Bundle Adjustment using Conjugate Gradients with Multiscale Preconditioning

Martin Byröd
byrod@maths.lth.se
Kalle Åström
kalle@maths.lth.se

Centre for Mathematical Sciences
Lund University
Lund, Sweden

Bundle adjustment is a key component of almost any feature based 3D reconstruction system, used to compute accurate estimates of calibration parameters and structure and motion configurations. The job of the bundle adjustment algorithm is typically to non-linearly minimize *e.g.* the L_2 norm of the reprojection errors. Recently there has been an increased interest in solving for the geometry of very large camera systems [1, 2, 4] and there is thus a need to research methods, which potentially scale better with problem size than the methods in use today.

In the paper, we develop new techniques for fast solution of the bundle adjustment problem using iterative linear solvers. In the commonly used Levenberg-Marquardt method the dominant step is forming and solving the normal equations typically using (sparse) Cholesky factorization. However, it has been hypothesized that for large problems the method of *conjugate gradients* could be a better choice [3, 5]. So far though one has mostly obtained rather disappointing convergence rates this way. This is likely due to the lack of suitable *preconditioners*. In the paper we show how multi scale representations, derived from the underlying geometric layout of the problem, can be used to dramatically increase the power of straight forward preconditioners such as Gauss-Seidel.

Let x denote the unknown parameters to be estimated and denote by $r(x)$ the column vector of residuals. Here $r(x)$ are non-linear functions of the parameters x living on a non-linear manifold. We will consider a Gauss-Newton approach for minimizing $f(x) = r(x)^T r(x)$. This means that in each (outer) iteration we will try to solve

$$J(x_k)\delta x = -r(x_k) \tag{1}$$

in a least squares sense, where $J(x)$ is the Jacobian, *i.e.* the partial derivatives of r with respect to local perturbations δx of the parameters.

In its basic form, the conjugate gradient method solves a square symmetric system $Ax = b$ and requires only multiplication of the matrix A with a vector. The basic way to apply the conjugate gradient algorithm to the bundle adjustment problem is to form the normal equations $J^T J\delta x = -J^T r$ and set $A = J^T J, b = -J^T r$. The crucial issue when applying the conjugate gradient method is the conditioning of A. Whenever the condition number $\kappa(A)$ is large convergence will be slow. In the case of least squares, $A = J^T J$ and thus $\kappa(A) = \kappa(J)^2$, so we will almost inevitably face a large condition number. In these cases one can apply *preconditioning*, which in the case of the conjugate gradient method means pre-multiplying from left and right with a matrix E to form

$$E^T AE\hat{x} = E^T b.$$

The idea is to select E so that $\hat{A} = E^T AE$ has a smaller condition number than A.

For the bundle adjustment problem, standard preconditioners such as the Jacobi or Gauss-Seidel preconditioners do not seem sufficient to obtain a competetive algorithm [5]. Apparently, more domain knowledge needs to be applied. In particular, iterative algorithms seem to have problems with correcting large scale, global deformations. What we propose in the paper is to tackle this situation by introducing an overcomplete multiscale representation taylored to the problem.

A multiscale representation can *e.g.* be obtained by hierarchically splitting the set of unknowns. In each step the set of unknown variables is split into two (approximately equally sized) pieces. This gives a dyadic multi-scale representation of the problem. On each level (scale) new variables are introduced representing transformations of whole groups of variables together. These transformations can be *e.g.* translation, rotation or scaling.

To get a manageble sized problem, we factor out the 3D point variables leaving only the camera variables. Now, given a set of cameras with approximately known camera centers $t_1, \cdots, t_m$ we construct a multiscale representation matrix P using a hierarchical binary partitioning of the cameras.

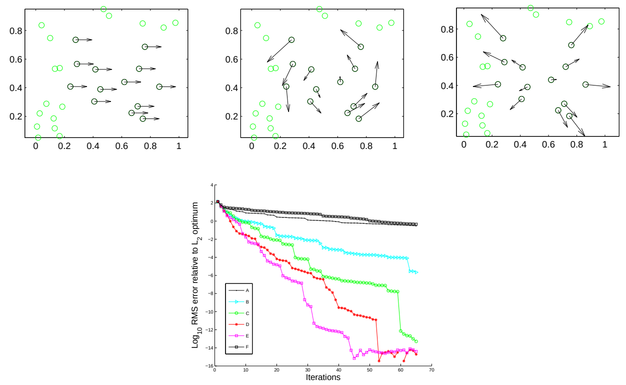

Figure 1: **Top:** Displacement basis vectors for a subset of *e.g.* camera centers. From left to right: translation, rotation and scaling. **Bottom:** Log_{10} residual error relative to the optimal solution using various forms of preconditioning. A: Jacobi, B: Gauss-Seidel (GS), C: Multiscale representation + GS, D: Multiscale with rotation + GS, E: Multiscale with rotation and scaling + GS F: Multiscale + Jacobi.

For each partition $c_i \subset \{t_1, \ldots, t_m\}$, we now add a set of basis vectors x_i, y_i, z_i representing translational displacement to the basis P. For instance, the basis vector x_i would consist of ones for each position corresponding to an x coordinate of $t_i \in c_i$ and zeros otherwise. Optionally, we also add basis vectors corresponding to rotation in three different planes, $t_i^{xy}, t_i^{yz}, t_i^{zx}$ and scaling s_i. See Figure 1 for an illustration of these basis vectors. The basis vectors are collected in a matrix

$$P = [x_1, y_1, z_1, \ldots, x_m, y_m, z_m, \ldots], \tag{2}$$

used to allow multiscale preconditioning. By changing basis according to

$$\tilde{A}_s = P^T A_s P, \quad x = P\tilde{x}, \quad \tilde{b} = P^T b \tag{3}$$

we obtain $\tilde{A}_s\tilde{x} = \tilde{b}$, where A_s is the Shur complement matrix for the camera part of $J^T J$ (for details, see the paper). We can now also apply standard preconditioning to the matrix $\tilde{A}_s$.

The synthetic as well as real experiments we have performed so far both indicate that vastly improved convergence rates can be obtained by iterating on a multiscale representation of A_s. However, the results are so far preliminary and we have yet to show reliable numbers that demonstrate state of the art performance compared to existing implementations of the Levenberg-Marquardt method. See Figure 1 (bottom) for results from a synthetic experiment. The conclusion is that multiscale representations may well have the possibility to become a very effective tool for large scale bundle adjustment, but that more investigation is needed in order to exploit these results and to obtain efficient algorithms.

[1] N. Cornelis, B. Leibe, K. Cornelis, and L. Van Gool. 3d urban scene modeling integrating recognition and reconstruction. *Int. Journal of Computer Vision*, 78(2-3):121–141, July 2008.

[2] Akbarzadeh F. Mordohai. Towards urban 3d reconstruction from video, 2006.

[3] K. Ni, D. Steedly, and F. Dellaert. Out-of-core bundle adjustment for large-scale 3d reconstruction. In *Proc. 11th Int. Conf. on Computer Vision, Rio de Janeiro, Brazil*, pages 1–8, 2007.

[4] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from Internet photo collections. *Int. Journal of Computer Vision*, 80(2): 189–210, November 2008.

[5] W. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment: A modern synthesis. In *Vision Algorithms: Theory and Practice*, LNCS. Springer Verlag, 2000.