Resource-Aware Large-Scale Cooperative 3D Mapping from Multiple Cell Phones

Chao X. Guo, Ryan C. DuToit, Kourosh Sartipi, Georgios Georgiou, Ruipeng Li, John O'Leary, Esha D. Nerurkar, Joel A. Hesch and Stergios I. Roumeliotis

Multiple Autonomous Robotic Systems (MARS) Lab - University of Minnesota

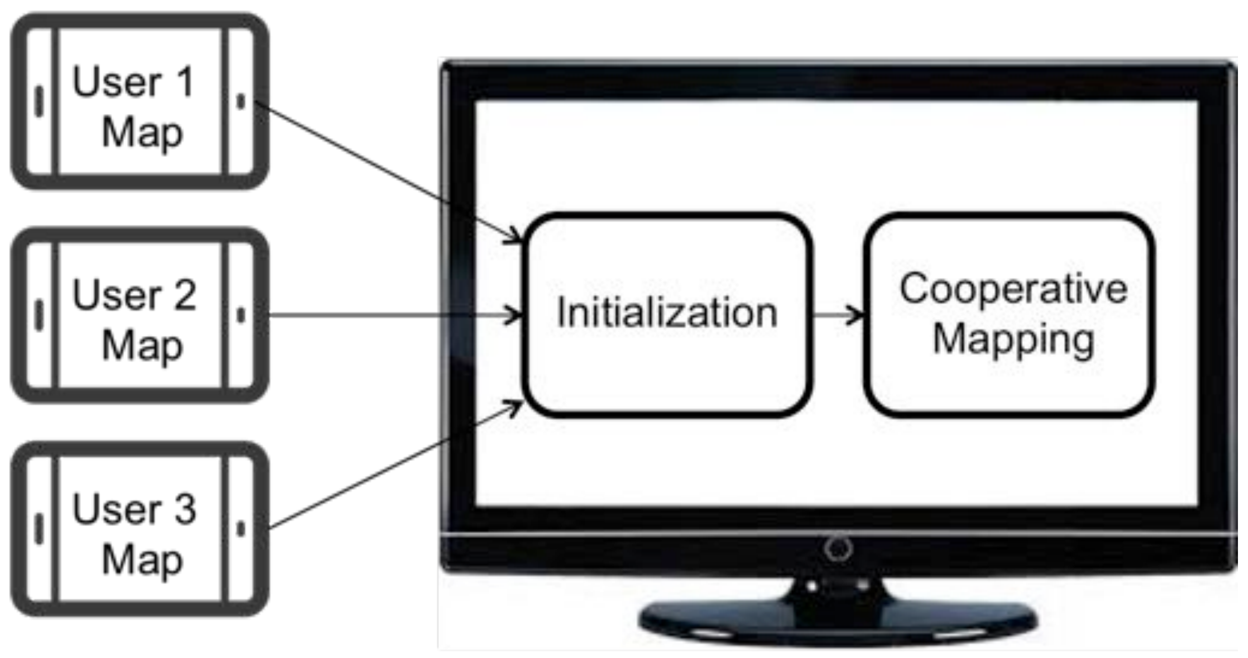
Problem Description

- Create a global map utilizing visual-inertial data collected independently by multiple users
- The relative poses between users/devices are unknown
- There exist common features between the users' maps

Main Contributions

- We formulate cooperative mapping (CM) as a constrained optimization problem, where the cost function comprises cost terms from each dataset, and constraints relate landmarks observed by multiple users. This formulation allows trading estimation accuracy for computational cost by selecting a subset of commonly-observed landmarks.
- The proposed algorithm is modular (i.e., maps can be added or removed), lends itself to parallel implementation, and is able to leverage each individual user's intermediate mapping results to reduce the processing cost.
- We provide a robust method for determining an initial estimate for the relative transformation between all users.

Algorithm Overview



The objective of this work is to find a ***batch least-squares (BLS) solution*** over all users' trajectories and maps. Our algorithm can be divided into three main steps:

- Obtain a BLS solution for each user's trajectory and map independently, using meas/nts from only their dataset.
- Generate an initial estimate of the users' relative poses, using their visual meas/nts to common landmarks.
- Find the optimal BLS solution of all users' trajectories and maps utilizing all available sensor data, and either a subset, or the entire set, of constraints that arise from commonly-observed landmarks.

Acknowledgements

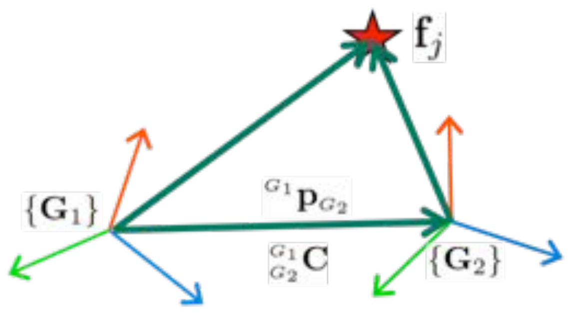
This work was supported by Google ATAP, Project Tango.

Initial Estimate of Users' Relative Poses

Find 4 d.o.f. transformation ${}^{G_1}\mathbf{p}_{G_2}$ and ${}^{G_1}_{G_2}\mathbf{C}$ between 2 users

Geometric constraints

$$ {}^{G_1}\mathbf{f}_1 = {}^{G_1}\mathbf{p}_{G_2} + {}^{G_1}_{G_2}\mathbf{C}\,{}^{G_2}\mathbf{f}_1 $$
$$ \vdots $$
$$ {}^{G_1}\mathbf{f}_M = {}^{G_1}\mathbf{p}_{G_2} + {}^{G_1}_{G_2}\mathbf{C}\,{}^{G_2}\mathbf{f}_M $$



Subtract the first constraint

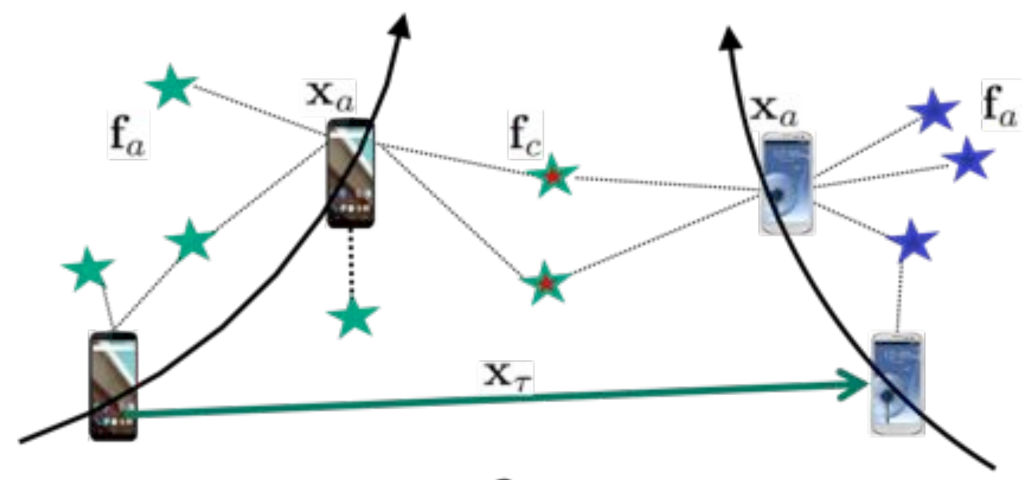
$$ {}^{G_1}\mathbf{f}_j - {}^{G_1}\mathbf{f}_1 = {}^{G_1}_{G_2}\mathbf{C}({}^{G_2}\mathbf{f}_j - {}^{G_2}\mathbf{f}_1) \quad \text{where } {}^{G_1}_{G_2}\mathbf{C} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} $$

Obtain ${}^{G_1}_{G_2}\mathbf{C}$ by solving

$$ \underset{\mathbf{v}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{v} - \mathbf{b}\| \quad \text{s.t. } \|\mathbf{v}\|^2 = 1 \qquad \text{where } \mathbf{v} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} $$

Then
$$ {}^{G_1}\mathbf{p}_{G_2} = \frac{1}{M}\sum_{j=1}^{M}\left({}^{G_1}\mathbf{f}_j - {}^{G_1}_{G_2}\mathbf{C}\,{}^{G_2}\mathbf{f}_j\right) $$

Cooperative Mapping



$\mathbf{x}_a$: all robot poses $\quad$ $\mathbf{f}_a$: feat. observed by only one user

$\mathbf{x}_\tau$: trans. between users $\quad$ $\mathbf{f}_c$: feat. observed by multiple users

Constrained optimization problem

$$ \underset{\mathbf{x}_a,\mathbf{f}_a,\mathbf{f}_c,\mathbf{x}_\tau}{\operatorname{argmin}} \sum_{i=1}^{N} \mathcal{C}_i $$
$$ \text{s. t. } \mathbf{f}_{c_i} = \mathbf{x}_\tau \oplus \mathbf{f}_{c_j}, \quad i,j = 1,\dots,N, i \neq j $$

where $\mathbf{f}_{c_i}$ and $\mathbf{f}_{c_j}$ are common features in the maps of users i and j

Linearized system

$$ \underset{\delta\mathbf{x}_\tau,\delta\mathbf{x}_{c_1},\dots,\delta\mathbf{x}_{c_N}}{\operatorname{argmin}} \sum_{i=1}^{N} ||\mathbf{J}_i\delta\mathbf{x}_{c_i} - \mathbf{b}_i||^2 $$
$$ \text{s. t. } \sum_{i=1}^{N}\mathbf{A}_i\delta\mathbf{x}_{c_i} + \mathbf{A}_\tau\delta\mathbf{x}_\tau = \mathbf{r} $$

where $\mathbf{x}_{c_i} \triangleq \begin{bmatrix}\mathbf{x}_{a_i}{}^T & \mathbf{f}_{a_i}{}^T & \mathbf{f}_{c_i}{}^T\end{bmatrix}^T$

Problem solution

KKT optimality conditions

$$ \left.\begin{aligned} &\mathbf{J}_i^T(\mathbf{J}_i\delta\mathbf{x}_{c_i} - \mathbf{b}_i) + \mathbf{A}_i^T\lambda = \mathbf{0} \quad i = 1,\dots,N \\ &\sum_{i=1}^{N}\mathbf{A}_i\delta\mathbf{x}_{c_i} + \mathbf{A}_\tau\delta\mathbf{x}_\tau - \mathbf{r} = \mathbf{0} \\ &\mathbf{A}_\tau^T\lambda = \mathbf{0} \end{aligned}\right\} $$

for N = 2
$$ \Longrightarrow \underbrace{\begin{bmatrix} \mathbf{J}_1^T\mathbf{J}_1 & & \mathbf{A}_1^T & \\ & \mathbf{J}_2^T\mathbf{J}_2 & \mathbf{A}_2^T & \\ \mathbf{A}_1 & \mathbf{A}_2 & & \mathbf{A}_\tau \\ & & \mathbf{A}_\tau^T & \end{bmatrix}}_{\mathbf{H}_{CM}} \begin{bmatrix} \delta\mathbf{x}_{c_1} \\ \delta\mathbf{x}_{c_2} \\ \lambda \\ \delta\mathbf{x}_\tau \end{bmatrix} = \begin{bmatrix} \mathbf{J}_1^T\mathbf{b}_1 \\ \mathbf{J}_2^T\mathbf{b}_2 \\ \mathbf{r} \\ \mathbf{0} \end{bmatrix} $$

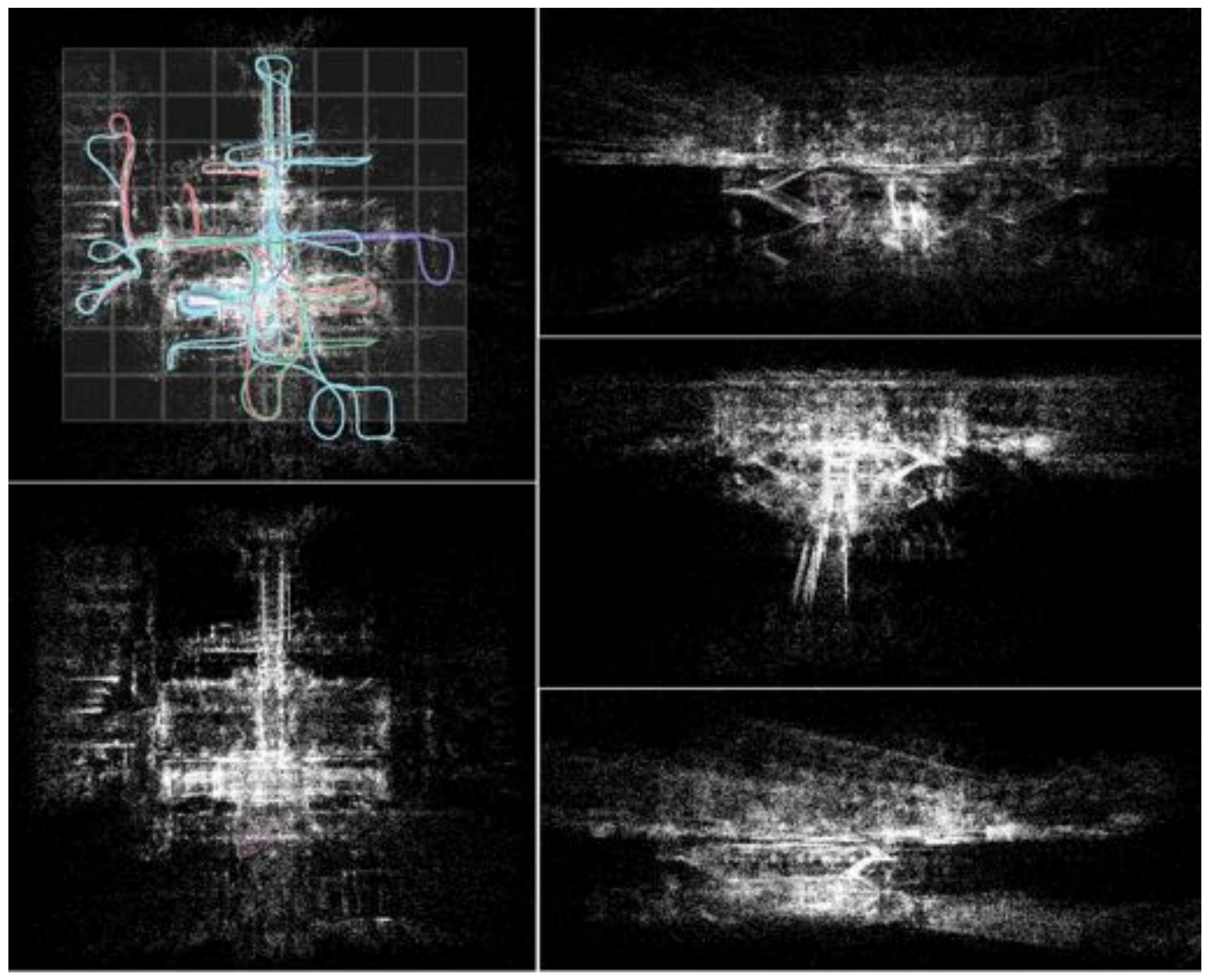
which is equivalent to two triangular back-substitutions

$$ \begin{bmatrix} \mathbf{G}_1 & & & \\ & \mathbf{G}_2 & & \\ \mathbf{K}_1^T & \mathbf{K}_2^T & \mathbf{T}_{11} & \\ & & \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{G}_1^T & & \mathbf{K}_1 & \\ & \mathbf{G}_2^T & \mathbf{K}_2 & \\ & & -\mathbf{T}_{11}^T & \mathbf{T}_{21}^T \\ & & & -\mathbf{T}_{22}^T \end{bmatrix} \begin{bmatrix} \delta\mathbf{x}_{c_1} \\ \delta\mathbf{x}_{c_2} \\ \lambda \\ \delta\mathbf{x}_\tau \end{bmatrix} = \begin{bmatrix} \mathbf{J}_1^T\mathbf{b}_1 \\ \mathbf{J}_2^T\mathbf{b}_2 \\ \mathbf{r} \\ \mathbf{0} \end{bmatrix} $$

Experimental Results

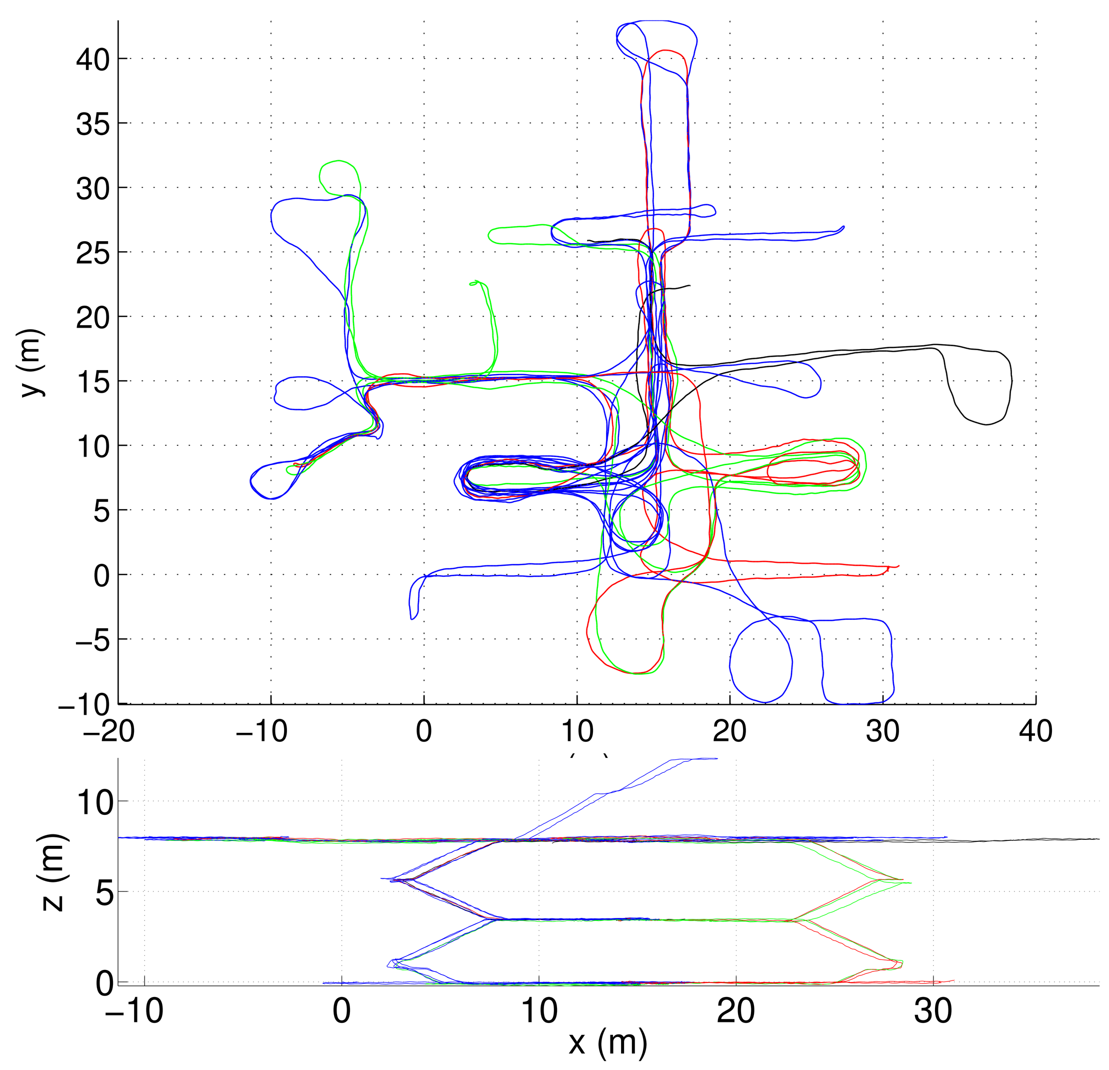
Data from a 2,000 m^2 building, spanning a total distance of approx. 2 km. Harris and KLT were used to generate consecutive image feature tracks. Loop closures found using feature descriptors (ORB, FREAK, or SIFT) and a vocabulary tree (VT).

Point cloud and trajectories



Interactive visualization: http://onionmaps.info

User trajectories (each color represents a different user)



Results validation

- 0.5% error in the estimated dimensions of the building compared to the blueprints
- By dropping constraints imposed by a subset of commonly-observed landmarks, we increase speed by 3, while introducing only 13 cm of additional RMS error

Re-localization using the created map

- Query the VT to determine feature matches between the current query image and the mapped images
- Employ PnP algorithm to localize within the CM
- Given sufficient feature matches and geometry, we are able to localize user within 0.5 m for over 95% queries

Reference

[1]: Chao X. Guo, Ryan C. DuToit, Kourosh Sartipi, Georgios Georgiou, Ruipeng Li, John O'Leary, Esha D. Nerurkar, Joel A. Hesch and Stergios I. Roumeliotis. "Large-Scale Cooperative 3D Visual-Inertial Mapping in a Manhattan World". IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS' 15), (submitted).