SLAM & 3D Gaussian Splatting

Literature Review

Shuqi XIAO

June 27, 2024

Outline

1 Overview

2 MonoGS
- Methodology

Overview

Timeline

11/2023 - 12/2024

- SplaTAM [1]
- GS-SLAM [2]
- Gaussian-SLAM [3]
- MonoGS [4]

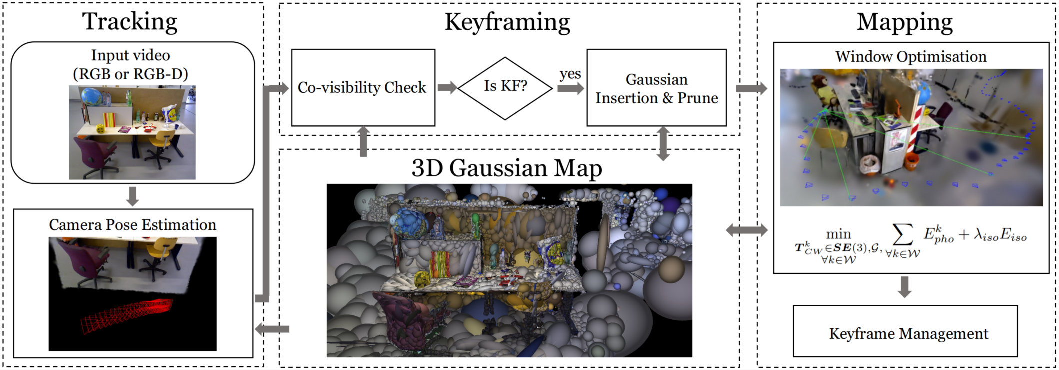
01/2024 - 04/2024

05/2024 - 06/2024

accepted to top conferences

MonoGS

Overview i



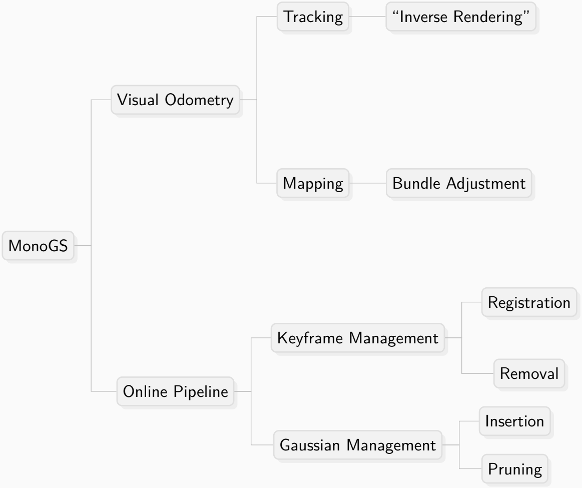
(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Overview ii

- MonoGS
  - Visual Odometry
    - Tracking
    - Mapping
  - Online Pipeline
    - Keyframe Management
    - Gaussian Management

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Overview ii



(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Overview ii

- MonoGS
  - Visual Odometry
    - Tracking
      - "Inverse Rendering"
        - Analytical Jacobians Derivation
        - Photometric RGB & Depth Loss
        - Optimizable Exposure
        - Penalization on non-edge/low-opacity pixels
    - Mapping
      - Bundle Adjustment
        - Photometric RGB & Depth Loss
        - Isotropic Regularization
        - Random Recall
  - Online Pipeline
    - Keyframe Management
      - Registration
        - Gaussian Covisibility
        - Relative Translation
      - Removal
        - Gaussian Overlap Coefficient
    - Gaussian Management
      - Insertion
        - Keyframing
      - Pruning
        - Gaussian Covisibility

key method

trick

convention

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Tracking: Overview

- Tracking
  - "Inverse Rendering"
    - Analytical Jacobians Derivation
    - Photometric RGB & Depth Loss
    - Optimizable Exposure
    - Penalization on non-edge/low-opacity pixels

Track camera poses,

- through the extended differentiable rendering pipeline,

---

key method   trick   convention

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Tracking: Overview

- Tracking
  - "Inverse Rendering"
    - Analytical Jacobians Derivation
    - Photometric RGB & Depth Loss
    - Optimizable Exposure
    - Penalization on non-edge/low-opacity pixels

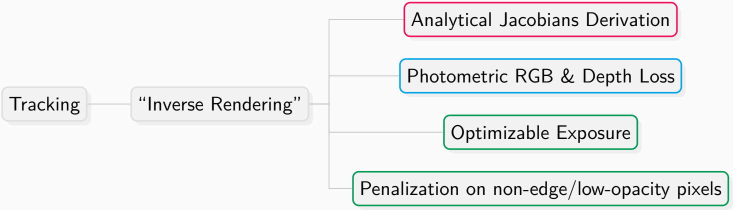
Track camera poses,

- through the extended differentiable rendering pipeline,
- by a direct optimization against fixed 3D Gaussians,

key method trick convention

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Tracking: Overview



Track camera poses,

- through the extended differentiable rendering pipeline,
- by a direct optimization against fixed 3D Gaussians,
- with some tricks to be more adaptive to brightness and more robust to noise.

---

key method   trick   convention

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Tracking: A Review of 3DGS

The projection from "ellipsoids" to "ellipses" in 3DGS,

$$\mathcal{N}\left(\mu_w, \Sigma_w\right) \overset{\pi}{\mapsto} \mathcal{N}\left(\mu_i, \Sigma_i\right), \tag{1}$$

is achieved by,

$$\mu_i = \pi\left(\mathbf{T}_{cw} \cdot \mu_w\right) \tag{2}$$

$$\Sigma_i = \mathbf{J}_\pi \mathbf{R}_{cw} \Sigma_w \mathbf{R}_{cw}^{\mathrm{T}} \mathbf{J}_\pi^{\mathrm{T}} \tag{3}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

The projection from “ellipsoids” to “ellipses” in 3DGS,

$$\mathcal{N}\left(\mu_w, \Sigma_w\right) \overset{\pi}{\mapsto} \mathcal{N}\left(\mu_i, \Sigma_i\right), \tag{1}$$

is achieved by,

$$\mu_i = \pi\left(\mathbf{T}_{cw} \cdot \underbrace{\mu_w}_{\in \mathbb{P}^3,\ \text{3D(world) mean}}\right) \tag{2}$$

$$\Sigma_i = \mathbf{J}_\pi \mathbf{R}_{cw} \Sigma_w \mathbf{R}_{cw}^{\mathrm{T}} \mathbf{J}_\pi^{\mathrm{T}} \tag{3}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

The projection from “ellipsoids” to “ellipses” in 3DGS,

$$\mathcal{N}\left(\mu_w, \Sigma_w\right) \overset{\pi}{\mapsto} \mathcal{N}\left(\mu_i, \Sigma_i\right), \tag{1}$$

is achieved by,

$$\mu_i = \pi\left(\underbrace{\mathbf{T}_{cw}}_{\in \mathrm{SE}(3),\ \text{camera pose}} \cdot \underbrace{\mu_w}_{\in \mathbb{P}^3,\ \text{3D(world) mean}}\right) \tag{2}$$

$$\Sigma_i = \mathbf{J}_\pi \mathbf{R}_{cw} \Sigma_w \mathbf{R}_{cw}^{\mathrm{T}} \mathbf{J}_\pi^{\mathrm{T}} \tag{3}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Tracking: A Review of 3DGS

The projection from "ellipsoids" to "ellipses" in 3DGS,

$$\mathcal{N}\left(\mu_w, \Sigma_w\right) \overset{\pi}{\mapsto} \mathcal{N}\left(\mu_i, \Sigma_i\right), \tag{1}$$

is achieved by,

$$\mu_i = \underbrace{\pi}_{\text{projection}} \left(\underbrace{\mathbf{T}_{cw}}_{\in \mathrm{SE}(3),\ \text{camera pose}} \cdot \underbrace{\mu_w}_{\in \mathbb{P}^3,\ \text{3D(world) mean}} \right) \tag{2}$$

$$\Sigma_i = \mathbf{J}_\pi \mathbf{R}_{cw} \Sigma_w \mathbf{R}_{cw}^{\mathrm{T}} \mathbf{J}_\pi^{\mathrm{T}} \tag{3}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Tracking: A Review of 3DGS

The projection from "ellipsoids" to "ellipses" in 3DGS,

$$\mathcal{N}\left(\mu_w, \Sigma_w\right) \overset{\pi}{\mapsto} \mathcal{N}\left(\mu_i, \Sigma_i\right), \tag{1}$$

is achieved by,

$$\mu_i = \pi\left(\mathbf{T}_{cw} \cdot \mu_w\right) \tag{2}$$

$\in \mathbb{P}^3$, 3D(world) mean

$\in \mathrm{SE}(3)$, camera pose

projection

$\in \mathbb{P}^2$, 2D(image) mean

$$\Sigma_i = \mathbf{J}_\pi \mathbf{R}_{cw} \Sigma_w \mathbf{R}_{cw}^{\mathrm{T}} \mathbf{J}_\pi^{\mathrm{T}} \tag{3}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

The projection from "ellipsoids" to "ellipses" in 3DGS,

$$\mathcal{N}\left(\mu_w, \Sigma_w\right) \overset{\pi}{\mapsto} \mathcal{N}\left(\mu_i, \Sigma_i\right), \tag{1}$$

is achieved by,

$$\underbrace{\mu_i}_{\in \mathbb{P}^2,\ \text{2D(image) mean}} = \underbrace{\pi}_{\text{projection}} \left(\underbrace{\mathbf{T}_{cw}}_{\in \mathrm{SE}(3),\ \text{camera pose}} \cdot \underbrace{\mu_w}_{\in \mathbb{P}^3,\ \text{3D(world) mean}} \right) \tag{2}$$

$$\Sigma_i = \mathbf{J}_\pi \mathbf{R}_{cw} \underbrace{\Sigma_w}_{\in \mathbb{R}^{3\times 3},\ \text{3D(world) covariance}} \mathbf{R}_{cw}^{\mathrm{T}} \mathbf{J}_\pi^{\mathrm{T}} \tag{3}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Tracking: A Review of 3DGS

The projection from "ellipsoids" to "ellipses" in 3DGS,

$$\mathcal{N}\left(\mu_w, \Sigma_w\right) \overset{\pi}{\mapsto} \mathcal{N}\left(\mu_i, \Sigma_i\right), \tag{1}$$

is achieved by,

$$\underbrace{\mu_i}_{\in \mathbb{P}^2,\ \text{2D(image) mean}} = \underbrace{\pi}_{\text{projection}} \left(\underbrace{\mathbf{T}_{cw}}_{\in \mathrm{SE}(3),\ \text{camera pose}} \cdot \underbrace{\mu_w}_{\in \mathbb{P}^3,\ \text{3D(world) mean}} \right) \tag{2}$$

$$\Sigma_i = \mathbf{J}_\pi \underbrace{\mathbf{R}_{cw}}_{\in \mathrm{SO}(3),\ \text{rotation component of } \mathbf{T}_{cw}} \underbrace{\Sigma_w}_{\in \mathbb{R}^{3\times3},\ \text{3D(world) covariance}} \mathbf{R}_{cw}^{\mathrm{T}} \mathbf{J}_\pi^{\mathrm{T}} \tag{3}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

The projection from "ellipsoids" to "ellipses" in 3DGS,

$$\mathcal{N}\left(\mu_w, \Sigma_w\right) \overset{\pi}{\mapsto} \mathcal{N}\left(\mu_i, \Sigma_i\right), \tag{1}$$

is achieved by,

$$\underbrace{\mu_i}_{\in \mathbb{P}^2,\ \text{2D(image) mean}} = \underbrace{\pi}_{\text{projection}} \left(\underbrace{\mathbf{T}_{cw}}_{\in \mathrm{SE}(3),\ \text{camera pose}} \cdot \underbrace{\mu_w}_{\in \mathbb{P}^3,\ \text{3D(world) mean}} \right) \tag{2}$$

$$\Sigma_i = \underbrace{\mathbf{J}_\pi}_{\in \mathbb{R}^{2\times 3},\ \text{Jacobian of the linear approximation of } \pi} \underbrace{\mathbf{R}_{cw}}_{\in \mathrm{SO}(3),\ \text{rotation component of } \mathbf{T}_{cw}} \underbrace{\Sigma_w}_{\in \mathbb{R}^{3\times 3},\ \text{3D(world) covariance}} \mathbf{R}_{cw}^{\mathrm{T}} \mathbf{J}_\pi^{\mathrm{T}} \tag{3}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Tracking: A Review of 3DGS

The projection from "ellipsoids" to "ellipses" in 3DGS,

$$\mathcal{N}\left(\mu_w, \Sigma_w\right) \overset{\pi}{\mapsto} \mathcal{N}\left(\mu_i, \Sigma_i\right), \tag{1}$$

is achieved by,

$$\mu_i = \pi\left(\mathbf{T}_{cw} \cdot \mu_w\right) \tag{2}$$

$\in \mathbb{P}^3$, 3D(world) mean

$\in \mathrm{SE}(3)$, camera pose

projection

$\in \mathbb{P}^2$, 2D(image) mean

$$\Sigma_i = \mathbf{J}_\pi \mathbf{R}_{cw} \Sigma_w \mathbf{R}_{cw}^{\mathrm{T}} \mathbf{J}_\pi^{\mathrm{T}} \tag{3}$$

$\in \mathbb{R}^{3\times 3}$, 3D(world) covariance

$\in \mathrm{SO}(3)$, rotation component of $\mathbf{T}_{cw}$

$\in \mathbb{R}^{2\times 3}$, Jacobian of the linear approximation of π

$\in \mathbb{R}^{2\times 2}$, 2D(image) covariance

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

The chain rule,

$$\frac{\partial \mu_i}{\partial \mathbf{T}_{cw}} = \frac{\partial \mu_i}{\partial \mu_c} \frac{\partial \mu_c}{\partial \mathbf{T}_{cw}} \tag{4}$$

$$\frac{\partial \Sigma_i}{\partial \mathbf{T}_{cw}} = \frac{\partial \Sigma_i}{\partial \mathbf{J}_\pi} \frac{\partial \mathbf{J}_\pi}{\partial \mu_c} \frac{\partial \mu_c}{\partial \mathbf{T}_{cw}} + \frac{\partial \Sigma_i}{\partial \mathbf{R}_{cw}} \frac{\partial \mathbf{R}_{cw}}{\partial \mathbf{T}_{cw}} \tag{5}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

The chain rule,

$$\frac{\partial \mu_i}{\partial \mathbf{T}_{cw}} = \frac{\partial \mu_i}{\partial \mu_c}\frac{\partial \mu_c}{\partial \mathbf{T}_{cw}} \tag{4}$$

$$\frac{\partial \Sigma_i}{\partial \mathbf{T}_{cw}} = \frac{\partial \Sigma_i}{\partial \mathbf{J}_\pi}\frac{\partial \mathbf{J}_\pi}{\partial \mu_c}\frac{\partial \mu_c}{\partial \mathbf{T}_{cw}} + \frac{\partial \Sigma_i}{\partial \mathbf{R}_{cw}}\frac{\partial \mathbf{R}_{cw}}{\partial \mathbf{T}_{cw}} \tag{5}$$

The Lie Algebra,

$$\frac{\partial \mu_c}{\partial \mathbf{T}_{cw}} = \begin{bmatrix} \mathbf{I} & -\mu_c^\times \end{bmatrix} \tag{6}$$

$$\frac{\partial \mathbf{R}_{cw}}{\partial \mathbf{T}_{cw}} = \begin{bmatrix} \mathbf{0} & -\mathbf{R}_{cw}^\times(:,1) \\ \mathbf{0} & -\mathbf{R}_{cw}^\times(:,2) \\ \mathbf{0} & -\mathbf{R}_{cw}^\times(:,3) \end{bmatrix} \tag{7}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Online Pipeline: Overview

- Online Pipeline
  - Keyframe Management
    - Registration
      - Gaussian Covisibility
      - Relative Translation
    - Removal
      - Gaussian Overlap Coefficient
  - Gaussian Management
    - Insertion
      - Keyframing
    - Pruning
      - Gaussian Covisibility

Keyframe Management:

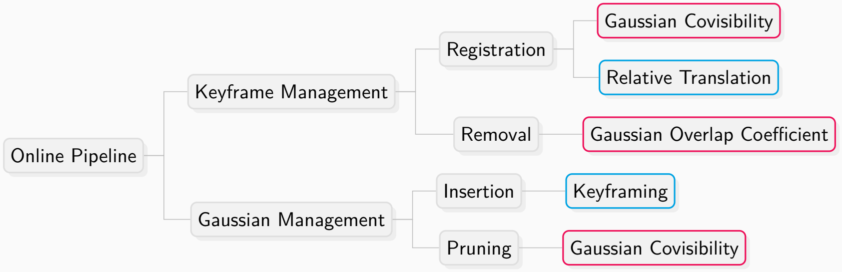
---

key method trick convention

(arXiv, 2016) DSO: Direct Sparse Odometry

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Online Pipeline: Overview



Keyframe Management:

- Classic strategies, e.g. covisibility & overlap, from DSO [5].

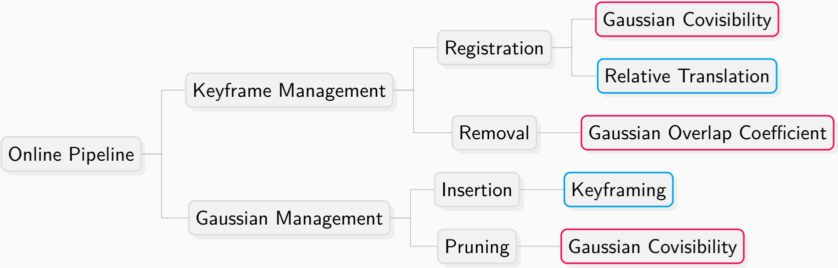
---

key method trick convention

(arXiv, 2016) DSO: Direct Sparse Odometry

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Online Pipeline: Overview



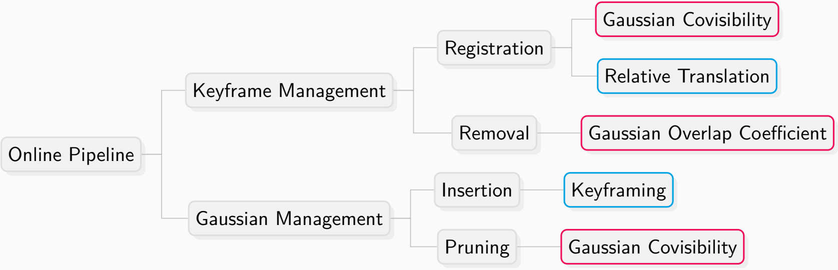
Keyframe Management:

- Classic strategies, e.g. covisibility & overlap, from DSO [5].
- Off-the-shelf occlusion-aware Gaussian visibility is leveraged to construct metrics.

key method trick convention

(arXiv, 2016) DSO: Direct Sparse Odometry

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM



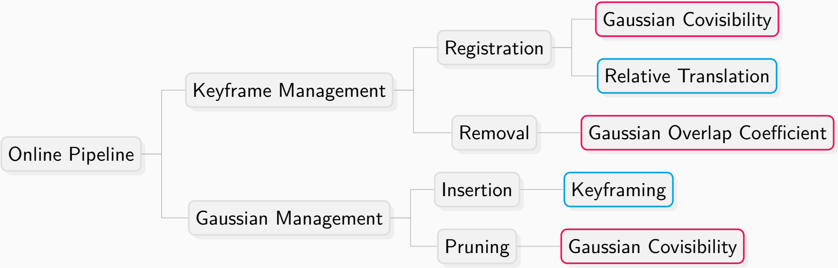
Gaussian Management:

key method   trick   convention

(arXiv, 2016) DSO: Direct Sparse Odometry

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Online Pipeline: Overview



Gaussian Management:

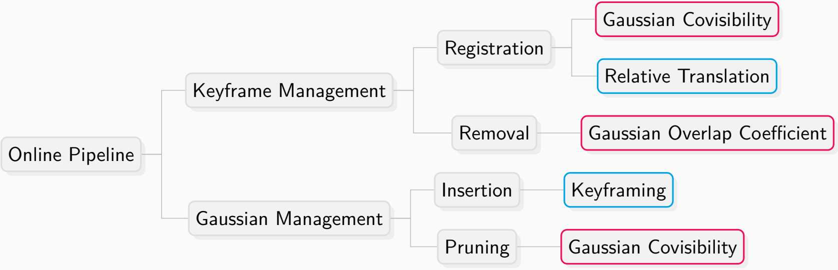
- Insertion: triggered by keyframing, followed by Gaussian initialization.

key method trick convention

(arXiv, 2016) DSO: Direct Sparse Odometry

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Online Pipeline: Overview



Gaussian Management:

- Insertion: triggered by keyframing, followed by Gaussian initialization.
- Pruning: to remove unstable/incorrect Gaussians by covisibility in a monocular setting.

---

key method trick convention

(arXiv, 2016) DSO: Direct Sparse Odometry
(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Keyframe Management: Prerequistes

1. What is keyframing or keyframe management?

Keyframe Management: Prerequistes

1. What is keyframing or keyframe management?
   - A strategy of selecting and utilizing a crucial subset of frames.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Keyframe Management: Prerequistes

1. What is keyframing or keyframe management?
   - A strategy of selecting and utilizing a crucial subset of frames.
2. Why do we need keyframing?

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Keyframe Management: Prerequistes

1. What is keyframing or keyframe management?
   - A strategy of selecting and utilizing a crucial subset of frames.
2. Why do we need keyframing?
   - Infeasible to optimize jointly on all frames online.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Keyframe Management: Prerequistes

1. What is keyframing or keyframe management?
   - A strategy of selecting and utilizing a crucial subset of frames.
2. Why do we need keyframing?
   - Infeasible to optimize jointly on all frames online.

     (a trade-off between efficiency and accuracy/robustness/...)

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Keyframe Management: Prerequistes

1. What is keyframing or keyframe management?
   - A strategy of selecting and utilizing a crucial subset of frames.
2. Why do we need keyframing?
   - Infeasible to optimize jointly on all frames online.

     (a trade-off between efficiency and accuracy/robustness/...)
3. How should we select keyframes?

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Keyframe Management: Prerequistes

1. What is keyframing or keyframe management?
   - A strategy of selecting and utilizing a crucial subset of frames.
2. Why do we need keyframing?
   - Infeasible to optimize jointly on all frames online.

     (a trade-off between efficiency and accuracy/robustness/...)
3. How should we select keyframes?
   - non-redundant and observing the same area.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Keyframe Management: Prerequistes

1. What is keyframing or keyframe management?
   - A strategy of selecting and utilizing a crucial subset of frames.
2. Why do we need keyframing?
   - Infeasible to optimize jointly on all frames online.

     (a trade-off between efficiency and accuracy/robustness/...)
3. How should we select keyframes?
   - non-redundant and observing the same area.
   - spanning a wide baseline for better multi-view constraints.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Keyframe Management: Registration

If any of the following conditions is true...

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

If any of the following conditions is true...

Small Gaussian Covisibility **Condition i, Keyframe Registration**

Gaussian covisibility between the current frame and the previous keyframe drops below a threshold.

$\subset \mathcal{G}$, **visible** Gaussians from frame j

$$\frac{|\,\mathrm{v}(\mathcal{G},\mathcal{F}_i) \cap \mathrm{v}(\mathcal{G},\mathcal{F}_j)\,|}{|\,\mathrm{v}\left(\mathcal{G},\mathcal{F}_i\right) \cup \mathrm{v}\left(\mathcal{G},\mathcal{F}_j\right)\,|} < \tau_1$$

the previous keyframe

the current frame

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Keyframe Management: Registration

If any of the following conditions is true...

Small Gaussian Covisibility — **Condition i, Keyframe Registration**

Gaussian covisibility between the current frame and the previous keyframe drops below a threshold.

$$\frac{|\,\mathrm{v}\left(\mathcal{G}, \mathcal{F}_i\right) \cap \mathrm{v}\left(\mathcal{G}, \mathcal{F}_j\right)\,|}{|\,\mathrm{v}\left(\mathcal{G}, \mathcal{F}_i\right) \cup \mathrm{v}\left(\mathcal{G}, \mathcal{F}_j\right)\,|} < \tau_1 \tag{8}$$

Large Relative Translation — **Condition ii, Keyframe Registration**

Translation from the previous keyframe w.r.t. to the median depth reaches a threshold.

$$\frac{\left\|\mathbf{t}_{\mathcal{F}_i\mathcal{F}_j}\right\|_2}{\bar{D}_{\mathcal{F}_i\mathcal{F}_j}} > \tau_2, \quad \bar{D}_{\mathcal{F}_i\mathcal{F}_j} = \frac{1}{2\,\underbrace{H}_{\text{image height}}\,\underbrace{W}_{\text{image width}}} \sum^{\{\mathcal{F}_i,\mathcal{F}_j\}} \sum_{h=0}^{H} \sum_{w=0}^{W} \underbrace{d(h, w)}_{\text{depth of pixel } (h,\, w)}$$

Keyframe Management: Registration

If any of the following conditions is true...

Small Gaussian Covisibility — **Condition i, Keyframe Registration**

Gaussian covisibility between the current frame and the previous keyframe drops below a threshold.

$$\frac{\left|\, \mathrm{v}\left(\mathcal{G}, \mathcal{F}_i\right) \cap \mathrm{v}\left(\mathcal{G}, \mathcal{F}_j\right) \,\right|}{\left|\, \mathrm{v}\left(\mathcal{G}, \mathcal{F}_i\right) \cup \mathrm{v}\left(\mathcal{G}, \mathcal{F}_j\right) \,\right|} < \tau_1 \tag{8}$$

Large Relative Translation — **Condition ii, Keyframe Registration**

Translation from the previous keyframe w.r.t. to the median depth reaches a threshold.

$$\frac{\left\|\mathbf{t}_{\mathcal{F}_i \mathcal{F}_j}\right\|_2}{\bar{D}_{\mathcal{F}_i \mathcal{F}_j}} > \tau_2, \quad \bar{D}_{\mathcal{F}_i \mathcal{F}_j} = \frac{1}{2HW} \sum^{\{\mathcal{F}_i, \mathcal{F}_j\}} \sum_{h=0}^{H} \sum_{w=0}^{W} d(h, w) \tag{9}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Keyframe Management: Removal

If any of the following conditions is true...

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Keyframe Management: Removal

If any of the following conditions is true...

Beyond Window Capacity **Condition i, Keyframe Removal**

Remove the earliest keyframe out of the sliding window if the capacity is exceeded.

$$|\mathcal{W}| < \tau_3 \tag{10}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Keyframe Management: Removal

If any of the following conditions is true...

Beyond Window Capacity — **Condition i, Keyframe Removal**

Remove the earliest keyframe out of the sliding window if the capacity is exceeded.

$$|\mathcal{W}| < \tau_3 \tag{10}$$

Low Gaussian Overlap Coefficient — **Condition ii, Keyframe Removal**

Remove the previous keyframe if the "Gaussian overlap coefficient" between the previous frame and the new keyframe drops below a threshold.

$$\frac{|\,\mathrm{v}\,(\mathcal{G},\mathcal{F}_i) \cap \mathrm{v}\,(\mathcal{G},\mathcal{F}_j)\,|}{\min\left(|\,\mathrm{v}\,(\mathcal{G},\mathcal{F}_i)\,|, |\,\mathrm{v}\,(\mathcal{G},\mathcal{F}_j)\,|\right)} < \tau_4 \tag{11}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Insertion

- Why do we need "Gaussian insertion"?

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Insertion

- Why do we need "Gaussian insertion"?
  - SLAM is for robotic exploration.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Insertion

- Why do we need "Gaussian insertion"?
  - SLAM is for robotic exploration.

- When do we need "Gaussian insertion"?

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Insertion

- Why do we need "Gaussian insertion"?
  - SLAM is for robotic exploration.
- When do we need "Gaussian insertion"?

Keyframing — **Condition i, Gaussian Insertion**

Insertion is triggered for every new keyframe.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Initialization

- How do we insert Gaussians?

In practice, “low”: 0.2σ; “high”: 0.5σ, where σ is the standard deviation of the rendered depth map.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Initialization

- How do we insert Gaussians?
  - Gaussian insertion is Gaussian initialization.

In practice, “low”: 0.2σ; “high”: 0.5σ, where σ is the standard deviation of the rendered depth map.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Initialization

- How do we insert Gaussians?
  - Gaussian insertion is Gaussian initialization.

If Depth Available	Gaussian Initialization
Back-project in a per-pixel, per-Gaussian approach.	

In practice, "low": 0.2σ; "high": 0.5σ, where σ is the standard deviation of the rendered depth map.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Initialization

- How do we insert Gaussians?
  - Gaussian insertion is Gaussian initialization.

If Depth Available — **Gaussian Initialization**

Back-project in a per-pixel, per-Gaussian approach.

If Depth Unavailable — **Gaussian Initialization**

Leverage the rendered depth map.

In practice, "low": 0.2σ; "high": 0.5σ, where σ is the standard deviation of the rendered depth map.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Initialization

- How do we insert Gaussians?
  - Gaussian insertion is Gaussian initialization.

If Depth Available **Gaussian Initialization**

Back-project in a per-pixel, per-Gaussian approach.

If Depth Unavailable **Gaussian Initialization**

Leverage the rendered depth map.

- for pixels with depth: use the rendered depth and assign a “low” covariance.

In practice, “low”: 0.2σ; “high”: 0.5σ, where σ is the standard deviation of the rendered depth map.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Initialization

- How do we insert Gaussians?
  - Gaussian insertion is Gaussian initialization.

If Depth Available	**Gaussian Initialization**
Back-project in a per-pixel, per-Gaussian approach.	

If Depth Unavailable — **Gaussian Initialization**

Leverage the rendered depth map.

- for pixels with depth: use the rendered depth and assign a "low" covariance.
- for pixels w/o depth: use the median of rendered depth and assign a "high" covariance.

In practice, "low": 0.2σ; "high": 0.5σ, where σ is the standard deviation of the rendered depth map.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Pruning

- Why do we need “Gaussian Pruning” **if depth unavailable**?

If no pruning, although the majority of incorrect Gaussians vanish quickly in following optimization, there are some survivals.

In practice, the opacity threshold is 0.7.

In practice, the pruned Gaussians are inserted in the last 3 keyframes and unobserved by any other 3 keyframes in the sliding window.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Pruning

- Why do we need "Gaussian Pruning" **if depth unavailable**?
  - Too many incorrect/unstable newly inserted Gaussians.

If no pruning, although the majority of incorrect Gaussians vanish quickly in following optimization, there are some survivals.

In practice, the opacity threshold is 0.7.

In practice, the pruned Gaussians are inserted in the last 3 keyframes and unobserved by any other 3 keyframes in the sliding window.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Pruning

- Why do we need “Gaussian Pruning” **if depth unavailable**?
    - Too many incorrect/unstable newly inserted Gaussians.

Low Gaussian Opacity — **Condition i, Gaussian Pruning**

The Gaussians with a “low” opacity are pruned.

If no pruning, although the majority of incorrect Gaussians vanish quickly in following optimization, there are some survivals.

In practice, the opacity threshold is 0.7.

In practice, the pruned Gaussians are inserted in the last 3 keyframes and unobserved by any other 3 keyframes in the sliding window.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Gaussian Management: Pruning

- Why do we need "Gaussian Pruning" **if depth unavailable**?
  - Too many incorrect/unstable newly inserted Gaussians.

Low Gaussian Opacity	**Condition i, Gaussian Pruning**
The Gaussians with a "low" opacity are pruned.	

Low Gaussian Covisibility	**Condition ii, Gaussian Pruning**
For "just" inserted Gaussians but unobserved by "some other" keyframes, are pruned out.	

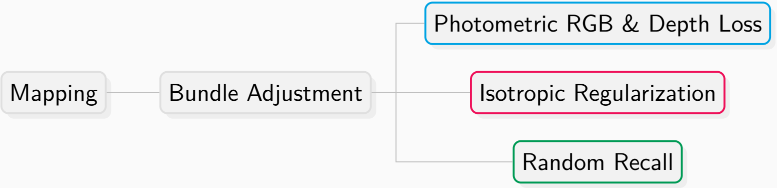
If no pruning, although the majority of incorrect Gaussians vanish quickly in following optimization, there are some survivals.

In practice, the opacity threshold is 0.7.

In practice, the pruned Gaussians are inserted in the last 3 keyframes and unobserved by any other 3 keyframes in the sliding window.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Mapping: Overview



key method trick convention

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Mapping: Overview

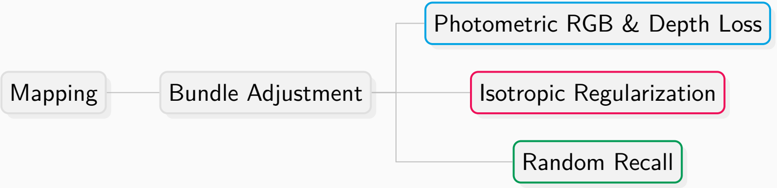
- Mapping
  - Bundle Adjustment
    - Photometric RGB & Depth Loss
    - Isotropic Regularization
    - Random Recall

- Why do we need mapping in **3DGS** SLAM?

key method trick convention

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Mapping: Overview



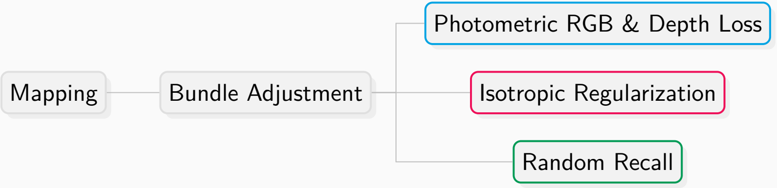
- Why do we need mapping in **3DGS** SLAM?
  - Local Mapping: Optimize newly inserted 3D Gaussians.

---

key method trick convention

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Mapping: Overview



- Why do we need mapping in **3DGS** SLAM?
  - Local Mapping: Optimize newly inserted 3D Gaussians.
  - Global Mapping: Reconstruct a 3D-coherent structure.

key method trick convention

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Mapping: Bundle Adjustment & Random Recall

Bundle Adjustment

keyframes in the sliding window

$$\underset{\mathcal{G},\{\mathbf{T}_{cw}(\mathcal{F}_k)|\mathcal{F}_k\in\mathcal{W}\}}{\operatorname{argmin}} \sum_{\mathcal{F}_k}^{\boxed{\mathcal{W}}} \mathcal{L}_{pho}\left(\mathcal{F}_k\right) \tag{12}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Mapping: Bundle Adjustment & Random Recall

Bundle Adjustment

keyframes in the sliding window

$$\underset{\mathcal{G}, \{\mathbf{T}_{cw}(\mathcal{F}_k) | \mathcal{F}_k \in \mathcal{W}\}}{\operatorname{argmin}} \sum_{\mathcal{F}_k}^{\mathcal{W}} \mathcal{L}_{pho}\left(\mathcal{F}_k\right) \tag{12}$$

Random Recall — A trick for global mapping

Besides $\mathcal{W}$, "some" randomly selected past keyframes are also leveraged in BA to avoid forgetting the global map.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Mapping: Isotropic Regularization

- Why do we need “isotropic regularization”?

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

- Why do we need "isotropic regularization"?
  - Observation: isotropic Gaussians behave better than anisotrophic.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Mapping: Isotropic Regularization

- Why do we need "isotropic regularization"?
  - Observation: isotropic Gaussians behave better than anisotrophic.
  - Analysis: no constraints on the elongation along the viewing ray direction, **even with depth**.

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Mapping: Isotropic Regularization

- Why do we need "isotropic regularization"?
  - Observation: isotropic Gaussians behave better than anisotrophic.
  - Analysis: no constraints on the elongation along the viewing ray direction, **even with depth**.

Isotropic Regularization

$$\mathcal{L}_{iso} = \sum_{i=1}^{|\mathcal{G}|} \|\mathbf{s}_i - \bar{\mathbf{s}}_i\|_1, \quad \text{where } \bar{\mathbf{s}}_i = \frac{1}{3}\left(s_i^x + s_i^y + s_i^z\right). \tag{13}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Mapping: Conclusion

The Overall Optimization for Mapping

$$\underset{\mathcal{G},\{\mathbf{T}_{cw}(\mathcal{F}_k)|\mathcal{F}_k\in\mathcal{W}^+\}}{\operatorname{argmin}} \sum_{\mathcal{F}_k}^{\mathcal{W}^+} \mathcal{L}_{pho}\left(\mathcal{F}_k\right) + \lambda_{iso}\mathcal{L}_{iso} \tag{14}$$

(CVPR Highlight, 2024) MonoGS: Gaussian Splatting SLAM

Appendix

References i

[1] N. Keetha, J. Karhade, K. M. Jatavallabhula, *et al.*, *SplaTAM: Splat, track & map 3d gaussians for dense RGB-d SLAM*, Apr. 16, 2024. arXiv: 2312.02126[cs]. [Online]. Available: http://arxiv.org/abs/2312.02126 (visited on 05/20/2024) (cit. on p. iv).

[2] C. Yan, D. Qu, D. Wang, *et al.*, *GS-SLAM: Dense visual SLAM with 3d gaussian splatting*, Nov. 21, 2023. arXiv: 2311.11700[cs]. [Online]. Available: http://arxiv.org/abs/2311.11700 (visited on 12/26/2023) (cit. on p. iv).

[3] V. Yugay, Y. Li, T. Gevers, and M. R. Oswald, *Gaussian-SLAM: Photo-realistic dense SLAM with gaussian splatting*, Mar. 22, 2024. arXiv: 2312.10070[cs]. [Online]. Available: http://arxiv.org/abs/2312.10070 (visited on 03/27/2024) (cit. on p. iv).

[4] H. Matsuki, R. Murai, P. H. J. Kelly, and A. J. Davison, *Gaussian splatting SLAM*, Apr. 14, 2024. arXiv: 2312.06741[cs]. [Online]. Available: http://arxiv.org/abs/2312.06741 (visited on 05/20/2024) (cit. on p. iv).

[5] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," in *arXiv:1607.02565*, Jul. 2016 (cit. on pp. xxiv–xxvi).