Submap-based Bundle Adjustment for 3D Reconstruction from RGB-D Data

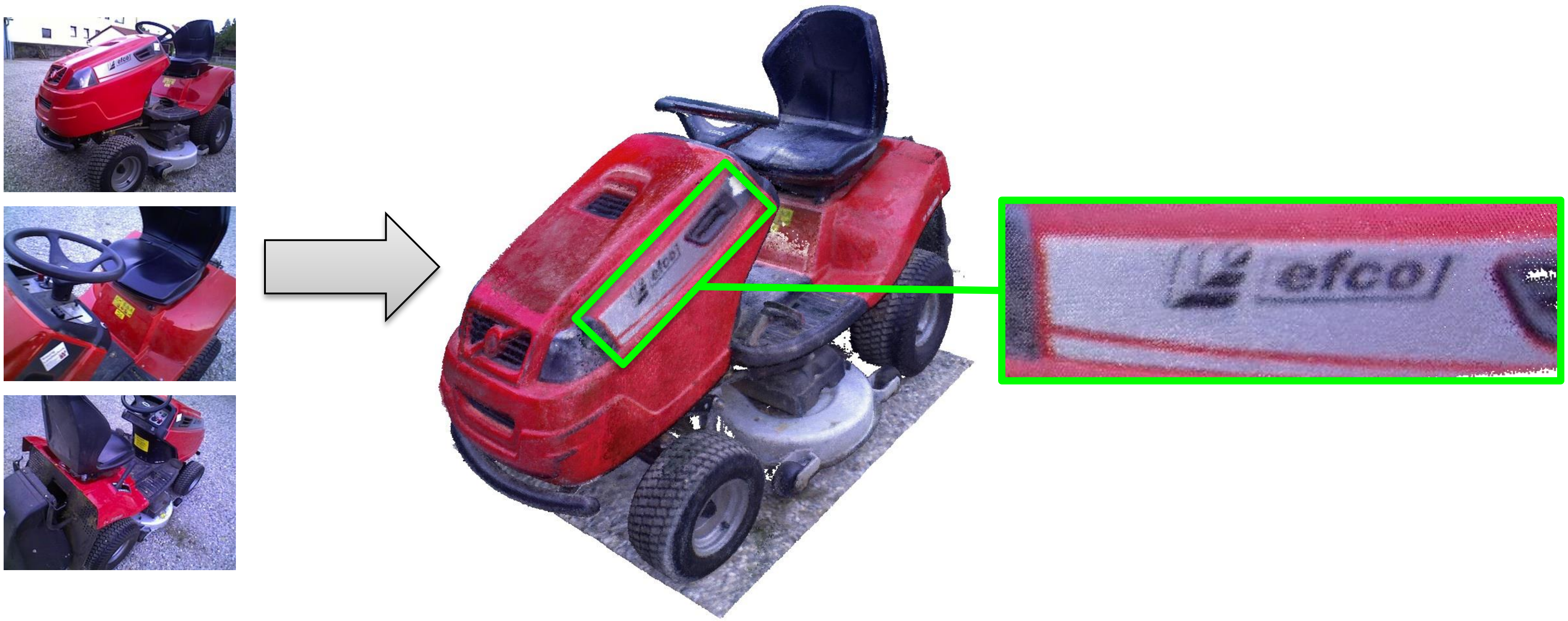
Robert Maier, Jürgen Sturm, Daniel Cremers

German Conference on Pattern Recognition (GCPR) 2014

September 3, 2014

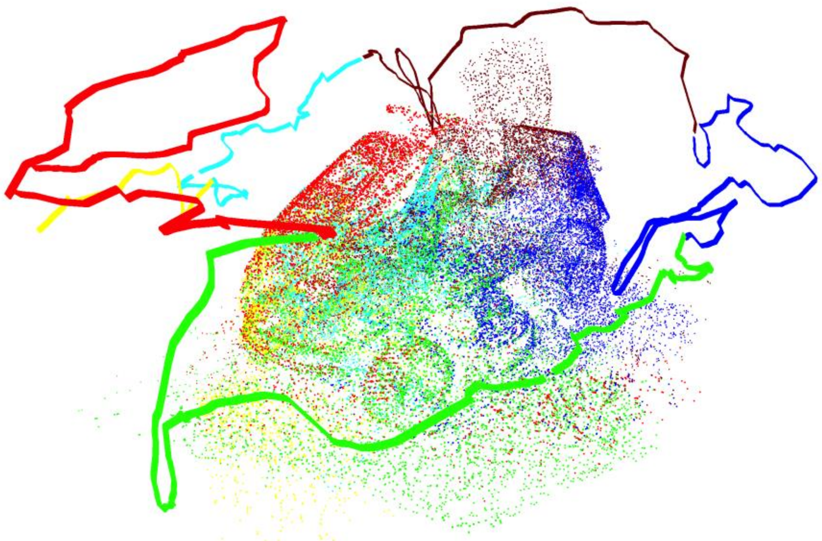
Motivation

- Given: Low-cost RGB-D sensors
- Wanted: 3D reconstruction of highly accurate 3D models (e.g. for reverse-engineering)



Submap-based Bundle Adjustment

- Problem:
  - Incremental tracking and mapping methods prone to drift
  - Full bundle adjustment (BA) too slow
- Our solution: Novel submap-based BA method for RGB-D based 3D reconstruction



Related Work

Related Work

- RGB-D SLAM systems
  - An evaluation of the RGB-D SLAM system [Endres et al., ICRA 2012]
  - RGB-D mapping: Using Kinect-Style Depth Cameras for Dense 3D Modeling of Indoor Environments [Henry et al., IJRR 2012]
  - Using depth in visual simultaneous localisation and mapping [Scherer et al., ICRA 2012]

Pose Graph Optimization

Sparse Bundle Adjustment

3D Bundle Adjustment

Related Work

- RGB-D SLAM systems
  - An evaluation of the RGB-D SLAM system [Endres et al., ICRA 2012]
  - RGB-D mapping: Using Kinect-Style Depth Cameras for Dense 3D Modeling of Indoor Environments [Henry et al., IJRR 2012]
  - Using depth in visual simultaneous localisation and mapping [Scherer et al., ICRA 2012]

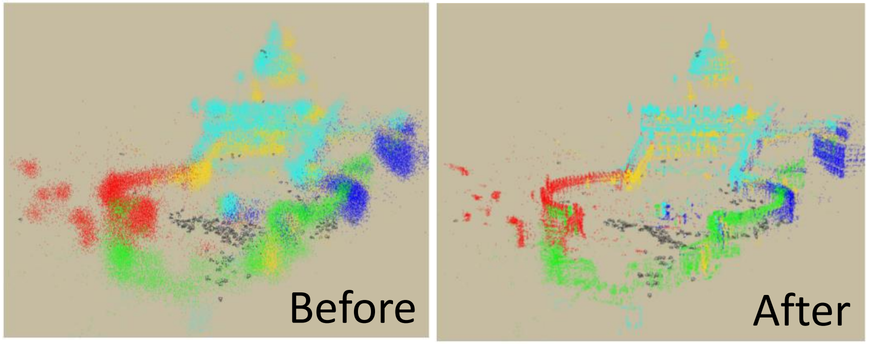
Pose Graph Optimization

Sparse Bundle Adjustment

3D Bundle Adjustment

- Out-of-core bundle adjustment for large-scale 3D reconstruction [Ni et al., ICCV 2007]

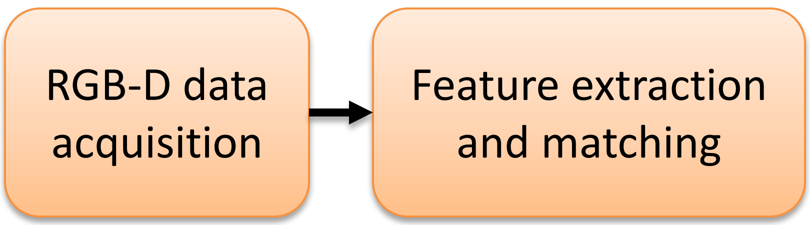
Submap-based Bundle Adjustment



Feature-based 3D Reconstruction System

RGB-D data acquisition

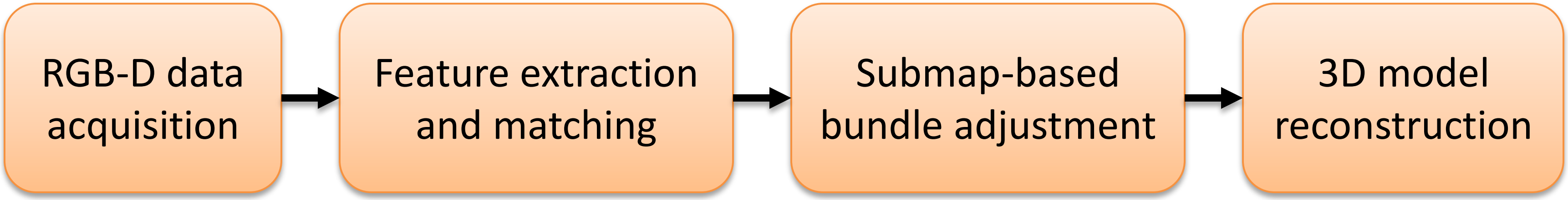
Feature-based 3D Reconstruction System



Feature-based 3D Reconstruction System

RGB-D data acquisition → Feature extraction and matching → Submap-based bundle adjustment

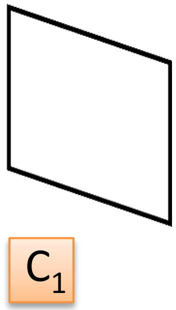
Feature-based 3D Reconstruction System



Graph-based map representation

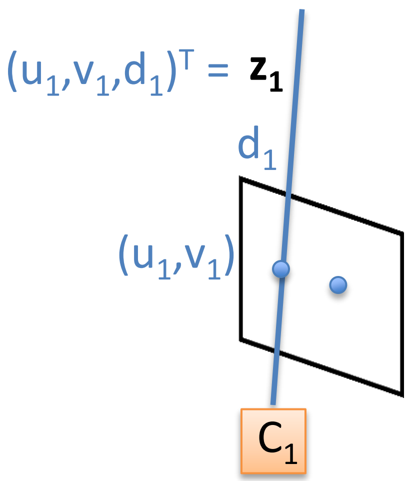
Graph-based map representation

- Camera poses $C_i \in SE(3)$ (with $i \in 1 \ldots M$)



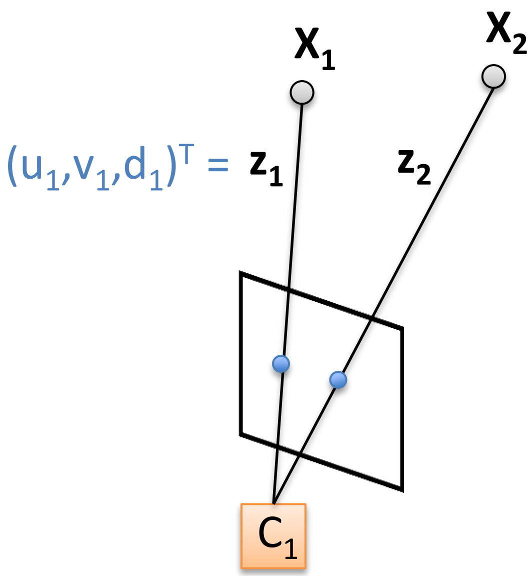
Graph-based map representation

- Camera poses $C_i \in SE(3)$ (with $i \in 1 \ldots M$)
- Landmark observations $\mathbf{z}_k = (u_k, v_k, d_k)^\top \in \mathbb{R}^3$ (with $k \in 1 \ldots K$)



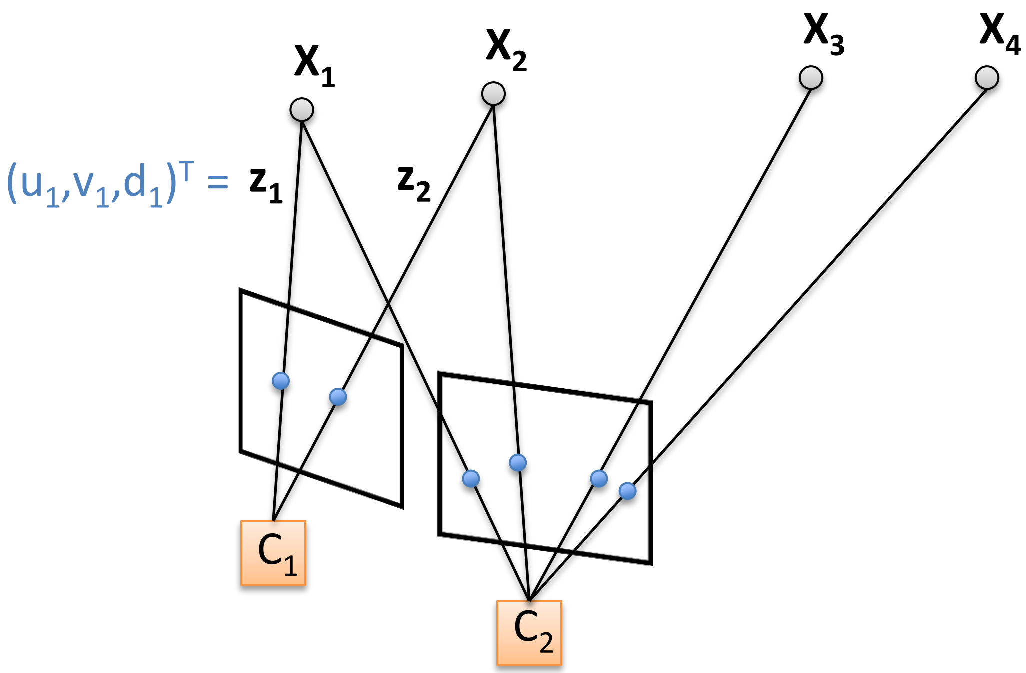
Graph-based map representation

- Camera poses $C_i \in SE(3)$ (with $i \in 1 \ldots M$)
- Landmark observations $\mathbf{z}_k = (u_k, v_k, d_k)^\top \in \mathbb{R}^3$ (with $k \in 1 \ldots K$)
- 3D landmarks $\mathbf{X}_j \in \mathbb{R}^3$ (with $j \in 1 \ldots N$)



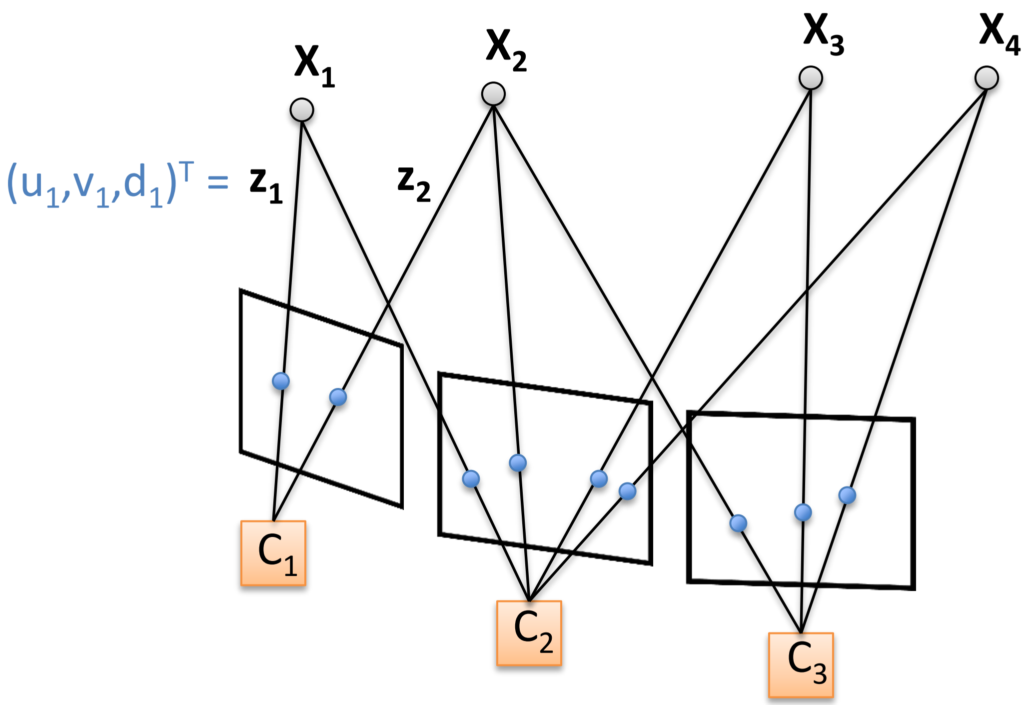
Graph-based map representation

- Camera poses $C_i \in SE(3)$ (with $i \in 1 \ldots M$)
- Landmark observations $\mathbf{z}_k = (u_k, v_k, d_k)^\top \in \mathbb{R}^3$ (with $k \in 1 \ldots K$)
- 3D landmarks $\mathbf{X}_j \in \mathbb{R}^3$ (with $j \in 1 \ldots N$)



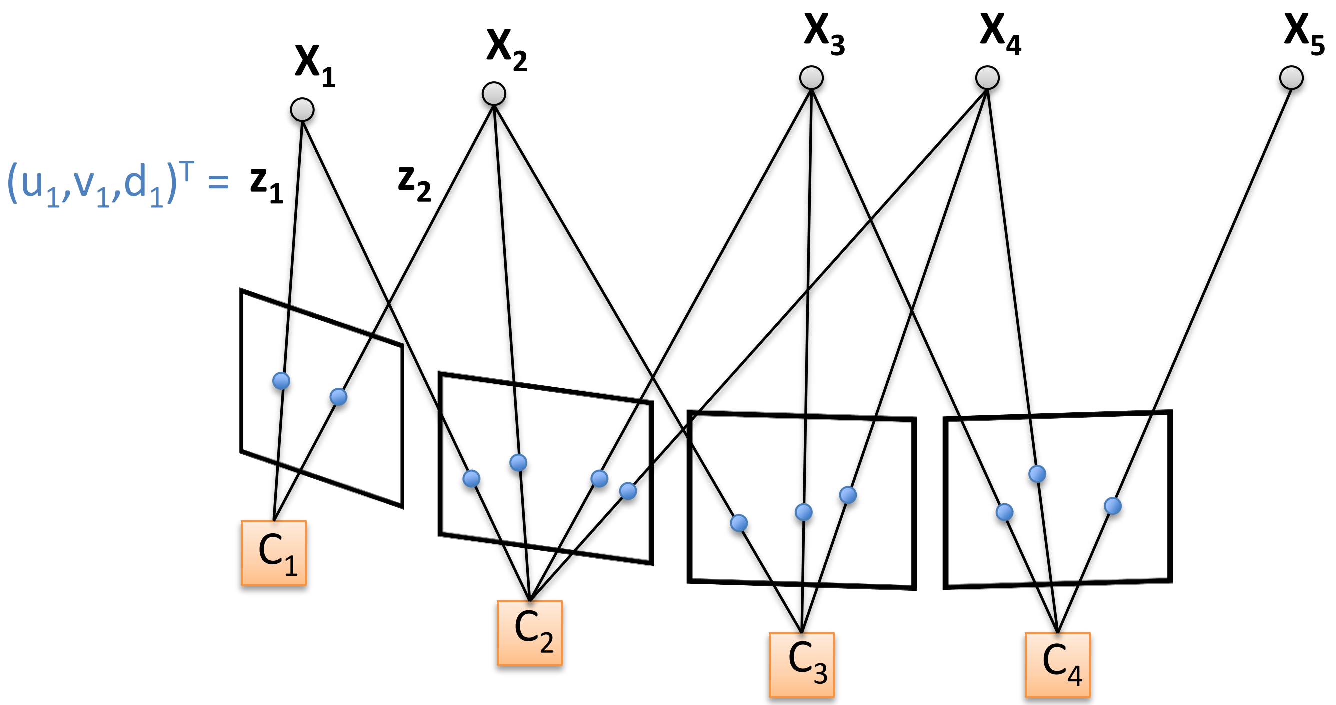
Graph-based map representation

- Camera poses $C_i \in SE(3)$ (with $i \in 1 \dots M$)
- Landmark observations $\mathbf{z}_k = (u_k, v_k, d_k)^\top \in \mathbb{R}^3$ (with $k \in 1 \dots K$)
- 3D landmarks $\mathbf{X}_j \in \mathbb{R}^3$ (with $j \in 1 \dots N$)



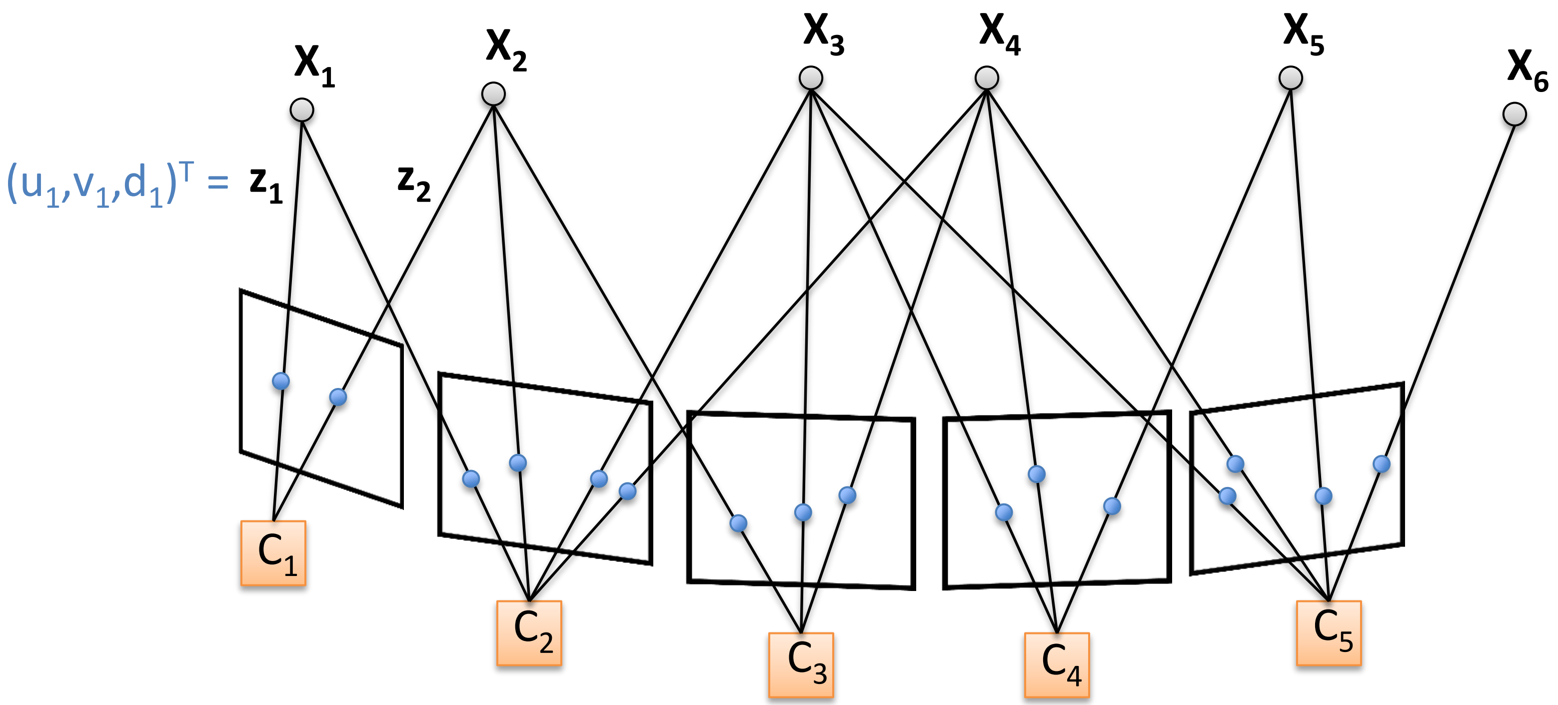
Graph-based map representation

- Camera poses $C_i \in SE(3)$ (with $i \in 1 \ldots M$)
- Landmark observations $\mathbf{z}_k = (u_k, v_k, d_k)^\top \in \mathbb{R}^3$ (with $k \in 1 \ldots K$)
- 3D landmarks $\mathbf{X}_j \in \mathbb{R}^3$ (with $j \in 1 \ldots N$)



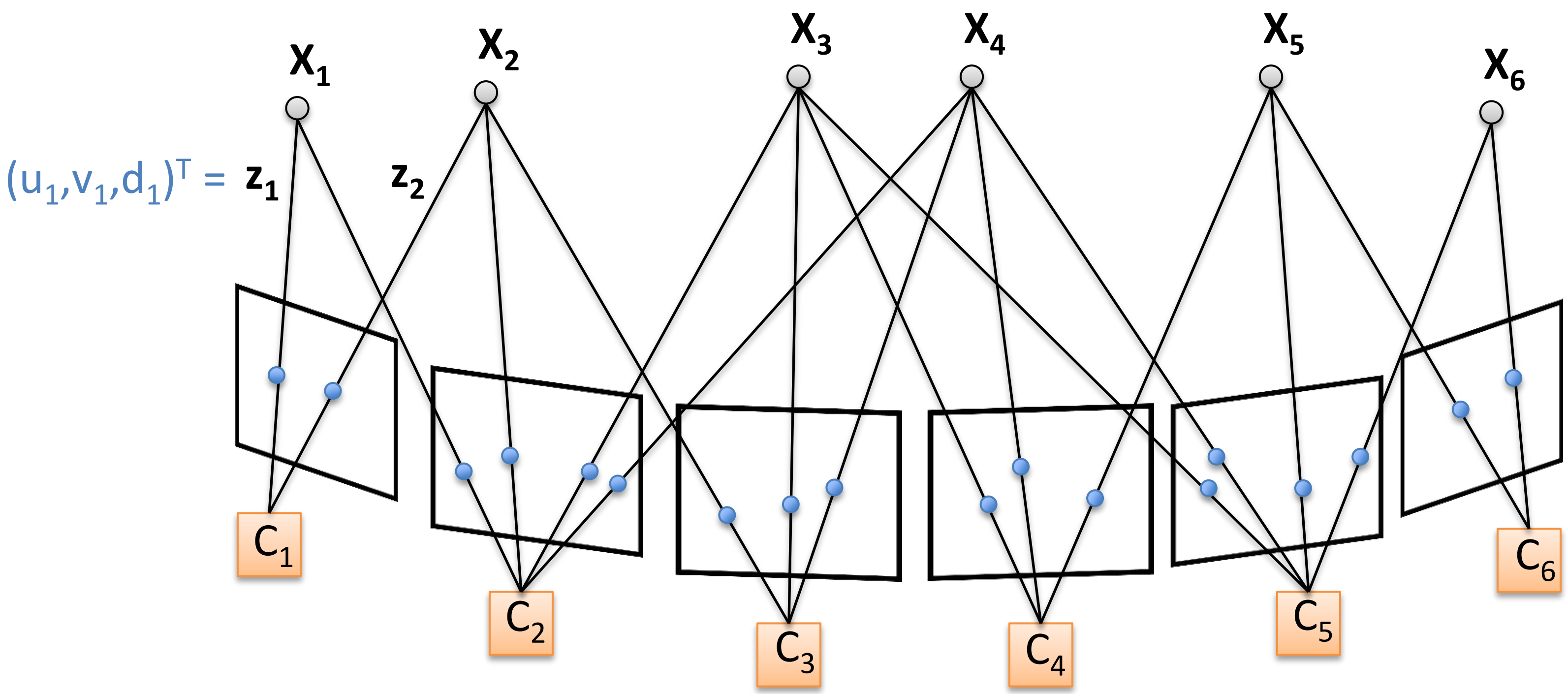
Graph-based map representation

- Camera poses $C_i \in SE(3)$ (with $i \in 1 \dots M$)
- Landmark observations $\mathbf{z}_k = (u_k, v_k, d_k)^\top \in \mathbb{R}^3$ (with $k \in 1 \dots K$)
- 3D landmarks $\mathbf{X}_j \in \mathbb{R}^3$ (with $j \in 1 \dots N$)



Graph-based map representation

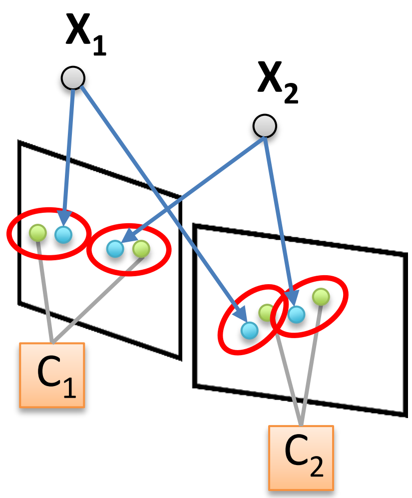
- Camera poses $C_i \in SE(3)$ (with $i \in 1 \ldots M$)
- Landmark observations $\mathbf{z}_k = (u_k, v_k, d_k)^\top \in \mathbb{R}^3$ (with $k \in 1 \ldots K$)
- 3D landmarks $\mathbf{X}_j \in \mathbb{R}^3$ (with $j \in 1 \ldots N$)



Full Bundle Adjustment for RGB-D Sensors

2D reprojection error

$$\min_{C_{i_k}, \mathbf{X}_{j_k}} \sum_{k=1}^{K} ||\pi(\mathcal{T}^{-1}(C_{i_k}, \mathbf{X}_{j_k})) - (u_k, v_k)^\top||^2$$



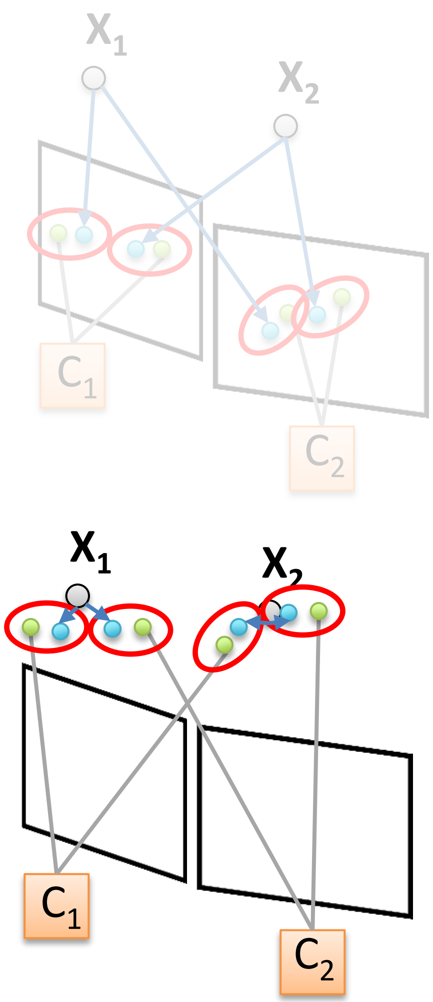
Full Bundle Adjustment for RGB-D Sensors

2D reprojection error

$$\min_{C_{i_k}, \mathbf{X}_{j_k}} \sum_{k=1}^{K} ||\pi(\mathcal{T}^{-1}(C_{i_k}, \mathbf{X}_{j_k})) - (u_k, v_k)^\top||^2$$

3D alignment error

$$\min_{C_{i_k}, \mathbf{X}_{j_k}} \sum_{k=1}^{K} ||\mathcal{T}^{-1}(C_{i_k}, \mathbf{X}_{j_k}) - \rho(u_k, v_k, d_k)||^2$$



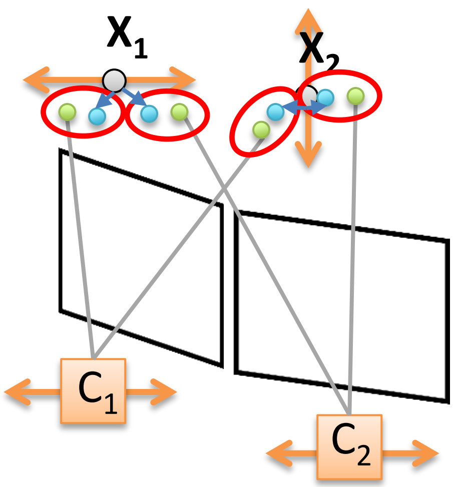
Full Bundle Adjustment for RGB-D Sensors

2D reprojection error

$$\min_{C_{i_k}, \mathbf{X}_{j_k}} \sum_{k=1}^{K} ||\pi(\mathcal{T}^{-1}(C_{i_k}, \mathbf{X}_{j_k})) - (u_k, v_k)^\top||^2$$

3D alignment error

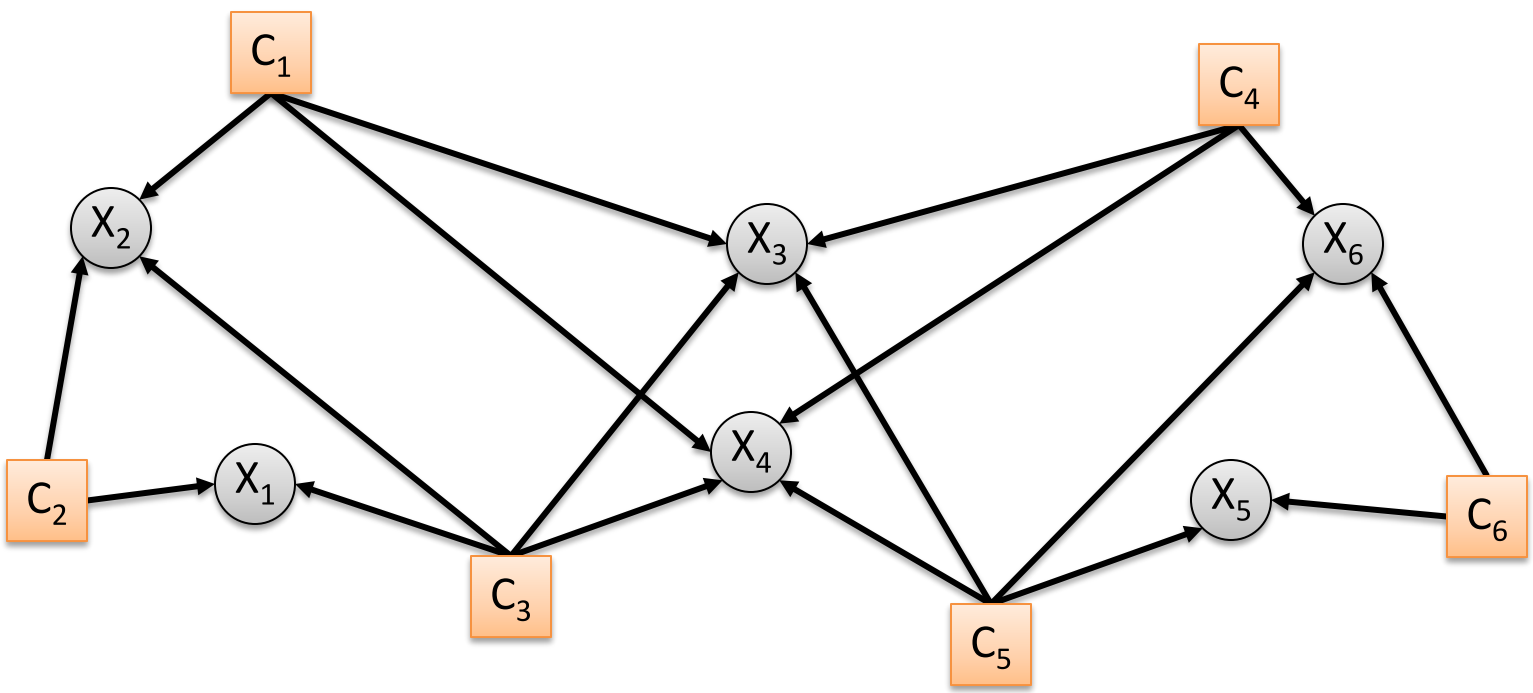
$$\min_{C_{i_k}, \mathbf{X}_{j_k}} \sum_{k=1}^{K} ||\mathcal{T}^{-1}(C_{i_k}, \mathbf{X}_{j_k}) - \rho(u_k, v_k, d_k)||^2$$



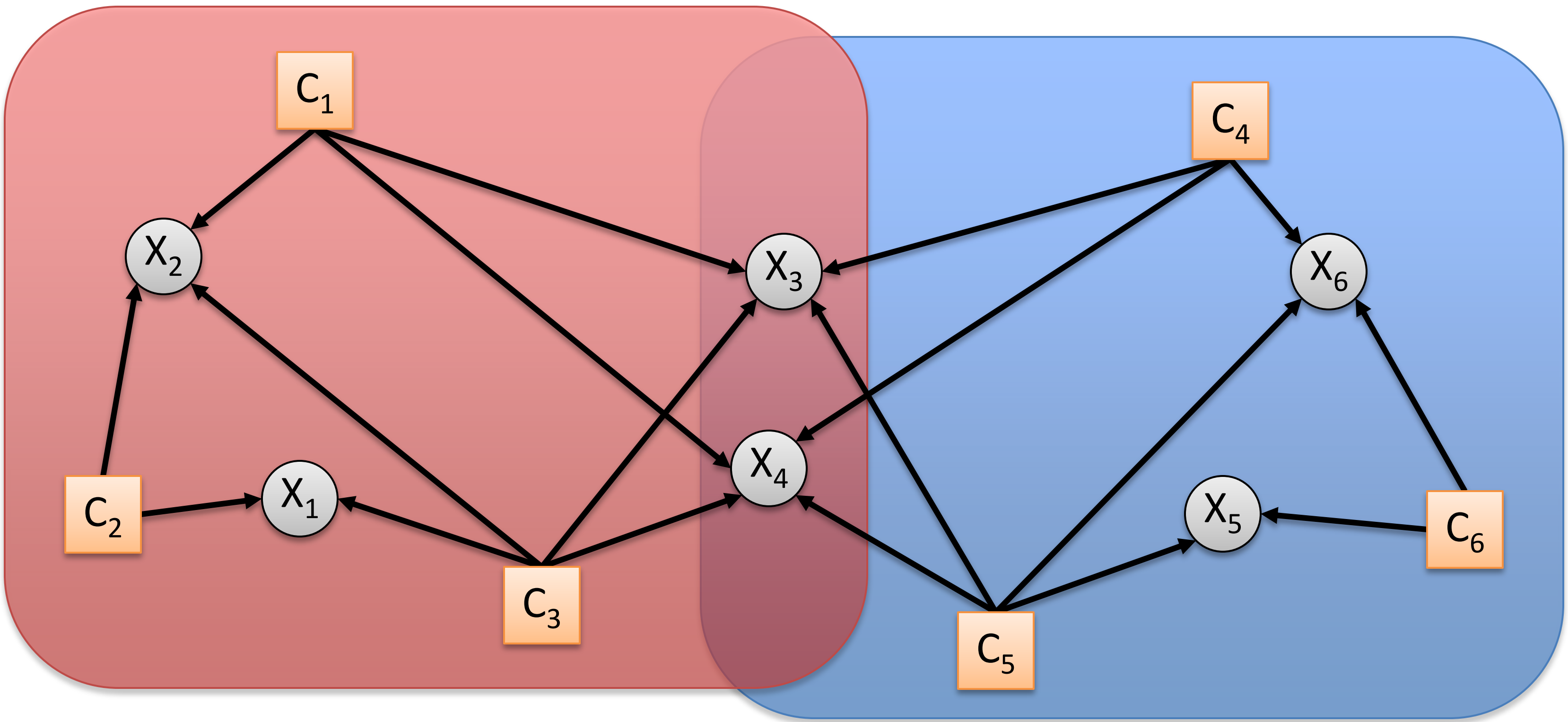
Efficient Bundle Adjustment for RGB-D Sensors using Submapping

1. Graph partitioning into submaps
2. Submap optimization
3. Global submaps alignment
4. Submap optimization with fixed separator

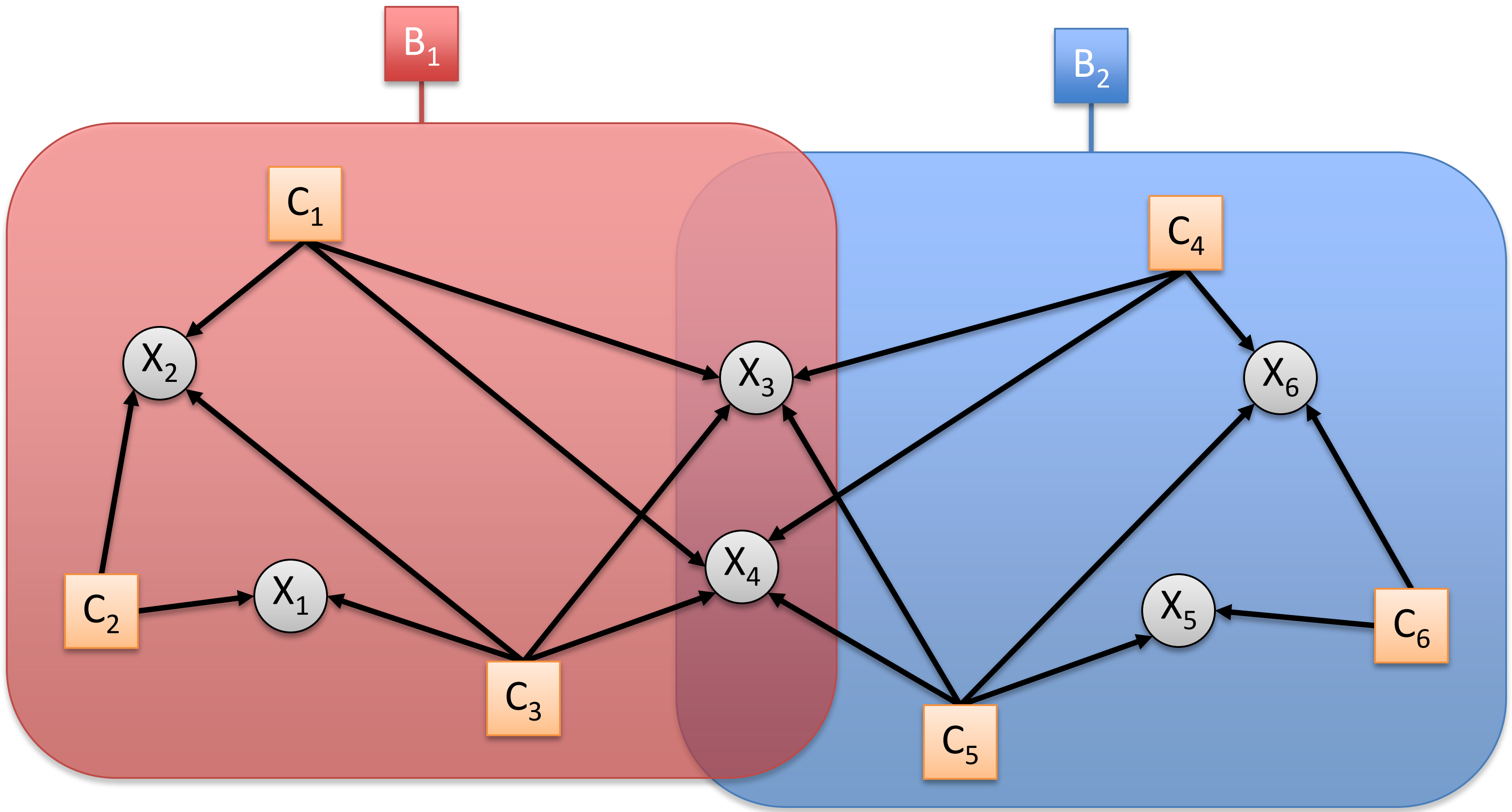
Stage 1: Graph partitioning into submaps



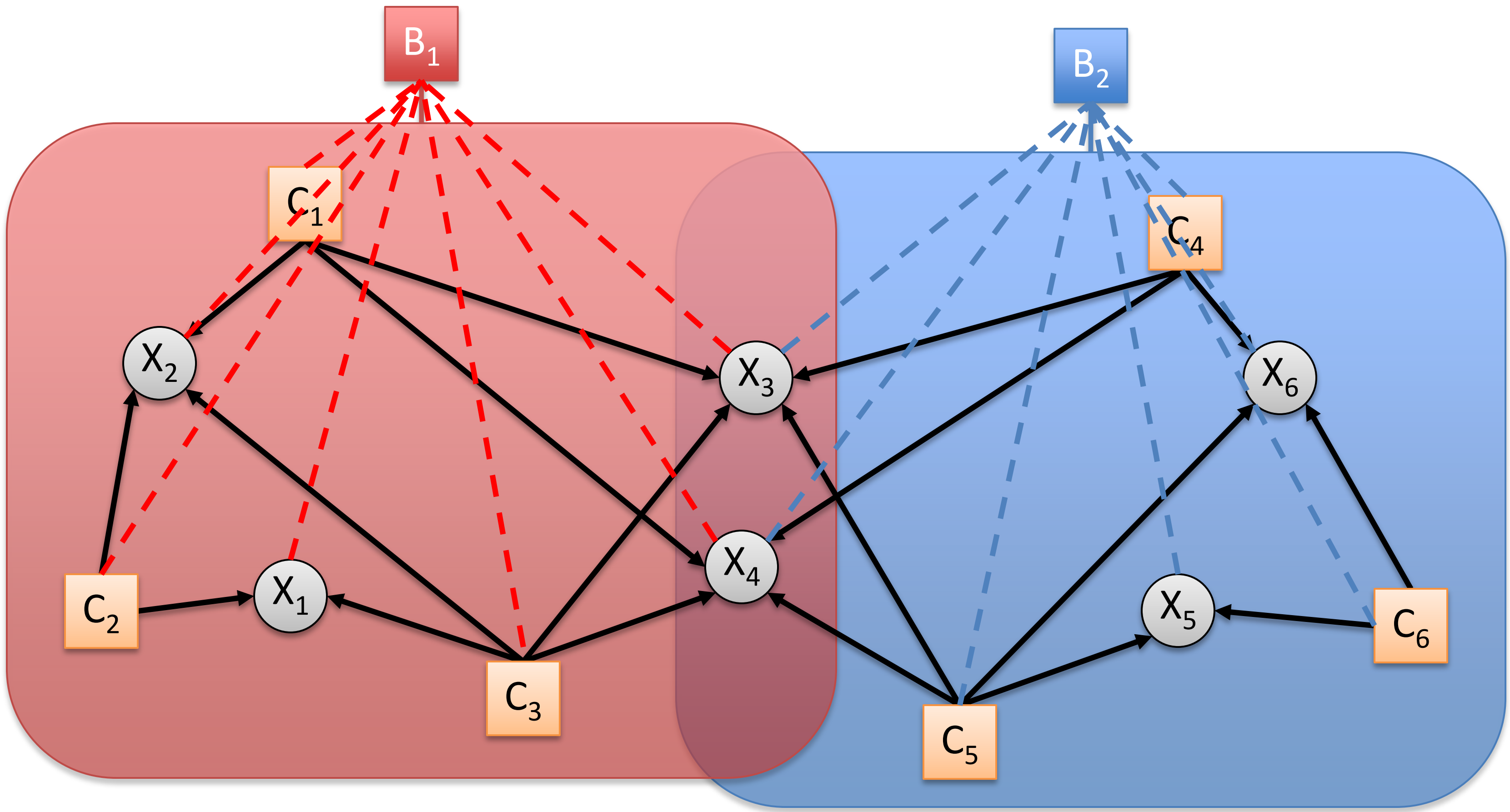
Stage 1: Graph partitioning into submaps



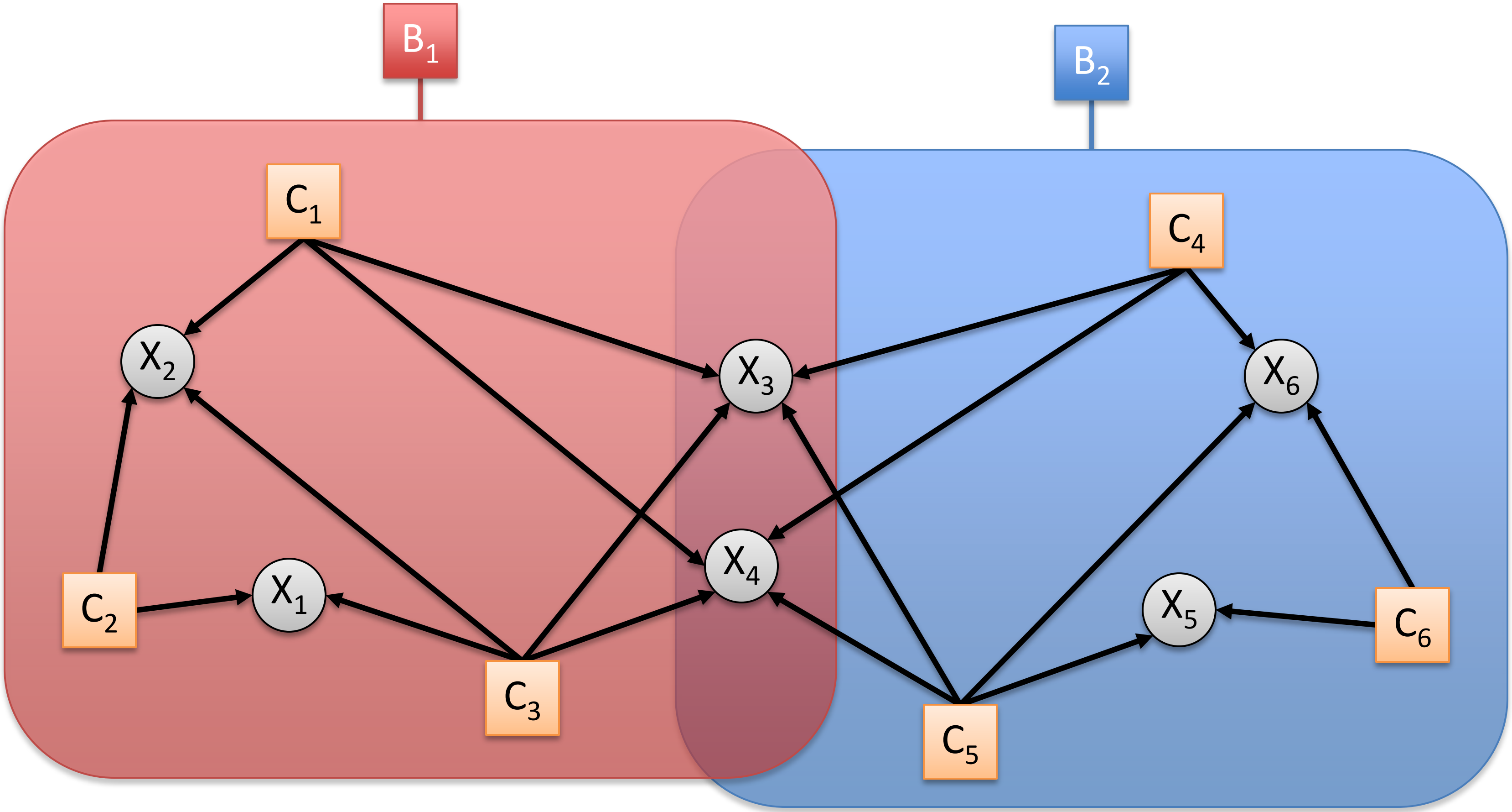
Stage 1: Graph partitioning into submaps



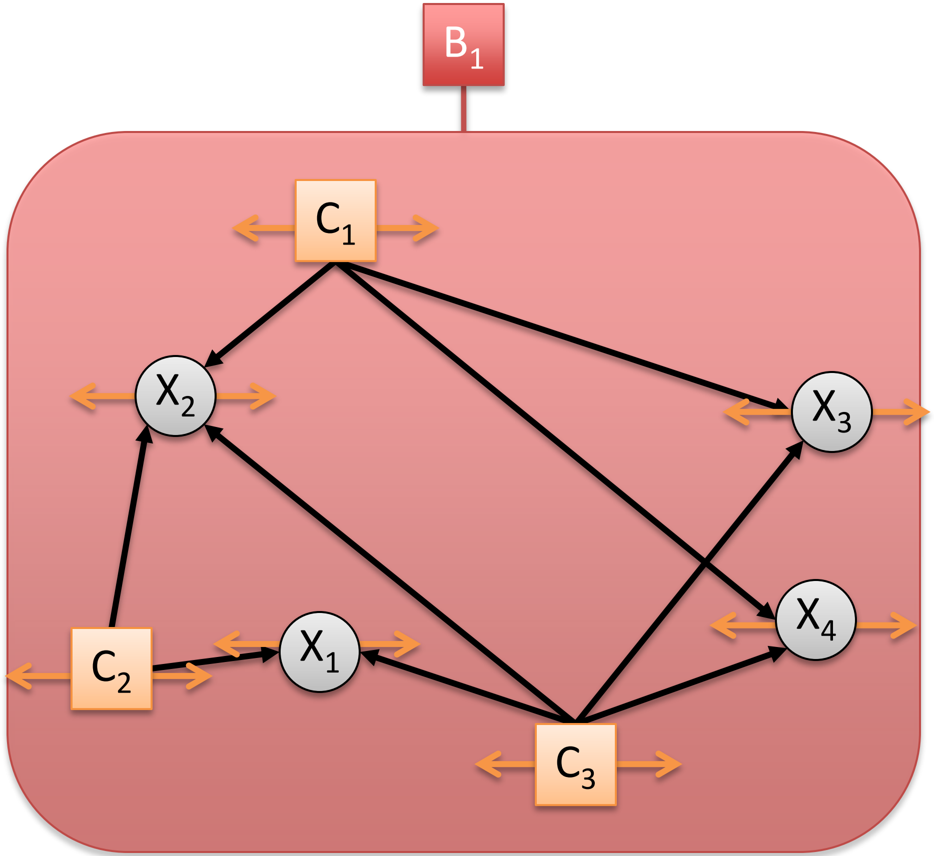
Stage 1: Graph partitioning into submaps



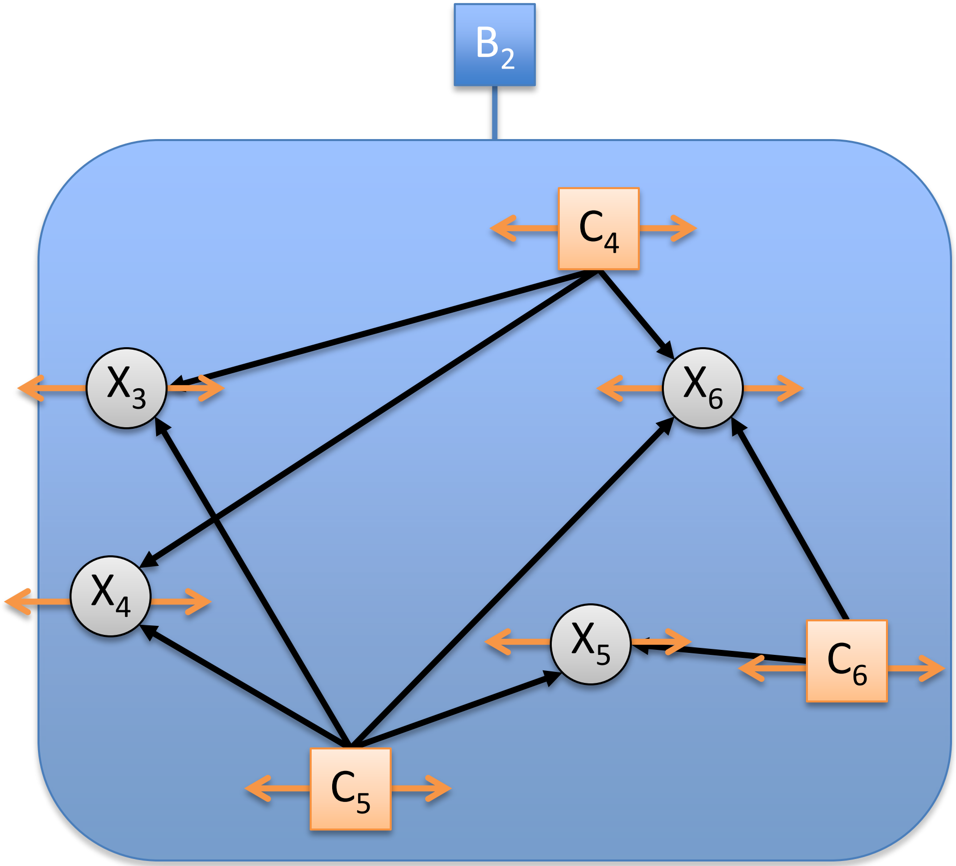
Stage 2: Submap optimization



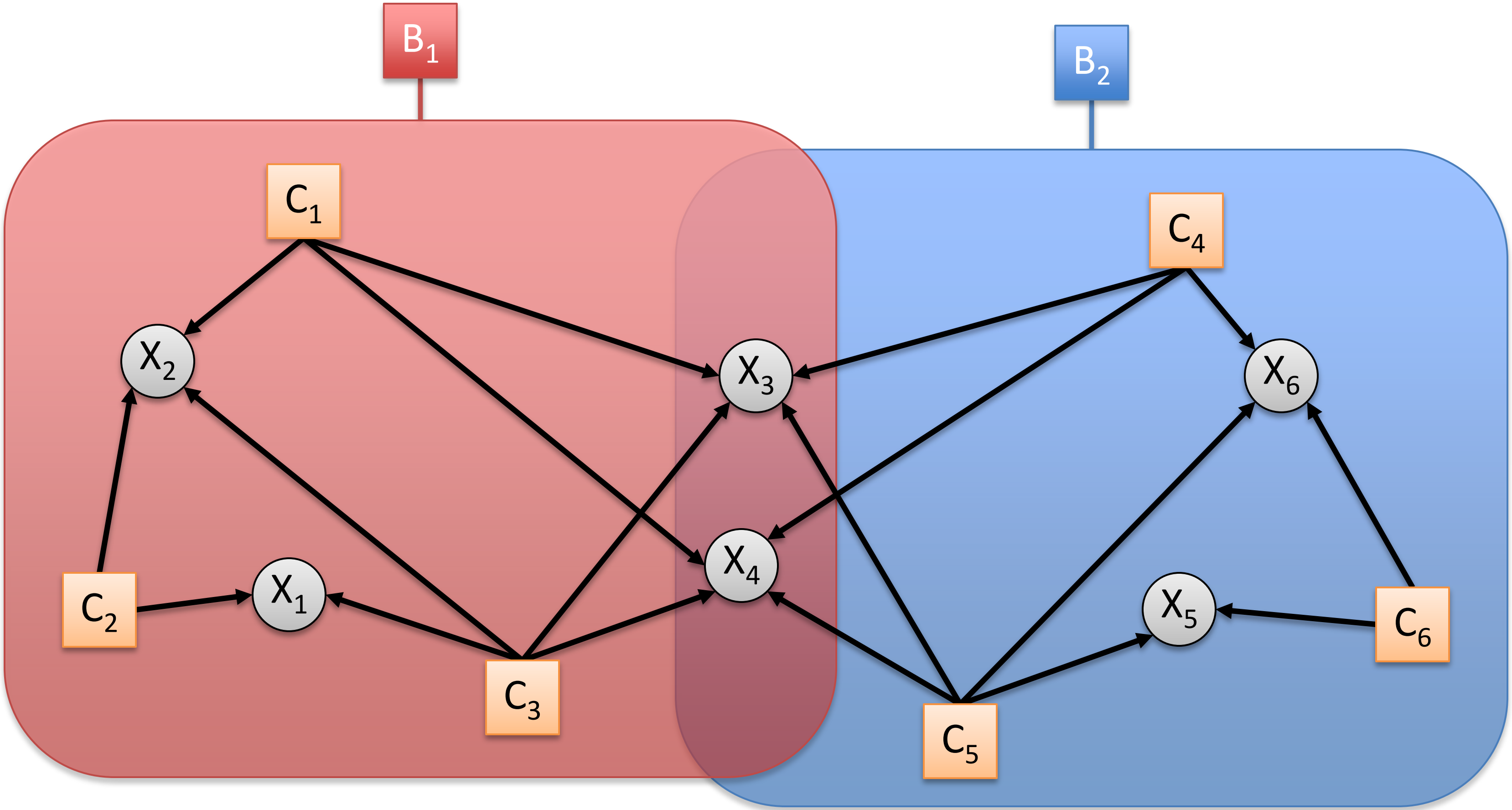
Stage 2: Submap optimization



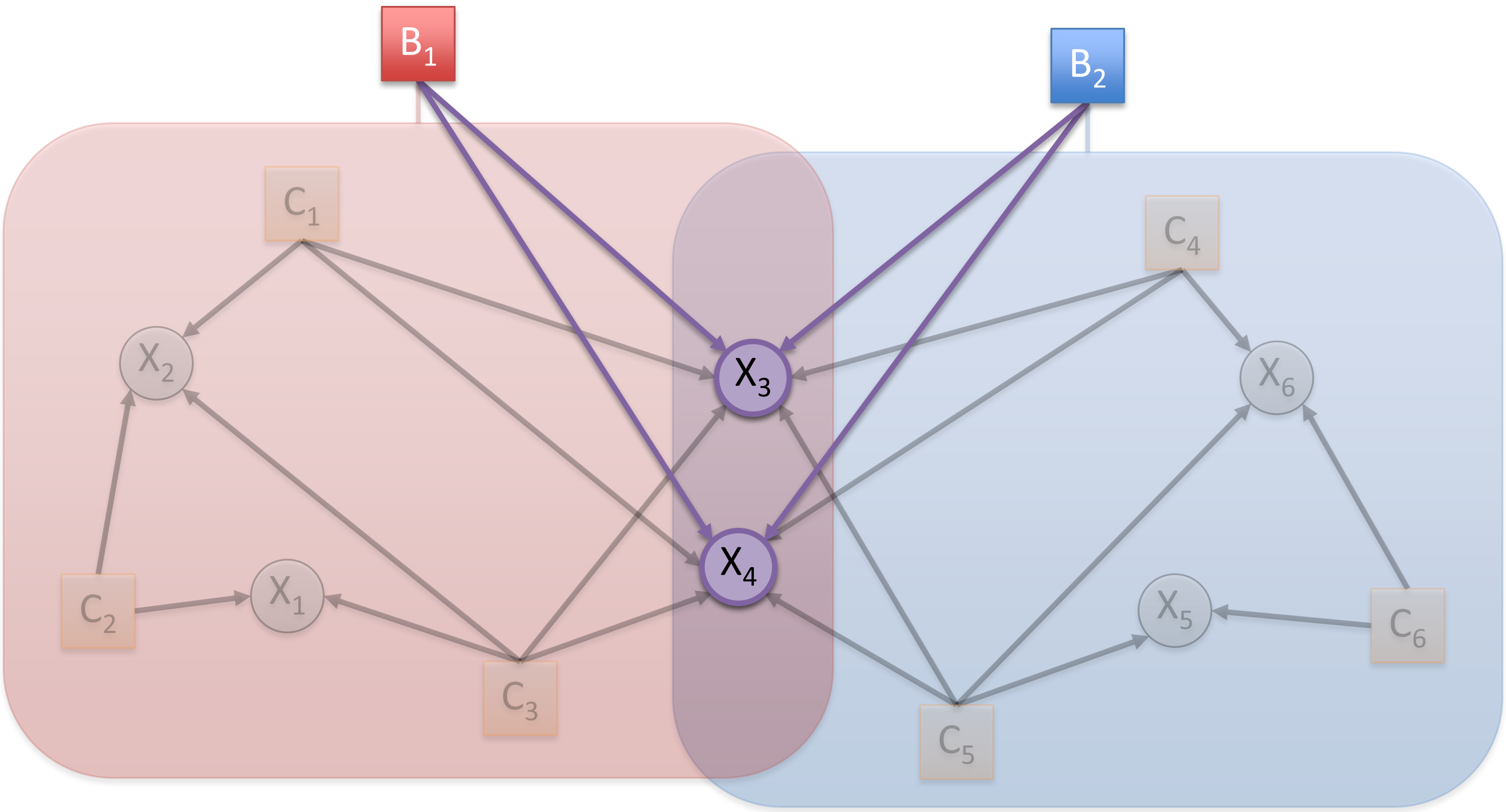
Stage 2: Submap optimization



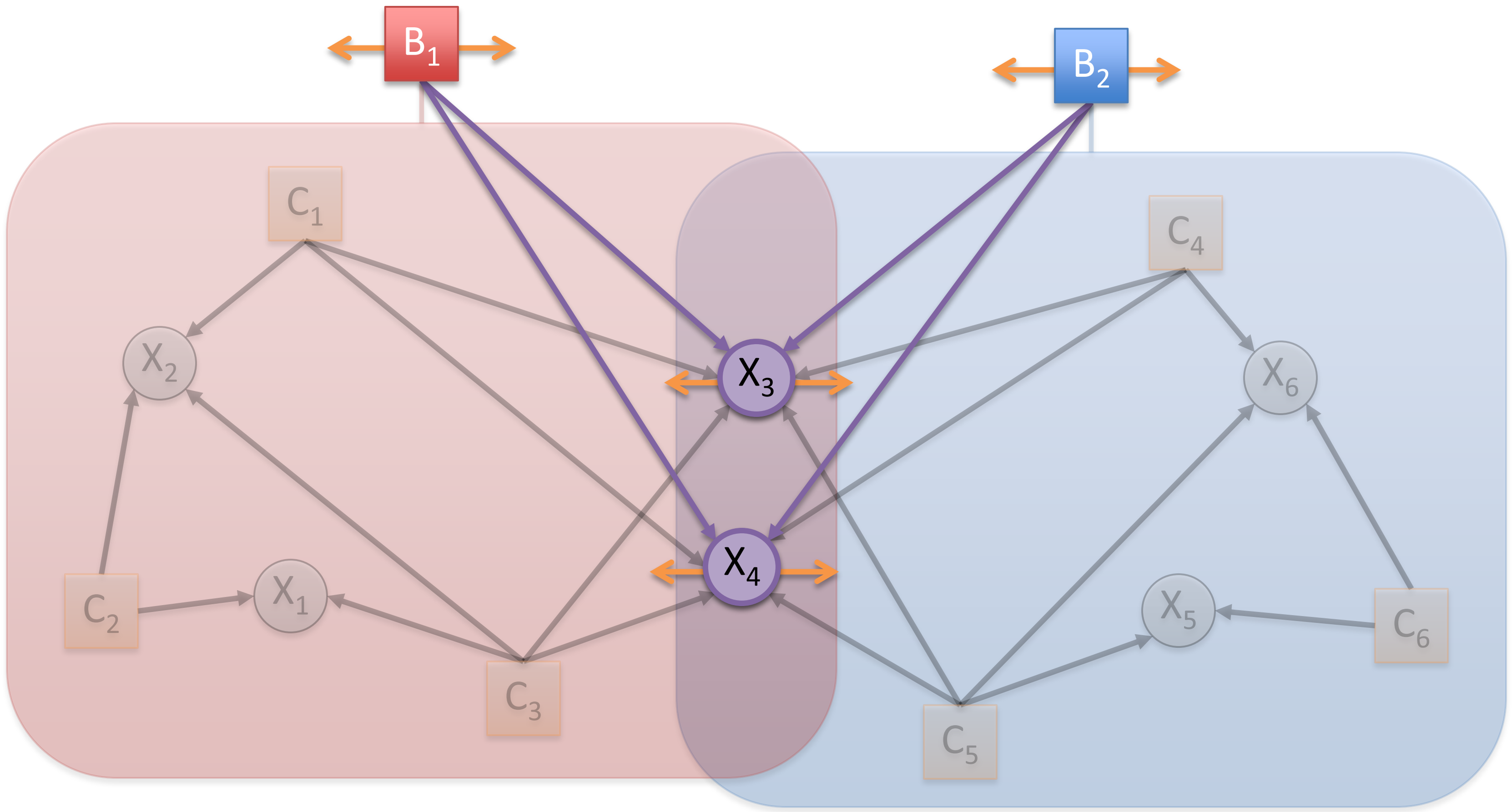
Stage 3: Global submaps alignment



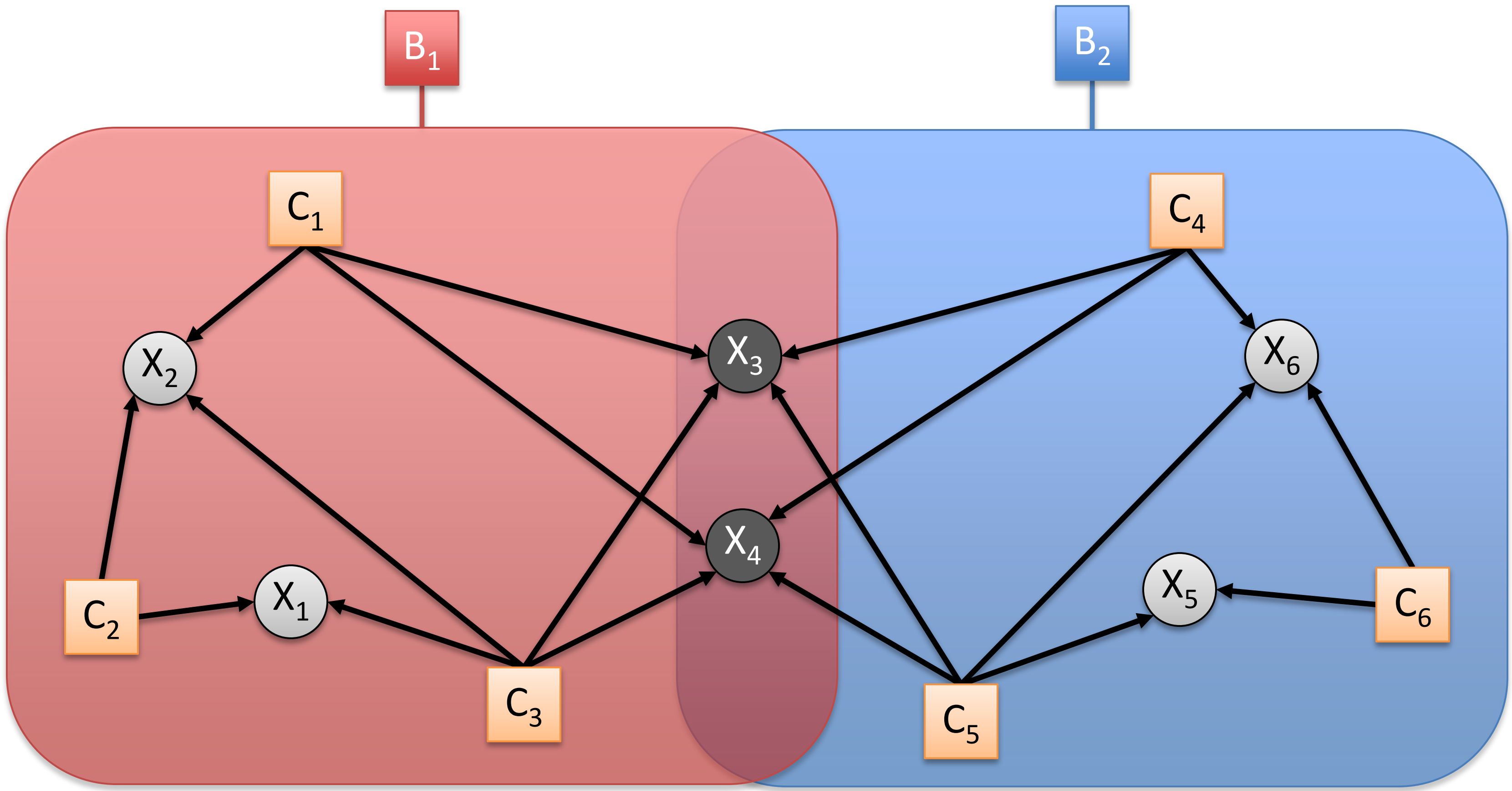
Stage 3: Global submaps alignment



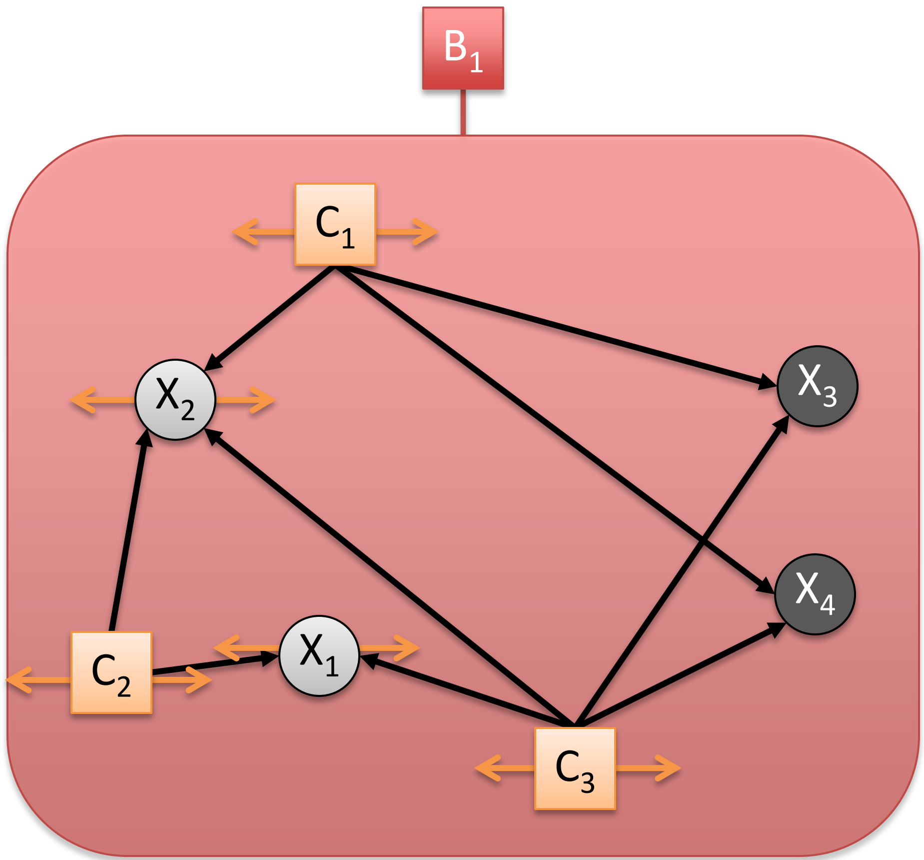
Stage 3: Global submaps alignment



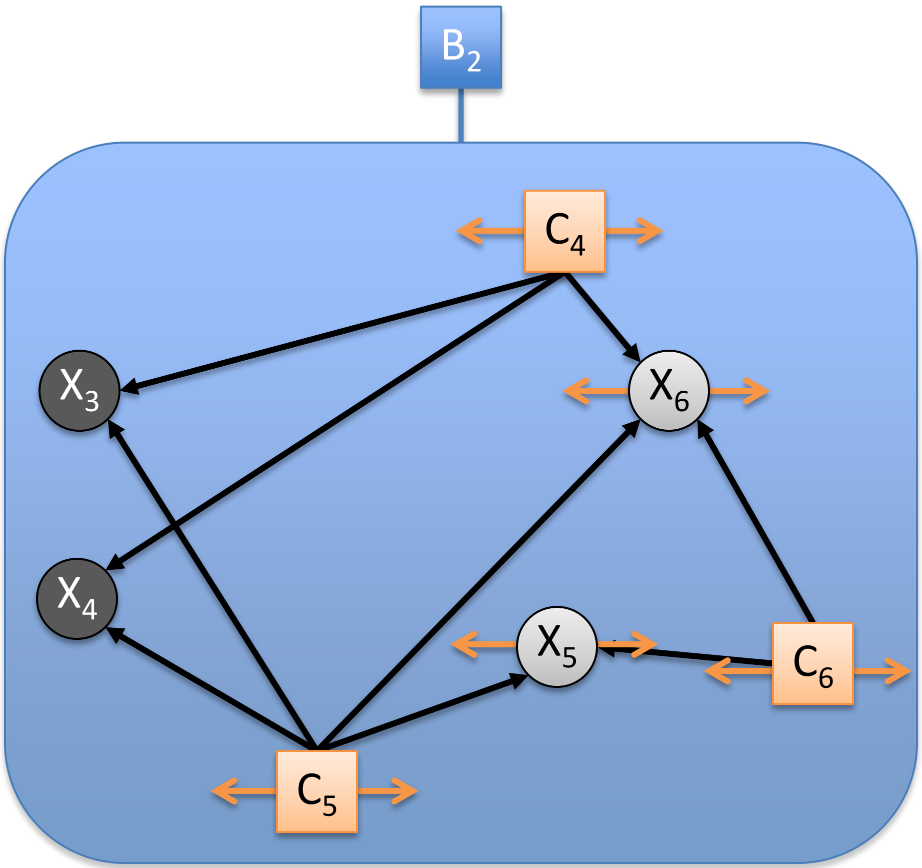
Stage 4: Internal submap update



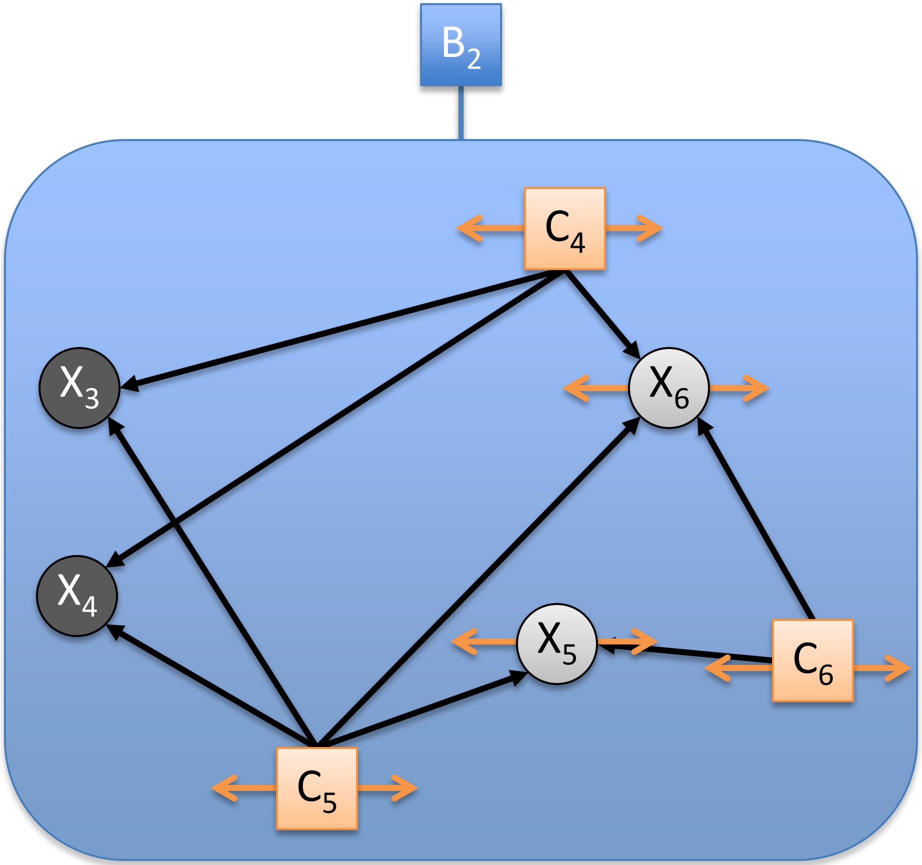
Stage 4: Internal submap update



Stage 4: Internal submap update



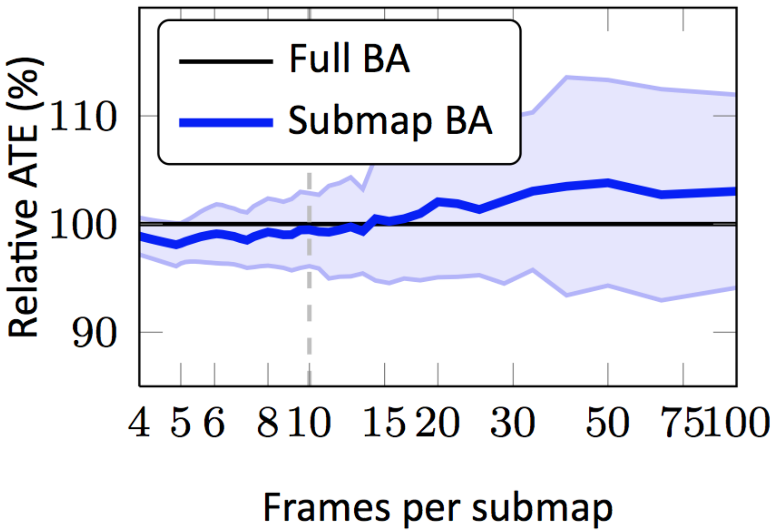
Stage 4: Internal submap update

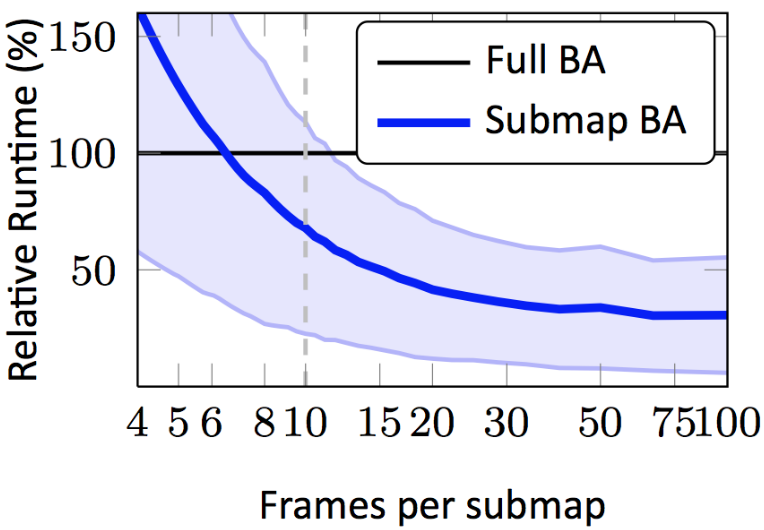


➔ Use final camera poses to fuse RGB-D frames into 3D octree model

Evaluation: Size of Submaps

- Evaluation of Absolute Trajectory Error (ATE) over 10 sequences of TUM RGB-D benchmark [Sturm et al., IROS 2012]





- Small submaps: smaller ATE than full BA
- Large submaps: increase efficiency but decrease accuracy
- Good speed/accuracy trade-off: 10 frames per submap

Evaluation: Performance

- Benchmark sequences (4 of 10 sequences):

Sequence	No BA	Full 2D	Full 3D		Submap-based 3D BA				
	ATE	ATE	ATE	time	submaps	ATE	$\pm(\%)$	time	$\pm(\%)$
FR1/desk2	0.098	0.044	0.030	27.23	62	0.031	+3.4	21.36	-21.5
FR1/room	0.275	0.228	0.085	125.46	135	0.086	+1.7	77.30	-38.4
FR2/desk	0.201	0.080	0.079	2355.26	289	0.076	-3.3	372.20	-84.2
FR3/office	0.176	0.039	0.036	1290.24	248	0.035	-3.0	242.88	-81.2
...	...	...	...	...	...	...	...	...	...
average	0.129	0.066	0.047			0.047	-0.5		-32.0

  - Similar accuracy as Full 3D BA at reduced cost (-32%)
  - Runtime improvement of up to 84% for long sequences

  - Comparison with state-of-the-art approaches:
    - RGB-D SLAM [Endres et al., ICRA 2012]: 13% (0.047m vs. 0.054m)
    - Direct SDF tracking [Bylow et al., RSS 2013]: 17% (0.047m vs. 0.058m)

Examples of Submap-based 3D Reconstructions

- Soil auger

Examples of Submap-based 3D Reconstructions

- Soil auger

Examples of Submap-based 3D Reconstructions

- Farm tractor

Examples of Submap-based 3D Reconstructions

- Farm tractor



Conclusion

- Our contribution: Submap-based bundle adjustment for RGB-D data
- Global optimization exploits available depth information

- Evaluation on benchmark datasets:
  - Accuracy similar to full bundle adjustment
  - Average runtime reduced by 32%
  - Higher accuracy than other state-of-the-art approaches
- Reconstructed 3D models: compelling visual quality and metric accuracy

Conclusion

- Our contribution: Submap-based bundle adjustment for RGB-D data
- Global optimization exploits available depth information

- Evaluation on benchmark datasets:
  - Accuracy similar to full bundle adjustment
  - Average runtime reduced by 32%
  - Higher accuracy than other state-of-the-art approaches
- Reconstructed 3D models: compelling visual quality and metric accuracy

Thank you!