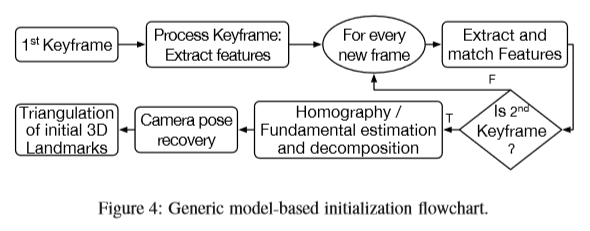
creates the initial 3D map on startup.



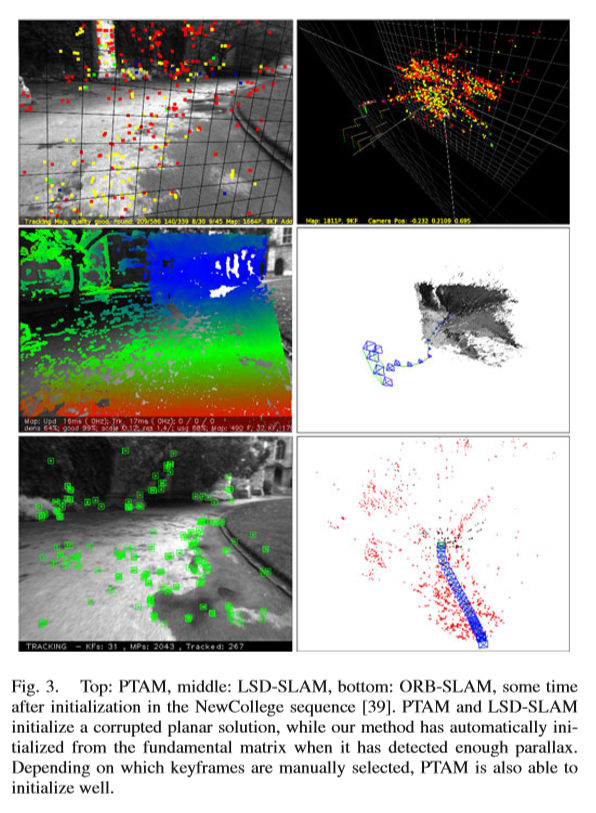
the ﬁrst frame processed by the KSLAM system is typically set as the ﬁrst keyframe. Subsequent frames are processed by establishing 2D-2D data associations, which are monitored to decide whether the new frame is the second keyframe or not. The decision criteria is based on the 2D distances between the found matches in both images. The matches are then used to estimate a Homography (degenerate for nonplanar scenes) or a Fundamental matrix (degenerate for planar scenes) using a robust model ﬁtting method (RANSAC or MLESAC). The estimated Homography or the Fundamental matrix are then decomposed as described in [35 (R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2003.)] into an initial scene structure and initial camera poses. To mitigate degenerate cases, random depth initialization (shown in Fig.3F), as its name suggests, initializes a KSLAM by randomly assigning depth values with large variance to a single initializing keyframe. The random depth is then iteratively updated over subsequent frames until the depth variance converges.

[38] suggested in **LSD SLAM**, and later in DSO, a randomly initialized scene’s depth from the ﬁrst viewpoint, Both systems use an initialization method that does not require two view geometry; it takes place on a single frame: pixels of interest (i.e., image locations that have high intensity gradients) in the ﬁrst keyframe are given a random depth value with an associated

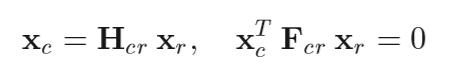
large variance. This results in an initially erroneous 3D map. The pose estimation methods are then invoked to estimate the pose of newly incoming frames using the erroneous map, which in return results in erroneous pose estimates. However, as the system process more frames of the same scene, the originally erroneous depth map converges to a stable solution. The initialization is considered complete when the depth variance of the initial scene converges to a minimum.

**LSD SLAM** initializes by giving the first keyframe a random depth map and a large variance.

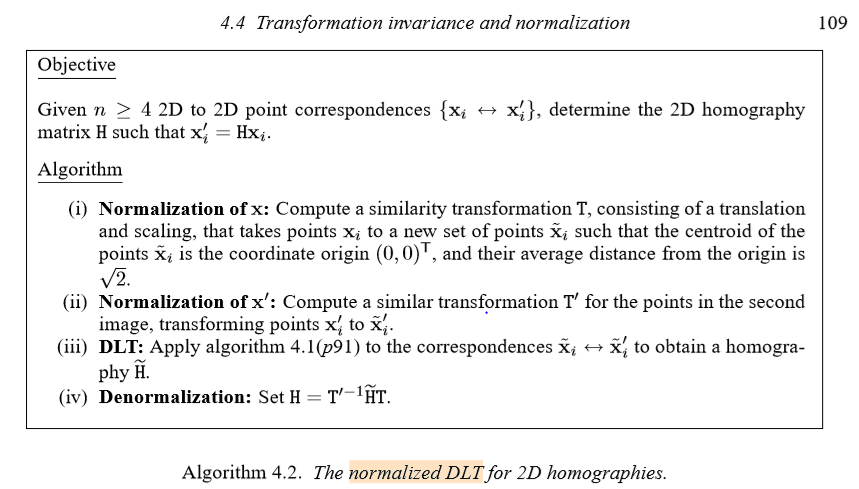
**ORB2** initializes via the following:



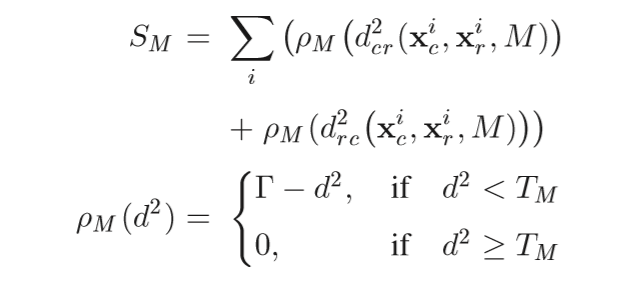
The goal of the map initialization is to compute the relative pose between two frames to triangulate an initial set of map points. This method should be independent of the scene (planar or general) and should not require human intervention to select a good two-view conﬁguration, i.e., a conﬁguration with signiﬁcant parallax. We propose to compute in parallel two geometrical models: a homography assuming a planar scene and a fundamental matrix assuming a nonplanar scene. We then use a heuristic to select a model and try to recover the relative pose with a speciﬁc method for the selected model. Our method only initializes when it is certain that the two-view conﬁguration is safe, detecting low-parallax cases and the well-known twofold planar ambiguity [27], avoiding to initialize a corrupted map. The steps of our algorithm are as follows.

1. Find initial correspondences: Extract ORB features(only at the ﬁnest scale) in the current frame Fc and search for matches xc ↔xr in the reference frame Fr. If not enough matches are found, reset the reference frame.
2. 2) Parallel computation of the two models: Compute in parallel threads a homography Hcr and a fundamental matrix Fcr as

with the normalized DLT and eight-point algorithms, respectively, as explained in [2]

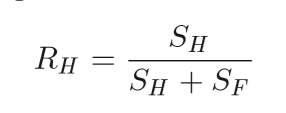


inside a RANSAC scheme. To make homogeneous the procedure for both models, the number of iterations is preﬁxed and the same for both models, along with the points to be used at each iteration: eight for the fundamental matrix, and four of them for the homography. At each iteration, we compute a score SM for each model M (H for the homography, F for the fundamental matrix)



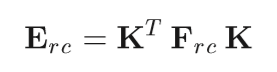
where d2cr and d2rc are the symmetric transfer errors [2] from one frame to the other. TM is the outlier rejection threshold based on the χ2 test at 95% (TH =5 .99, TF =3 .84, assuming a standard deviation of 1 pixel in the measurement error). Γ is deﬁned equal to TH so that both models score equally for the same d in their inlier region, again to make the process homogeneous. We keep the homography and fundamental matrix with the highest score. If no model could be found (not enough inliers), we restart the process again from step 1.

1. Model selection: If the scene is planar, nearly planar or there is low parallax, it can be explained by a homography. However, a fundamental matrix can also be found, but the problem is not well constrained [2], and any attempt to recover the motion from the fundamental matrix would yield wrong results. We should select the homography as the reconstruction method will correctly initialize from a plane or it will detect the low parallax case and refuse the initialization. On the other hand, a nonplanar scene with enough parallax can only be explained by the fundamental matrix, but a homography can also be found explaining a subset of the matches if they lie on a plane or they have low parallax (they are far away). In this case, we should select the fundamental matrix. We have found that a robust heuristic is to compute



and select the homography if RH > 0.45, which adequately captures the planar and low parallax cases. Otherwise, we select the fundamental matrix.

1. Motion and structure from motion recovery: Once a model is selected, we retrieve the motion hypotheses associated. In the case of the homography, we retrieve eight motion hypotheses using the method of Faugeras and Lustman [23]. The method proposes cheirality tests to select the valid solution. However, these tests fail if there is low parallax as points easily go in front or back of the cameras, which could yield the selection of a wrong solution. We propose to directly triangulate the eight solutions and check if there is one solution with most points seen with parallax, in front of both cameras and with low reprojection error. If there is not a clear winner solution, we do not initialize and continue from step 1. This technique to disambiguate the solutions makes our initialization robust under low parallax and the twofold ambiguity conﬁguration and could be considered the key of the robustness of our method. In the case of the fundamental matrix, we convert it in an essential matrix using the calibration matrix K as



And then retrieve four motion hypotheses with the singular value decomposition method explained in [2]. We triangulate the four solutions and select the reconstruction as done for the homography.

1. Bundle adjustment: Finally, we perform a full BA (see the Appendix for details) to reﬁne the initial reconstruction. An example of a challenging initialization in the outdoor New College robot sequence [39] is shown in Fig. 3. It can be seen how PTAM and LSD-SLAM have initialized all points in a plane, while our method has waited until there is enough parallax, initializing correctly from the fundamental matrix.

ORB2 Rephrased:

1. Get 2 frames and match features
2. Compute 2 models in parallel threads, with the same # of iterations for each model:
   1. Homography
      1. Assumes planar scene
      2. Algorithm: Normalized DLT
      3. 4 points used at each iteration
   2. Fundamental Matrix
      1. Assumes non-planar scene
      2. Algorithm: Eight-Point algorithm
      3. 8 points used at each iteration

At each iteration, compute a score for each model – keep the H and F with the highest score.

1. Select either the F or the H – this will depend on whether a planar or non-planar model better suits the scene
2. Motion and Structure from Motion Recovery to get the essential matrix E.
3. Full Bundle Adjustment to refine the initialization