Theory

1. Why do we need the initial pose estimation? What happens, if we just run bundle adjustment?
Since we implement Levenberg Marquardt or Gaussian Newton method for optimization. Both of them may suffer from slow convergence or converge to a local minimum which gives an unacceptable match of observed data if a poor initial guess is used.

2. How can more than three cameras be reconstructed? How can the reconstruction grow beyond what is visible in the first two cameras?
In case more cameras are added, the sparse jacobian matrix would be changed, namely, for each added camera, we need added a new block to represent the relationship between the residuals and the calibration matrix parameters of the camera. Then,m we could estimate the position of camera with a initial guess and do the optimization based on that.

3. What would have to be changed, to also estimate radial distortion in the bundle adjustment?
In case that radial distortion happens, that parameters of calibration matrix should be changed. Normally, we use five parameters to define the K matrix, and assume first two columns are independent, but with radial distortion, the first two columns should be correlated with each other, hence we need 7 parameters to represent K matrix.