

Report on Thesis of Lukáš Polok – Accelerated sparse matrix operations in nonlinear least squares solvers

Richard Hartley

January 23, 2017

1 Overall comments

The thesis is an extraordinarily thorough investigation of the numerical analysis methods that come up in the SLAM and Bundle-adjustment problems, two of the important problems in Computer Vision and Robotics. Although the thesis at times extends its scope to more general least-squares systems or to problems of sparse-matrix numerics in general, it is clear that the main focus is these two problems.

A systematic analysis is undertaken of numerous aspects of the central least-squares solver. After an introductory chapter on the applications of least-squares solvers a very informative summary of techniques for sparse-matrix representation and then a review of existing linear solvers are given. I found these two chapters particularly informative.

The author then goes ahead to describe his innovations in many different facets of the problem:

1. a new format for sparse matrix representation, particularly suited to the computations involved in sparse least-squares solvers;
2. incremental and batch bundle-adjustment methods;
3. covariance calculation;
4. GPU acceleration of the required computations.

The parts of the thesis that most captured my attention were the incremental methods, which arise from the addition of new measurements to the system, both in the methods applied to the resolving of the numerical system (incremental Cholesky factorization) and the incremental computation of covariance calculation. The techniques described in these chapters were both interesting and innovative.

Overall evaluation. Overall, the thesis was very comprehensive, showing a very detailed knowledge of the problem and the previous work, and also proposing important advances to the state of knowledge and the computational methods in this area. As such the thesis is an excellent piece of work, and easily fulfills the requirements for the award of a PhD.

2 More specific comments

The comprehensive nature of this thesis makes it suitable as an extended tutorial of the topic of bundle-adjustment and SLAM. I feel that it would be interesting to many readers, for the description of the issues involved in BA/SLAM, the overview of prior art, and the innovations that the authors have made.

In some areas, issues are discussed in a way that would be difficult for a person not already knowledgeable in the area to follow. Some topics are not sufficiently introduced in a way that a non-initiate would find easy, or even possible to follow.

My feeling is that the general topic of sparse non-linear least squares is not sufficiently prepared. The thesis would be improved by a discussion of the sources and patterns of sparseness of the systems that need to be solved in the SLAM and SfM contexts.

There are really two different types of sparseness. The major source of sparseness in structure-from-motion (SfM) is the fact that a given observation or measurement involves (usually) a single point (or geometric primitive) and a single camera (or sensor). In a large SfM system this can make the system 99% sparse, or more. However, as is pointed out, this type of sparseness is well-handled by the Schur-complement method. This type of sparseness has been well described in various places, such as the recent book on Photogrammetric Computer Vision by Förstner et al. or other sources such as the Hartley-Zisserman book.

A second source of sparseness more relevant in the context of SLAM than in the large-scale 3D reconstruction problems that are done from on-line image data sets is the *banded* nature of the observations taken from a single moving camera. This means that the matrix produced by Schur complement (in the upper-left position of the normal equations, used to solve just for camera parameters) is itself sparse. In fact, a block with coordinates (i, j) corresponding to parameters of cameras i and j is non-zero exactly when the cameras i and j see some point in common. If the camera follows a long track, then points fall out of the visible set and this matrix is roughly banded around the diagonal blocks, until loop-closure occurs. This is where techniques such as the skyline Cholesky methods become important. If one understands the source of this bandedness, then one can better appreciate the problems of fill-in caused by loop closure, for instance.

It would be useful also to understand the issues that make the Schur-complement method not always applicable, and why we may need more general, less problem-specific methods to handle sparseness. As the author points out, the Schur complement method requires a bipartite nature for the measurements.

Each measurement should depend on two entities (each parametrized), usually a point and a camera. It would be instructive to point out where this breaks down. In particular, the SC (Schur complements) handles any sort of relationship (shared parameters, distances, restriction to lie on a line or circle) between the different **cameras**. However, it does not handle so easily any sort of similar relationships between the **points**, since it destroys the advantage of SC that the updates to the point parameters can be computed independently and easily. Nevertheless, such restrictions can often be handled as *soft* constraints.

A very brief hint at the issue is hidden away somewhere I can not locate again, where it is pointed out (a point I had not sufficiently appreciated) that in SLAM problems the savings from SC are not as substantial, because on may typically have thousands or tens of thousands of cameras (i.e. camera positions) but maybe only a hundred or so points (landmarks). This is the opposite of the BA situation where one may have millions of points, but only thousands of cameras. The obvious expedient of swapping the roles of points and cameras may not be effective because of relationships between the cameras, such as odometry measurements. Still, this is mentioned very briefly, and probably the point would be lost on almost anyone who has not actually implemented a SC BA system.

A further example is the discussion of the computation of covariance. This is undertaken without any real preparation. It is indicated that the covariance is related to inverting the normal equations (or the SC equation), but the exact formula is not given. Hence the discussion proceeds without the possibility of a full understanding by the reader, unless he/she knows the formula for the covariance that is being discussed.

Generally, what I am saying here is that a more careful and analytical discussion of the causes and patterns of sparseness would provide more context for the reader.

I note that the distinction between BA and SLAM made here, that SLAM problems may typically have many fewer points than BA is a distinction that has arisen from the current trend of making reconstructions from internet datasets or in situations in which the order of capture of the images is largely ignored. The problems considered as BA or 3D reconstruction problems in the 20th century as often as not involved sequential image capture, as with the *dinosaur* data set, or Nister's *Bundle adjustment rules* paper.

As another example, it is not really pointed out very clearly in the discussion of Cholesky factorization (round page 87) that the size of the incremental problem (specifically the matrix Ω in equation (7.1) depends on how close one can keep related parameters together, and hence is strongly influenced by the presence of absence of loop-closures.

Conjugate gradient methods are only mentioned in passing in the thesis. One feels that there is a lot more that could be said on this topic, since they avoid many of the problems that arise from the methods described in the thesis.

On the subject of Schur complements, I am confused by the statement on page 113 that using the SC does not reduce complexity. I am not sure in what sense (strictly mathematical?) the word complexity is used. Also, I am not

sure what SC is being compared with here. Is it alternative sparse solvers? In a strict sense, computing the SC should take time $O(n^2)$, and solving is a lot quicker, since it depends only linearly on the number of points (and cubically on the number of cameras). In the other hand, standard techniques would be $O(n^3)$ (though smaller using sparse systems). The argument, that SC is faster because it uses simple additions and multiplications, rather than divisions and square roots in Cholesky, is one that I cannot believe.

GPU processing. I read the chapters on GPU processing with some interest, but less understanding since I am not as familiar with this topic. Sometimes, however, it is hard to follow. I am not sure that it is really possible to understand the brief description given for the matrix multiplication or sorting routines.

3 Summary

I make some suggestions above about motivation of the discussions of sparseness and writing in a style more accessible for non-experts in the field. My suggestions are more about the expository style than the content of the thesis. This is more of a suggestion for any possible revision for later publication. I do not think that it is necessary to make such significant changes to the thesis itself.

The thesis does an excellent job of laying out a body of work and relating it to the prior art. In my opinion it goes well beyond what is expected, and shows a depth of understanding of the problems, and an intricate knowledge of current approaches to the topic, that are truly impressive.

4 Recommendation