 통한 아바타 얼굴 생성

ExAvatar : 간편한 전신 3D Gaussian avatar 생성

3DGStream : 실시간으로 동영상 시점 변경

DiffusionGS : 한 장의 이미지로 빠르게 3D 모델링

DiffusionGS : 프롬프트 기반 3D editing

...