THREE-DIMENSIONAL RECONSTRUCTION FROM HETEROGENEOUS VIDEO DEVICES WITH CAMERA-IN-VIEW INFORMATION

Simone Milani

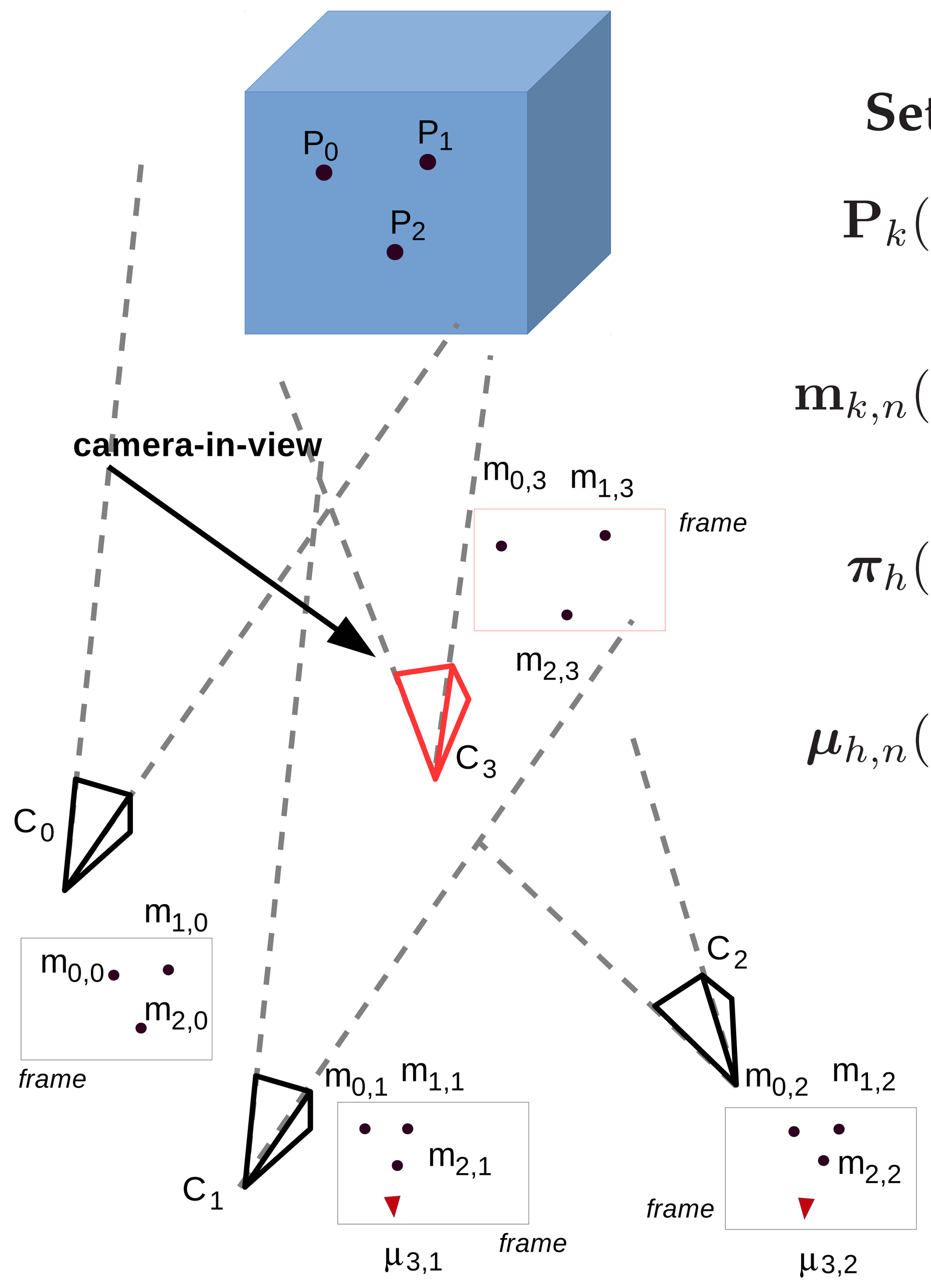
Department of Information Engineering, University of Padova, Italy - simone.milani@dei.unipd.it

DEPARTMENT OF INFORMATION ENGINEERING UNIVERSITY OF PADOVA

1-Overview

In this work, a 3D modelization of the surrounding environment is enabled with an improvised ad-hoc camera networks of both static and mobile devices (cloud vision network).
The estimation can be significantly improved whenever one or more cameras (named here camera-in-views) can be localized within the field of view of other devices.
The locations of camera-in-views (CIV) correspond to both scene points and extrinsic parameters. Image points and synchronization associated to CIV are obtained via a VLC signaling.
As a matter of fact, it is possible to modify a standard bundle adjustment algorithm to improve the accuracy and reduce the amount of iterations.
Experimental results show that this additional information can improve the accuracy of the system up to 17 %.

2-Scenario



Set-up:

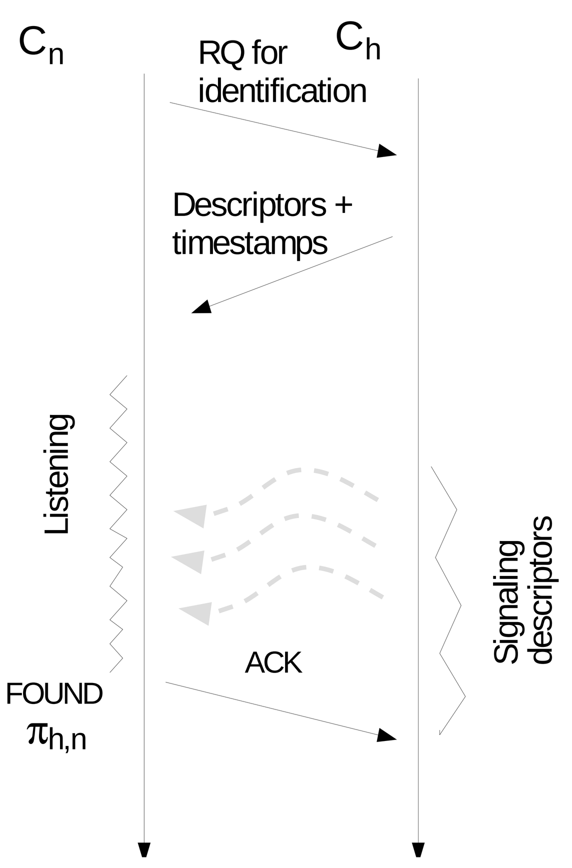
- $\mathbf{P}_k(t)$ 3D point in homogeneous coordinates;
- $\mathbf{m}_{k,n}(t)$ projection of $\mathbf{P}_k(t)$ on camera C_n;
- $\pi_h(t)$ location of C_h in 3D coordinates;
- $\mu_{h,n}(t)$ projection of location $\pi_h(t)$ on C_n;

$$\mathbf{m}_{k,n}(t) \sim K_n\,[R_n(t)|T_n(t)]\,\mathbf{P}_k(t)$$

$$\boldsymbol{\mu}_{h,n}(t) \sim K_n\,[R_n(t)|T_n(t)]\,\boldsymbol{\pi}_h(t)$$

The point $\mathbf{P}_k(t)$ and the camera C_h are projected on camera C_n.

3-Synchronization and localization of cameras



The localization of target objects within images can be performed in different ways:

- SIFT descriptors;
- VLC.

In our implementation, the synchronization is obtained using a VLC protocol (it exploits phone screens or vehicle lights).

Feature-based synchronization is possible as well.

The required accuracy depends on the motion level of the cameras.

4-Bundle adjustment with camera-in-views

Given a set of points $\mathbf{m}_{k,n}$, the bundle adjustment strategy finds $\mathbf{P}_k$, R_n, and T_n

$$\min_{R_n,T_n,K_n,\mathbf{P}_k\ \forall k,n} \sum_{k=0}^{N-1}\sum_{n=0}^{M-1} w_{k,n}\|\mathbf{m}_{k,n} - K_n[R_n|T_n]\mathbf{P}_k\|^2$$

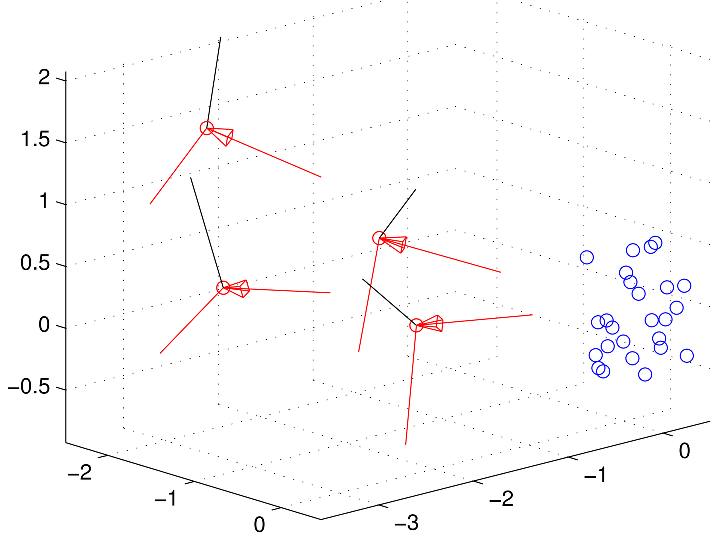
via an iterative two-steps minimization strategy.
If camera-in-views are known (i.e. $\mu_{h,n}$), the target function becomes

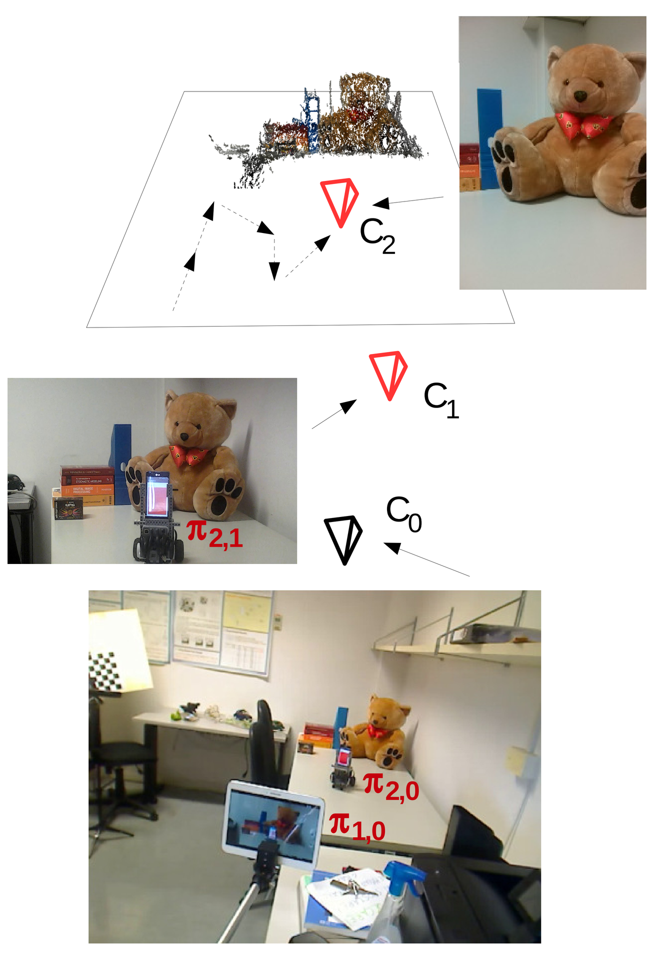
$$\sum_{h=0}^{M-1}\sum_{k=0}^{N-1}\sum_{n=0}^{M-1}\Big\{w_{k,n}\|\mathbf{m}_{k,n} - K_n[R_n|T_n]\mathbf{P}_k\|^2 + \lambda\,\omega_{h,n}\|\boldsymbol{\mu}_{h,n} - K_n[R_n|T_n]\boldsymbol{\pi}_h\|^2\Big\}\,,$$

where $\omega_{h,n}$ is equal to 1 in case the camera C_h is "in-view" with respect to camera C_n and 0 otherwise.

5-Results

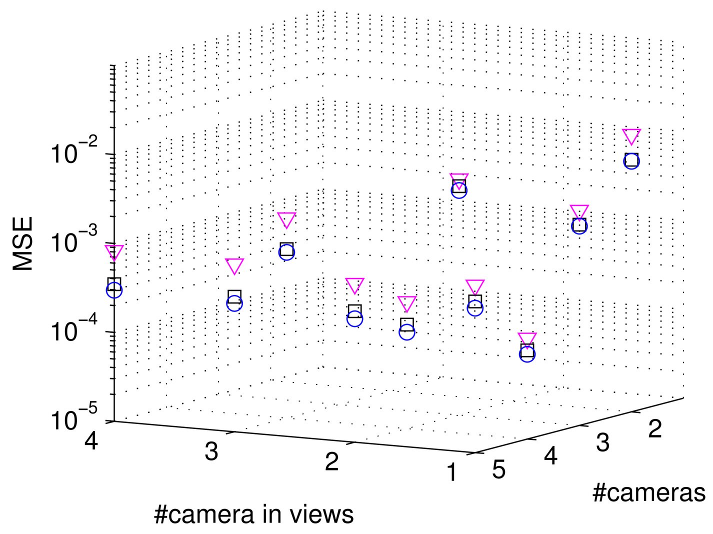
Experimental tests were run both on a synthetic setting and on a real one. The first scenario allows us to evaluate the performance of the approach with different camera settings, where N_T is the total number of cameras and N_{CIV} ($< N_T$) the number of camera-in-views available. In the real scenario, we have 3 cameras with 2 CIVs.



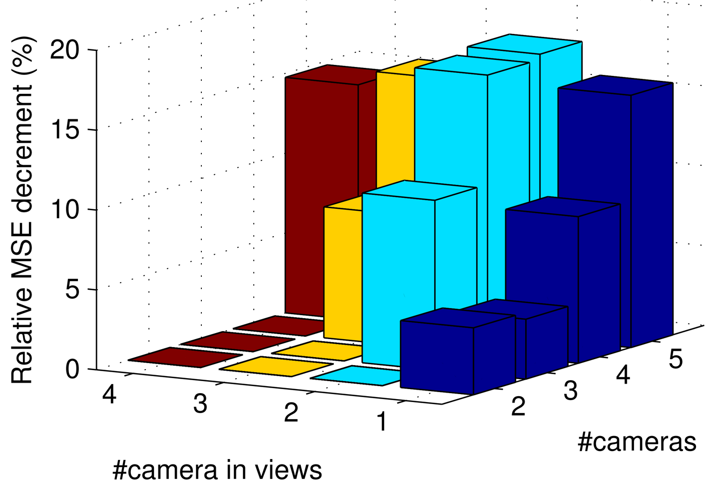


Video demo and further material is available on line.

Synthetic setting

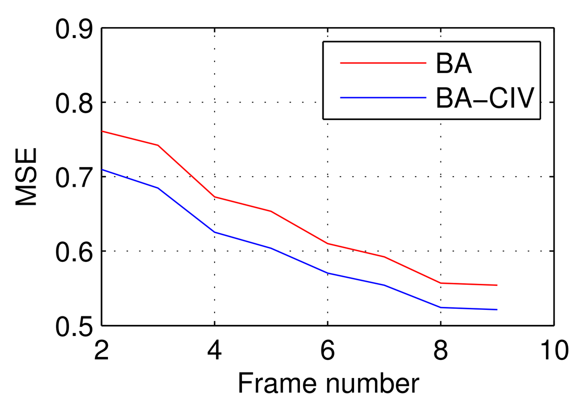


MSE of 3D points. triangle=no BA; square=BA; circle=BA-CIV.

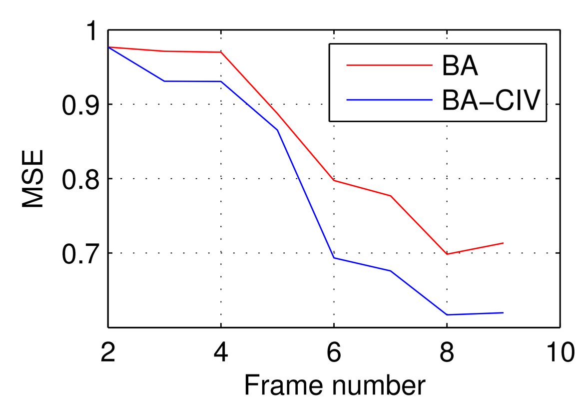


Relative MSE improvement.

Real scenario



C_1 and C_2.



C_0, C_1, and C_2.