SCENE RECONSTRUCTION FROM MULTIPLE IMAGES WITH SINGLE CENTRE OF PROJECTION

A. I. Audu¹ and A. H. Sadka²

¹Department of Computer Engineering, University of Maiduguri, Nigeria.
²Department of Electronic and Computer Engineering, Brunel University, London, United Kingdom

Corresponding author's e-mail address: lm324fairchild@yahoo.com; lm324fairchild@gmail.com

Abstract
One problem which has remained deeply rooted in multi-view geometry is the reconstruction of a three-dimensional (3D) scene starting with multiple two-dimensional (2D) representations of different parts of the same scene. In this paper, the main focus is to perform metric reconstruction of a three-dimensional scene from multiple image samples using sparse bundle adjustment algorithm. In most of the multi-view cases which have been documented in literature, acquisition cameras are assumed to have different centers of projection. The focus here is on the unique problem of refining visual reconstruction which yields camera pose and calibration, and three-dimensional structure estimate when multiple views have a single center of projection.

Key words: Metric reconstruction; sparse bundle adjustment; multiple images; single center of projection; synthetic and real scene; image segmentation.

1. Introduction
Sparse Bundle Adjustment (SBA) is an efficient algorithm used in computer vision. It provides a concise and straightforward introduction as well as a detailed coverage on the unique research problem of re-projection error minimization in a wide range of stereo imagery. It adequately addresses the problems of high computational and memory costs associated with least squares optimization techniques used in the early years of bundle adjustment. This is in view of the fact that the Jacobian of some parameters exhibit sparsity which can be tailored to improve computational speed. This concept has been successfully applied in parameter refinement problems and has allowed inconsistencies between stereo pairs to be removed. It has also enhanced the realistic descriptions and modeling of other sophisticated imaging system application (Triggs et al., 2000; Hu et al., 2012; Liu and Hu, 2012).

There are two trends which have been widely adopted in scene reconstruction algorithms based on SBA. One situation is when the scene is fixed and the camera motion is around it. This is of course simple to implement with hand-held cameras, otherwise a specialized rig is required. When a hand-held camera is used the recovery of homographies that map images to each other and consequently allow the images to be transformed and combined becomes a huge challenge. Alternatively, and the focus of this work, the scene object is placed on a turn-table and images are acquired at regular angular interval with a fixed camera. This is only realizable if the scene is a portable object or cast object.

2. Related Work
Understandably, Li (et al., 2011) has it that the stereo image data obtained by High-Resolution Imaging Science Experiment (HiRISE) and used in the characterization of the surface, subsurface, and atmosphere of Mars have been optimized using SBA, in view of the scientific relevance of the project. The project is aimed at obtaining high-precision topographic information which is at the core of most space exploration programmes.

SBA is also known to have been used in image mosaicking (McLauchlan and Jaenicke, 2002). This approach facilitates an accurate 3D reconstruction from multiple images of the same. This procedure features result which is statistically optimal based on the enforcement of hard geometric constraints.
constraints. Frame decimation discussed in (Nistér, 2000) provides for an automatic method which decides on the frame rate for any image sequence to be used in a structure from motion problem. This idea makes use of a coarse to fine, optical flow-based video mosaicking algorithm. Expansion procedure is adopted in (Furukawa and Ponce, 2010) to output a qualitative dense set of patches on the surface of an object. Feature points obtained by matching multiple images are used to generate initial patches. These are continuously expanded until dense patches are obtained. The optimality of the patches is determined by local photometric consistency and global visibility constraints.

In simultaneous localization and mapping (SLAM) problem proposed by (Konolige and Agrawal, 2008), classical bundle adjustment (BA) technique has been used to precisely describe and estimate the position of a mobile robot on a constructed map of the same environment.

Cost function is very critical in the formulation of SBA. The comprehensive theoretical background of SBA given in (Engels et al., 2006) defines a robust cost function as the square sum of all the dimensions of an error vector function. This is contrary to the conventional method of approximating the cost function locally with a quadratic Taylor expansion. The discussion also provides an entirely different proposition in which it is investigated that error-buildup can be described as a function of accuracy that is obtainable through the use of bundle adjustment and helps to improve the reliability of camera tracking. The discussion also focuses on how bundle adjustment facilitates real-time application.

It is observed that the definition of variables in bundle adjustment is critical when large error propagation is needed in order to correct global error at loop closure. This will involve constraining the expected solution. An adaptive bundle adjustment is proposed in (Sibley et al., 2009) which works in a metric-space defined by a connected Riemannian manifold. This is aimed at addressing the problem of single coordinate frame which is responsible for the high computational cost of bundle adjustment.

Since line features from image correspondences is an integral part of scene modeling, augmented reality, and visual servoing (Bartoli and Sturm, 2005), the ‘two points’ and the ‘two plane’ overparameterization which can cause gauge freedoms and/or internal constraints is addressed with plucker coordinates so that the feature lines can be represented in 3D.

In (Blonquist and Pack, 2011), solution instability due to the linear dependencies between parameters when perspective algorithms (collinearity based) are used has been identified to be common in imaging situations where long focal length lenses and narrow FOV play a prominent role. This problem is addressed by the use of a scaled orthographic projection model based on linear algebraic formulations. The mathematical model developed uses quaternions which translate to partial derivatives as well as the inner constraint equations for a scaled orthographic bundle adjustment.

In the method adopted by (Konolige and Garage, 2010), fast direct Cholesky decomposition techniques are employed to solve SBA problem with sparse linear subproblem. The technique involves the use of a compressed representation of large sparse matrices for efficiently handling the block data structures of SBA to take advantage of this representation. The performance evaluation is proved to surpass what is known to be obtainable from the method used in (Medioni and Kang, 2004).

2. Formulation of SBA Based on Levenberg-Marquardt (LMA) Algorithm

Problems arising in the contest of SBA have the possibility of being solved using Powell’s dog-leg or conjugate gradient technique. However, LMA remains widely known and attractive at solving unconstrained non-linear least squares problems (Marquardt, 1963). It can take advantage of the sparseness that exists at different levels in bundle adjustment. LMA can be thought of as a
A combination of steepest descent and the Newton method. When the current solution is far from the correct one, the algorithm behaves like a steepest descent method: Slow, but guaranteed to converge. When the current solution is close to the correct solution, it becomes a Newton’s method. The behavior, in terms of convergence, and computational cost of this process depend on the parameterization used to represent the problem, i.e. of structure and motion (Lourakis and Argyros, 2009).

Starting from the measured image pixel coordinates $m$ resulting from image formation by multi-view cameras with parameter $P \in \mathbb{R}$, an estimated measurement vector $\hat{m} \in \mathbb{R}$ can be formulated as

$$m^* = g(P)$$

(1)

Let the difference between the measured quantity $m$ and estimated quantity $\hat{m}$ be

$$m - m^* = \varepsilon$$

(2)

A minima value for the quantity $\varepsilon^T \varepsilon$ known as square distance can be used as a criteria in the determination of $P^*$ which satisfies the relation of (1) if an initial guess value $P_0$ is provided.

The fundamental basis of LMA is that for an infinitesimal change $\delta P$ in $P$, $g$ can be approximately expressed as

$$g(P + \delta P) = g(P) + J\delta P$$

(3)

where $J$ is the matrix of all first-order partial derivatives of the function $g$. It is called Jacobian. Equation (2) can now be expressed as

$$||m - g(P + \delta P)|| = ||\varepsilon - J\delta P||$$

(4)

For each $\delta P$, an iteration process is performed using (4) until when $J\delta P - \varepsilon$ is orthogonal to the column space of $J$. This means that

$$J^T(\delta P - \varepsilon) = J^TJ\delta P - J^T\varepsilon = 0$$

(5)

$$J^T\delta P = J^T\varepsilon$$

(6)

$J^T$ is the transpose of $J$. Equation (5) is referred as normal equations, (Golub and Van Loan; Lourakis and Argyros, 2009). $J^TJ$ constitute the Hessian of $\frac{1}{2}\varepsilon^T\varepsilon$ and the infinitesimal change $\delta P$ is called Gauss-Newton step. $J^T\varepsilon$ is the steepest descent components.

For the purpose of error reduction and to make sure that $J$ is not rank deficient or singularity of $J^TJ$, a modified form of (6) referred to as the augmented normal equations is solved as shown in (7).

This involves the manipulation of the diagonal elements of $J^TJ$ using $\sigma$ as a damping element. $\sigma$ is always chosen such that $J^TJ + \sigma I$ is non-singular and positive definite.

$$\left(J^TJ + \sigma I\right)\delta P = J^T\varepsilon$$

(7)

The unpredictability of $m$ described by covariance matrix $P_m$, can be factored into (7) such that we now have $\left(J^T\sum_{m}^{-1}J + \sigma I\right)\delta P = J^T\sum_{m}^{-1}\varepsilon$.

This variation allows for the minimization of squared $\sum_{m}^{-1}$-norm (Mahalanobis distance) as

$$||\varepsilon||^2_{P_m} = \varepsilon^T\sum_{m}^{-1}\varepsilon$$

(8)

The computational complexity is still cubic.

In multi-view setting, the inherent characteristic of lack of interaction among parameters for different 3D points and cameras results in the underlying normal equations exhibiting a sparse block structure (Lourakis and Argyros, 2009). Therefore, SBA is a variant of LMA which seeks to
reduce the computational cost of bundle adjustment by taking advantage of data sparseness. It is established in literature that SBA can cope with parameterization of the multi-view reconstruction problem such as arbitrary projective cameras, partially or fully intrinsically calibrated cameras, exterior orientation (i.e. pose) estimation from fixed 3D points, and refinement of intrinsic parameters.

If $P$ is a vector of parameters describing the $d$ projection matrices and the $c$ 3D points.

$$P = [a_1^T, ..., a_d^T, ..., b_1^T, ..., b_c^T]^T$$

Also if the vector $X$ of $m_{ij}$ is

$$m = [(m_{11})^T, ..., (m_{1d})^T, (m_{21})^T, ..., (m_{2d})^T, ..., (m_{c1})^T, ..., (m_{cd})^T]^T$$

Then the Jacobian of the vector of $X$, i.e $J = \frac{\partial m}{\partial P}$, exhibit sparseness based on the conditions of (11) and (12)

$$\frac{\partial m_{ij}}{\partial a_k} = 0, \quad k \neq i$$

$$\frac{\partial m_{ij}}{\partial b_k} = 0, \quad k \neq j$$

Let the $\delta_p$ be written in vector form in terms of camera and scene points (structure) parameters $\delta_a$ and $\delta_b$ respectively as

$$\delta_p = [\delta_a^T, \delta_b^T]^T \in \mathbb{R}^N$$

where $N$ equals the sum of $i$ and $j$.

Using (13), $(J^T \sum_m^{-1} J + \sigma I) \delta_p = J^T \sum_m^{-1} \varepsilon$ can be expressed as

$$[J^T \sum_m^{-1} J + \sigma I] [\delta_a \delta_b]^T = J^T \sum_m^{-1} \varepsilon$$

The partitioning of

$$[J^T \sum_m^{-1} J + \sigma I] [\delta_a \delta_b]^T = J^T \sum_m^{-1} \varepsilon$$

can be achieved by further simplification.

By combining the simplified forms of the right and left hand sides of

$$[J^T \sum_m^{-1} J + \sigma I] [\delta_a \delta_b]^T = J^T \sum_m^{-1} \varepsilon,$$

we have
Equation (15) is the partitioned form of the normal equations. We can simply write (15) as

$$DE = F$$  \hspace{1cm} (16)$$

Based on orthogonality principle, the identity matrix in the upper left and lower right corners of $D$ can be written as $A^T A$ and $B^T B$ respectively. The upper right corner containing element $Z$ elements equals $A^T B$ while the lower left corner with $Z$ elements equals $B^T A$, i.e the transpose of $A^T B$. For the purpose of further simplification, (15) is written in compact form as

$$
\begin{bmatrix}
X^* & Z \\
Z^T & Y^*
\end{bmatrix}
\begin{bmatrix}
\Delta_a \\
\Delta_b
\end{bmatrix}
= 
\begin{bmatrix}
\varepsilon_a \\
\varepsilon_b
\end{bmatrix}
$$  \hspace{1cm} (17)$$

$\Delta_a = \delta a$ and $\Delta_b = \delta b$. $X^*$ and $Y^*$ are the augmented forms of $X$ and $Y$ respectively. The pre-multiplication of (17) by

$$
\begin{bmatrix}
I & -ZY^{*^{-1}}T \\
0 & I
\end{bmatrix}
$$  \hspace{1cm} (18)$$

Gives

$$
\begin{bmatrix}
(X^* - ZY^{*^{-1}}Z^T) & 0 \\
Z^T & Y^*
\end{bmatrix}
\begin{bmatrix}
\Delta_a \\
\Delta_b
\end{bmatrix}
= 
\begin{bmatrix}
\varepsilon_a - (ZY^{*^{-1}}\varepsilon_b) \\
\varepsilon_b
\end{bmatrix}
$$  \hspace{1cm} (19)$$

The simplification of (19) gives two simultaneous equations. If we make $\Delta_a$ the subject of the formula, we get

$$
\Delta_a = \frac{\varepsilon_a - (ZY^{*^{-1}}\varepsilon_b)}{(X^* - ZY^{*^{-1}}Z^T)}
$$  \hspace{1cm} (20)$$

For $\Delta_b$ we have

$$
\Delta_b = \frac{\varepsilon_b - \Delta_a Z^T}{Y^*}
$$  \hspace{1cm} (21)$$

An important observation is that the invertible form of $Y^*$ can be seamlessly performed since it is block diagonal.$\Delta_a$ and $\Delta_b$ can now be used to update $P$ such that

$$P_{update} = P + [\Delta_a^T, \Delta_b^T]^T$$  \hspace{1cm} (22)$$

$P_{update}$ is further used to update the re-projection error $\varepsilon$ expressed (2), i.e

$$E_{update} = m - g (P_{update})$$  \hspace{1cm} (23)$$

Therefore, due to the sparseness of the primary structure, the implementation steps of bundle adjustment are entirely linearized. Hence a computational complexity of $O (c)$ is exhibited (Lourakis and Argyros, 2009).
We now test $\varepsilon_{\text{update}}$ by comparing its absolute value with the absolute value of $\varepsilon$. If $\varepsilon_{\text{update}}$ is less or equal to $\varepsilon$, i.e $(||\varepsilon_{\text{update}}|| \leq ||\varepsilon||)$, $\sigma$ is updated, otherwise the iteration process continues until convergence is achieved.

SBA is implemented using Linear Algebra PACKage (LAPACK) written in Fortran 90. LAPACK routines are written so that as much as possible the computation is performed by calls to the Basic Linear Algebra Subprograms (BLAS). It provides routines for solving least-squares solutions of linear systems of equations (Lourakis and Argyros, 2009).

4. Result and Discussion
The sample images with common center of projection, used in this work are given in figure 1. Segmented images of the corresponding images of figure 1 are shown in figure 2. The reconstructed 3D scene is presented in figure 3.
Figure 2 Segmented image samples.
Figure 3 Reconstructed scene from multiple images with SCP using SBA.

The task of scene reconstruction from the multiple images shown in Figure 1 is the main consideration of this work. It is carried out through the use of SBA, a C/C++ implementation of generic bundle adjustment based on the sparse Levenberg-Marquardt algorithm (Lourakis and Argyros, 2004).

In order to use SBA, the images have to be understood contextually. Context is said to be a consequence of theories about cognitive processes, not something observed. It has a theoretical role, not one of a measurable unit. The use of contextual constraints is considered to be indispensable for a capable vision system. It is observed that contextual constraints are ultimately necessary in the interpretation of visual information (Kandefer and Shapiro, 2008). It therefore means the scene contained in the image samples is best understood in the spatial and visual context of the objects in it; the objects are recognized in the context of object features at a lower level of representation; the object features are identified based on the context of primitives at an even lower level; and the primitives are extracted in the context of image pixels at the lowest level of abstraction (Li, 2009).

Markov random field (MRF) theory provides a convenient and consistent way for modeling context dependent entities such as image pixels and correlated features. This is achieved through characterizing mutual influences among such entities using conditional MRF distributions. The practical use of MRF models is extensively demonstrated in (Cross and Jain, 1983) in which a scene is segmented into regions based on texture. Hence, in this work, segmentation of sample images and other steps are implemented in Matlab based on the idea discussed in (Saxena et al., 2005; Saxena et al., 2007; Saxena et al., 2008; Saxena et al., 2007). The segmentation results for the different image samples are shown in Figure 2. Consequently, feature detection, extraction, and matching is performed on segment base rather than the whole image at a goal.

The other implementation steps are similar to that used in classical BA, except that the optimization process is performed using generic SBA (Lourakis and Argyros, 2009). This means that starting with initial estimates of the 3D coordinates corresponding to a set of points seen in image samples, as well as initial estimates for the viewing parameters pertaining to each image; SBA is a large optimization problem which takes advantage of the sparsity associated with the linear equations to be solved. It involves the simultaneous refinement of the 3D structure and viewing parameters (i.e. camera pose and possibly intrinsic calibration and radial distortion), aiming to obtain a reconstruction which is optimal under the assumption of zero-mean Gaussian regarding the noise pertaining to the observed image features. So long as outliers have been removed from the initial estimate of 3D points using geometry-based techniques, SBA algorithm can demonstrate a high degree of robustness.

At the heart of SBA algorithm is eucsbademo.c. This visual C file is modified in two ways. First it is modified so that motion and/or structure parameters are printed to an output file after the SBA has been run. Second an assignment to perform motion and structure optimization is set. The compilation of eucsbademo.c and other related files are then performed. This step allows some of the necessary libraries to be created. Whenever eucsbademo.c is run, metric reconstruction (Zhang et al., 2007) is performed using the optimized 3D points and camera parameters. During this process, mono estimation, robust pose and match generation, dense match, reinference, coherence texture processing, volumetric refining, and rendering steps are implemented (Saxena et al., 2007).

The numerical results of SBA simulation are output to a file. This is as a result of the modification mentioned earlier on. The file not only contains statistics regarding the minimization but also 3D coordinates of scene points, camera pose, and image pixel coordinates are. In the output file, the
first four lines summarize the statistics regarding the minimization that has been performed. It describes the number of 3D points that have been optimized, number of cameras or frames considered, number of image pixels, SBA method used i.e structure and motion (BA MOTSTRUCT), expert driver, analytic Jacobian, fixed intrinsics, and whether or not covariance is used. It also gives a summary of the number of iterations before convergence is achieved. The initial and final errors are also described. For example, SBA algorithm reduces the error from 1235.07 to 0.0108958 in 200msecs.

Motion parameters which describe the poses of the cameras are given in line six and seven in each of the sections. The pose of the reference camera is given in line six and consists of four elements quaternion and three elements translational matrix. Line seven contains the pose of the second camera.

A rendered 3D scene is shown in figure 3. The quality of the rendered scene is not very good as should be expected from SBA. There is the presence of holes. This is purely due to two fundamental reasons. First, image samples with SCP have been considered in this work since it is applicable to panoramic images. The second reason is related to the inaccuracy embedded in the initial estimate of 3D parameters. However, when the reconstructed image is zoomed into, the size of the observed holes is reduced.

5. Conclusion

3D reconstruction from multiple 2D representations of a scene has been investigated. This experiment involves the use of SBA which, in theory is expected to perform computationally better because of the sparsity associated with it, to achieve the above mentioned. The obtained numerical results actually pointed to this based on the statistics regarding the minimization that is performed through the use of SBA. The main challenge encountered is the visualization, as can be seen in figure 3, of the optimized numerical results in Meshlab software environment. The optimized 3D points have been rendered with some holes present. Understandably, this is attributable to the fact that image samples with SCP have been used and also the presence of minute number of outliers in the initial estimate of the 3D points parameters.

References


