

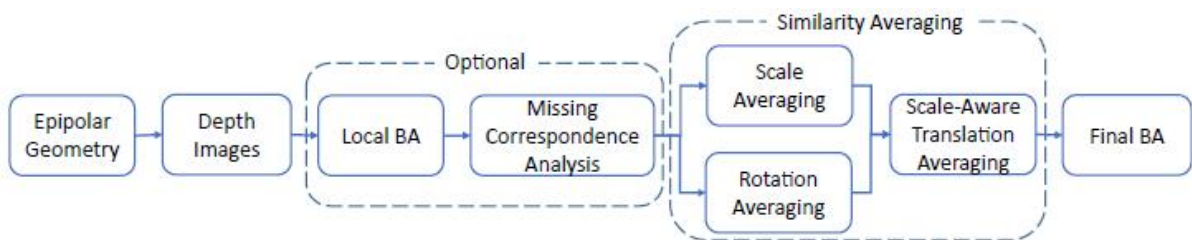
Global Structure-from-Motion by Similarity Averaging

Zhaopeng Cui and Ping Tan

(ICCV 2015)

Goal: To solve all camera positions and orientations in a global coordinate system.

Pipeline of the proposed approach



(Figure from the paper)

[Epipolar Geometry] Input is a set of images with known essential matrices. That is inputs are represented by an Epipolar Graph (see the figure (a) below), where each camera is a vertex and the two cameras are connected if the essential matrix between them is known.

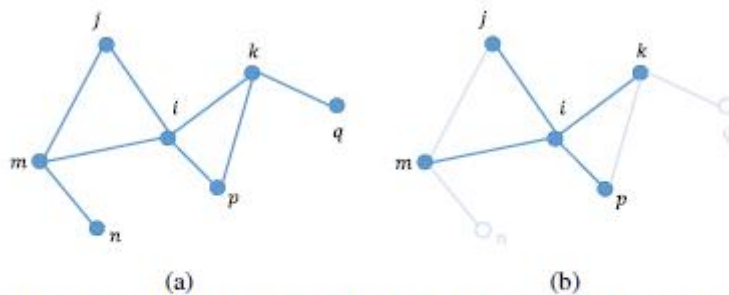


Figure 1. Left: an EG graph where each camera is a vertex and two cameras are connected if the essential matrix between them is known. Right: a stellate graph includes all vertices and edges directly linked to a center vertex i .

[Depth Images] Compute a sparse 'depth image' for each camera from a stellate graph (see the figure (b) above). For a depth image for camera i , the stellate graph includes all vertices and edges **directly** linked to i . A depth image contains the depth values at sparse features.

* With these depth images, we can transform an essential matrix to a similarity transformation. Because similarity transformation encodes rotation, translation and ***scale***, we can determine the **baseline lengths** between the two cameras from these scale ratio.

[Local BA/Optional] is applied to *images in each stellate graph* to improve pairwise relative motion and excludes some poor essential matrices.

[Missing correspondences/Optional] is applied between image pairs to exclude outlier essential matrices due to repetitive scene structures.

[Rotation Averaging] solves all camera orientations simultaneously from input pairwise relative rotations.

[Scale Averaging] solves global scale of each depth image.

[Multiple View Triangulation] computes the scene structure.

[Final BA] optimises the camera parameters and 3D points together.