Preventing the influence of mismatchings in reduced structure-less cost optimization

Antonio L. Rodríguez
University of Murcia, Spain
Email: alrl1@um.es

***Abstract*—Making bundle adjustment robust against the presence of mismatchings has a significant importance in SfM applications, as a small fraction of outliers can produce large errors in the estimated multiple view geometry.**

In this paper we propose two modifications of GEA, an efficient structure-less bundle adjustment method which precomputes a matching data reduction to speed up the epipolar cost error optimization. The first one is a simple change in the error expression which accelerates the data reduction significantly. The second is to incorporate a loss function in the cost that disables the influence of feature mismatchings during the optimization, while it preserves the computational advantages of the data reduction.

We also describe an accurate structure-less incremental motion estimation procedure which uses the modified GEA to reduce initialization errors in the intermediate steps. The input matchings for the optimization are detected using standard two-view feature matching techniques, demonstrating the robustness against mismatchings achieved with our proposal.

I. Introduction

Bundle adjustment (BA) [15, 3] is widely used to reduce initialization errors in the intermediate steps of incremental Structure from Motion (SfM) applications. This operation is in most occasions crucial to obtain valid and accurate estimations for the camera poses and scene structure, as it prevents error buildup and improves the chance of convergence towards the valid reconstruction configuration.

However, using BA under certain circumstances is unpractical, given the considerable computation cost of the reprojection error optimization. For example, using BA in medium to large scale reconstruction problems can be computationally demanding for reduced performance devices such as smartphones or tablet PC's.

Several techniques have been proposed to alleviate this problem. For example, using Preconditioned Conjugate Gradient (PCG) [1, 2, 16] or the Schur complement trick [3] reduces the computational requirements of the cost optimization. Structure-less BA [17, 12, 8] achieve this by substituting the reprojection error by an alternative cost which does not involve the structure parameters.

In [12] the authors described GEA, an efficient structure-less BA method based on a cost defined on epipolar constraints. Much of the computational efficiency of this method is a consequence of reducing the matching information in a preprocessing operation. Each epipolar constraint defined between a view pair encodes the feature matching information in a data matrix. This matrix has a variable size which depends on the number of feature correspondences detected between the two views, and can be substituted by a small matrix of fixed size, without affecting the results of the optimization [6].

Using GEA in a real SfM pipeline offers certain challenges. For example, the optimization should cope with the presence of outliers in the input set of feature correspondences, which appear in practice due to matching failures, and the existence of perceptual aliasing in the scene. To our knowledge, this problem has not been addressed until now.

In this paper we evaluate several improvements and applications for the GEA algorithm, which were initially proposed in [11]. We describe an alternative reduction which, unlike the method used in [12], does not require a time consuming matrix factorization to produce the reduced matrices. In most occasions GEA requires only a few iterations of the cost optimization to reach the optimal configuration. Hence, in practice the data reduction usually takes a large time to compute, in comparison with the time required by the optimization of the reduced cost. For this reason, our proposal can considerably decrease the overall time required to obtain the optimal camera poses.

Along with this improvement, we describe a method to prevent the influence of feature mismatchings in the quality of the camera poses obtained with GEA. In principle this could be done by applying a loss function individually to the epipolar residual contribution of each input matching in the optimized cost. For example, this approach is used in most classical bundle adjustment methods, where the loss function is applied to the residual of each measurement in the reprojection error. However, using loss functions with a reduced cost is not straightforward, as residuals for different matchings detected between a given view pair are evaluated jointly in each term of the cost function.

We can robustify such reduced cost using a loss function, as long as certain considerations are taken into account during the detection of the input image feature matchings. To demonstrate this, we propose an incremental motion estimation procedure that uses the modified version of GEA to obtain accurate camera pose estimations from a set of input images. In this procedure we use classical image descriptor matching and sample consensus methods to detect feature correspondences between the input images. It is well known that these methods produce mismatchings which are difficult to detect and eliminate. As we will demonstrate, the inclusion of the loss function in the optimized cost will prevent the influence of these mismatchings, which are usually produced by the estimation of incorrect epipolar geometries.

A. Structure of this document

In section II we describe GEA, and the original data reduction technique used to speed up the optimization. In section III we provide an alternative cost function which does not require the matrix factorization to reduce the matching data. Section IV discusses how to robustify the cost optimization against the influence of feature mismatchings. In section V we describe the incremental motion initialization method proposed.

Finally, sections VI and VII provide experimental results which demonstrate the performance of the new GEA implementation, as well as the conclusions for these results.

II. The GEA optimization

To improve the quality of the estimated camera poses, GEA optimizes the following cost function:

$$C_{GEA} = \sum_{i=1}^{n-1} \sum_{j=i}^{n} \sum_{\mathbf{p}\leftrightarrow\mathbf{q}\in\mathcal{M}_{ij}} \left(\mathbf{q}^T E_{ij}^{\dagger}\mathbf{p}\right)^2 \tag{1}$$

In this equation $\mathcal{M}_{ij} = \{\mathbf{p} \leftrightarrow \mathbf{q}\}$ is the set of pairwise point correspondences detected between the input views i and j. Matrix $E_{ij}^{\dagger}$ is the following essential parametrization for the camera poses of those two views:

$$E_{ij}^{\dagger} = \frac{1}{\|\mathbf{C}_j - \mathbf{C}_i\|} R_j \left[\mathbf{C}_j - \mathbf{C}_i\right]_\times R_i^T \tag{2}$$

R_i and $\mathbf{C}_i$ are the rotation and center for the i-th camera. The normalization factor $1/\|\mathbf{C}_j - \mathbf{C}_i\|$ prevents the convergence of the cost optimization to the trivial incorrect solution $\mathbf{C}_i = \mathbf{C}_j$ (with $i \neq j$).

The proposed cost can be rewritten into the following compact form:

$$C_{GEA} = \sum_{i=1}^{n-1} \sum_{j=i}^{n} \mathbf{e}_{ij}^T U_{ij}^T U_{ij} \mathbf{e}_{ij} \tag{3}$$

where $\mathbf{e}_{ij}$ is the vectorization of matrix $E_{ij}^{\dagger}$, and U_{ij} is the DLT transformation of the matching data in $\mathcal{M}_{ij}$ [5]. Each row vector $\mathbf{u}$ in U_{ij} is obtained from one of the feature matchings $(p_x, p_y, 1) \leftrightarrow (q_x, q_y, 1)$ in $\mathcal{M}_{ij}$ as follows:

$$\mathbf{u} = (q_x p_x, q_x p_y, q_x, q_y p_x, q_y p_y, q_y, p_x, p_y, 1)^T \tag{4}$$

A. Reduction of the data matrices

In [12] the authors suggested using the data matrix reduction of U_{ij} proposed in [6] to speed up the evaluation of the cost function in equation 3. Each matrix U_{ij} of size $|\mathcal{M}_{ij}| \times 9$ in this cost can be replaced by a $\tilde{U}$ matrix of size 9×9 such that:

$$U^T U = \tilde{U}^T \tilde{U}$$

The new cost is equivalent to the original, but the time required for its evaluation and optimization is significantly smaller. The matrix $\tilde{U}$ can be obtained from U using different factorizations. With the Cholesky decomposition $U^T U = LL^T$, the reduced matrix can be evaluated as $\tilde{U} = L^T$ Using the SVD decomposition $U = XDV^T$, the reduced matrix is $\tilde{U} = DV^T$. Once obtained, these reduced matrices can be reused in different iterations of the cost optimization.

III. Speeding up the data reduction

The matrix factorization in the original data reduction method requires a significant computation time, which we can save by changing the expression of the GEA cost function in equation 1 to:

$$C_{GEA} = \sum_{i=1}^{n-1} \sum_{j=i}^{n} \mathbf{e}_{ij}^T \Omega_{ij} \mathbf{e}_{ij} \tag{5}$$

The evaluation time for each reduced term in both the new and old expressions is approximately the same. Meanwhile, the reduction $\Omega = U^T U$ in the new expression does not require a time consuming factorization. Furthermore, we can evaluate the matrix Ω efficiently by exploiting the inherent redundancy of its elements. For example, we can use the following summatory to obtain this matrix from the input matchings in $\mathcal{M}$:

$$\Omega = \sum_{k=1}^{|\mathcal{M}|} \mathbf{u}_k \otimes \mathbf{u}_k \tag{6}$$

In this expression $\mathbf{u}_k$ is the DLT transformation for the k-th feature matching in $\mathcal{M}$, as denoted in equation 4. The elements in Ω are redundant because the product of many pairs of entries in $\mathbf{u}$ produces the same value. For example, the elements $(1, 9)$ and $(3, 7)$ in matrix $\mathbf{u}_k \otimes \mathbf{u}_k$ (and in the accumulated matrix Ω as well) are equal:

$$u_1 u_9 = u_3 u_7 = q_x p_x \tag{7}$$

Many entries in each matrix $\mathbf{u}_k \mathbf{u}_k^T$ can also be obtained by multiplying other elements in the matrix, which we can evaluate beforehand. For example, the element $(1, 4)$ can be obtained by multiplying the elements $(3, 7)$ and $(6, 7)$:

$$(u_3 u_7)\,(u_6 u_7) = (p_x q_x)\,(p_x q_y) = p_x^2 q_x q_y = u_1 u_4 \tag{8}$$

We can exploit these facts to design an efficient evaluation code for the Ω matrix, which performs the minimal number of operations to evaluate each term in the summatory of expression 6.

IV. Robustification of reduced costs

Most SfM applications check the epipolar consistency of the feature matchings detected between the input images, before using them to compute the multiple view geometry. This way they can reject a large number of mismatchings and improve the quality of the results obtained. Batch SfM applications such as [14] use RANSAC to estimate the epipolar geometry, and reject outliers with a large epipolar residual for this geometry. Real-time SfM applications, such as the one proposed in [10], can check the epipolar consistency of non-loop closing matchings with the camera poses estimated for the key-frames.

In occasions, certain mismatchings survive this consistency check. As long as the epipolar geometry estimated for a given view pair is correct, we can include the surviving outliers in the GEA optimization without expecting a significant degradation in the quality of the results obtained. These outliers will have a small epipolar residual for the true camera pose configuration. Hence, they will not penalize the convergence of the optimization towards this configuration.

However, RANSAC can eventually estimate an invalid epipolar geometry. For example, when the scene contains repeated structures, we can easily find an incorrect epipolar geometry which is supported by a large number of feature matchings. In this case, some outliers with large residuals for the true camera poses can survive the consistency check, and degrade the results obtained with GEA.

We can use the fact that these outliers have a high algebraic epipolar residual error, to detect those view pairs which are likely to contain them, and reduce their contribution to the cost error. The cost function in equation 5 can be rewritten as follows:

$$C_{GEA} = \sum_{\Omega_{ij} \in \mathcal{O}} \Phi_{|\mathcal{M}_{ij}|} \left(\mathbf{e}_{ij}^T \Omega_{ij} \mathbf{e}_{ij} \right) \tag{9}$$

In this expression $\Phi_n(r)$ is a *ramp* loss function which vanishes the contribution of terms with a large average epipolar residual:

$$\Phi_n(r) = \begin{cases} r & |\frac{r}{n}| < \mu \\ 0 & |\frac{r}{n}| \geq \mu \end{cases} \tag{10}$$

Figure 1 shows the plot for this function. In this cost r represents the epipolar residual of one of the terms in the GEA cost function, and n is the number of correspondences used to estimate the Ω matrix in this term. Terms in the GEA cost with a large $\Phi_n(r)$ value are likely to contain mismatchings obtained from incorrectly estimated epipolar geometries. Thanks to the robustification technique, the residual contribution of these terms is ignored during the cost optimization. Meanwhile, terms with an average absolute residual smaller than μ will usually contain matchings consistent with a valid epipolar geometry, and their contribution will remain unchanged by the loss function.

To include the loss function in our GEA implementation we did not modify the expression of the derivatives for the original cost. Instead, we changed the way the cost function is evaluated during the optimization. In our case we used a Quasi-Newton method to optimize the cost, where we simply ignored terms with large average epipolar residuals during the evaluation of the step equation. Solving the step equation, and updating the state vector was done as usual.

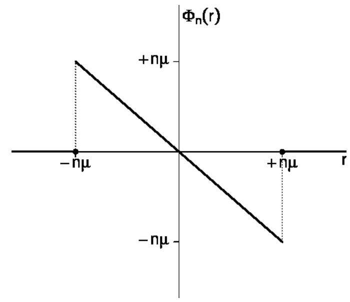

Fig. 1. Plot for the ramp loss function used in the proposed robustification method to reduce the influence of outliers in the reduced cost optimization.

As we will demonstrate in section VI, in the case of batch SfM applications the robustified GEA can successfully deal with mismatchings found in practice by image feature matching methods. In real-time SfM applications, the robustification technique can prevent motion estimation errors produced due to invalid loop closing evidence.

V. INCREMENTAL STRUCTURE-LESS MOTION ESTIMATION

We can use the new GEA implementation to develop a structure-less incremental motion estimation procedure, which obtains accurate camera poses from the pair-wise feature matchings detected at the input images using feature descriptor matching and RANSAC methods.

The procedure starts by estimating the camera poses for two views from the epipolar geometry of the feature matchings detected between them. The orientation and location of new camera poses are estimated in each iteration by averaging relative rotations [4] and camera centers [13] obtained from the epipolar geometries defined between these views, and the views already initialized. At the end of each iteration, a robustified GEA optimization corrects possible initialization errors produced by the averaging methods, and ensures convergence towards the valid optimal configuration. The process continues until convergence, obtaining accurate camera poses without merging pair-wise feature matchings into trackings, or estimating the 3D location of points in the structure.

VI. RESULTS

We tested the performance of the techniques here proposed in several experiments. In most of these experiments we used the reconstruction data-sets referenced in [1].

To evaluate the quality of the camera poses obtained with the new GEA implementation, we triangulate the structure points with these camera poses and the trackings contained in the data-sets using a linear method [7]. We then measure the reprojection error with both the camera poses and these estimated 3D scene points. The reprojection errors shown in this section were evaluated as the L_2 norm of the measurement residuals $\mathbf{r}$ for the image projections. For readability purposes, we scaled this value by a factor of 1000.0, divided by the square root of the number of elements $n_{\mathbf{r}}$ in the residuals vector.

The linear triangulation method is not completely reliable, and can produce incorrect 3D points with large residuals under certain circumstances, even if the estimated camera poses are highly accurate. For this reason, when we evaluated the reprojection error we ignored the 10% of the measurements in each view with the largest reprojection residuals. This way we ensure that the evaluated reprojection error is a fair measurement of the quality for the estimated camera poses, and not the 3D points obtained.

Our modified GEA implementation optimizes the cost function from equation 9 using the Gauss-Newton method. The step equation is solved using a preconditioned conjugate gradient. We used a symbolic math package to obtain an efficient Ω matrix evaluation code, which exploits the inherent redundancy of the terms in the summatory from equation 6. The c++ implementation of the modified GEA algorithm can be found here [1]. The tests were executed on an Intel i3 (3.20GHz) with 4Gb of RAM memory. No GPU hardware acceleration was used.

A. Reduction time speed-up

In a first set of experiments we measured the time required to reduce the data matrices with the method described in section III, and compared it with the time required by the method used in [12]. Figure 2 shows a comparison of both times, and demonstrates the significant speed increase of the proposed data reduction method.

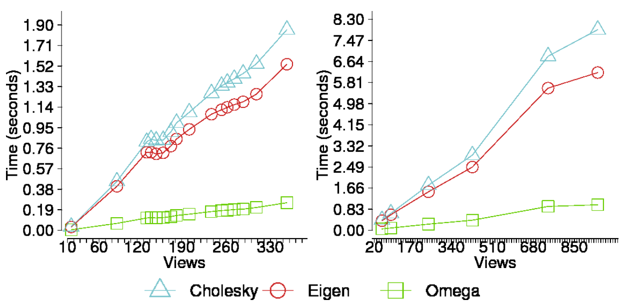

Fig. 2. Time (in seconds) required to reduce the matching information for the data-sets *dubrovnik* (left) and *venice* (right) with two methods: the one proposed in [6, 12] using both the Cholesky and eigen decompositions, and the method described here (Omega), which does not require a matrix factorization.

Figure 3 compares the optimization time for the new GEA implementation and SSBA, an efficient and freely available BA implementation [9]. As can be seen in the figure, the data reduction operation can require a significant computation time, but is not the most expensive step in GEA anymore.

B. Robustness to outliers

In a second set of experiments we evaluated the reprojection error obtained with and without the outlier robustification technique discussed here. We added to each data-set an increasing number of synthetic matchings in 10% of the view pairs. These matchings were designed to have an epipolar residual with

[1]http://perception.inf.um.es/gea

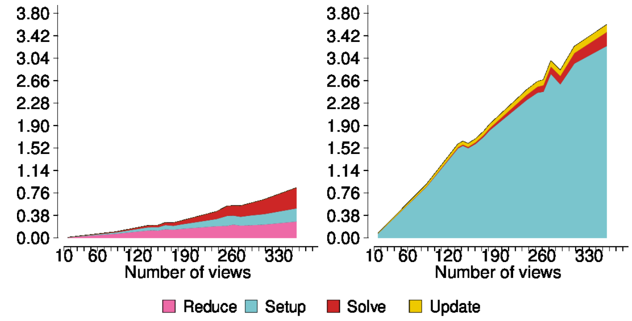

Fig. 3. Time (in seconds) required by each iteration of the cost optimization in GEA (left) and SSBA (right) for the data-set *dubrovnik*. The times are separated into the following stages. **Reduce:** data matching reduction using the technique proposed (GEA only). **Setup:** setting up the Gauss-Newton step equation. **Solve:** solving the step equation. **Update:** update the state vector (SSBA only).

the optimal GEA camera pose configuration 10 times larger than the residuals of the original matchings in the data-set. Hence, they simulate mismatchings obtained from incorrectly estimated epipolar geometries. In these tests the μ parameter for the outlier robustification was set to 10^{-4}.

Figures 4 and 5 show the accuracy of the camera poses obtained with GEA when we include these synthetic outliers in the optimization. Figure 4 compares the evolution of the error as the fraction of synthetic mismatchings increases. Either with a small or large number of outliers, the robustification technique provides accurate camera poses with a configuration very close to the optimal for the original set of matchings (mismatching free).

When the size of the problem grows, a few camera poses tend to contain large errors despite the robustification. However, the number of these incorrectly estimated camera poses is small, and the remaining camera poses are estimated accurately.

Figure 5 shows the error distribution per view for the optimal GEA configurations when the fraction of synthetic matchings is as large as 7%. Without the robustification technique, almost every camera pose estimated by GEA has a large average reprojection error. With the robustification, the error for most of the views obtained is small, almost equivalent to the error for the camera poses estimated without synthetic mismatchings. Only a small fraction of the camera poses remain after the robustified GEA optimization with a significant average reprojection error.

C. Testing the incremental motion estimation

Figure 6 shows dense reconstructions obtained from the image sets[2] *medusa* (19 frames) and *leuven castle* (28 images) using the software package PMVS2[3]. The camera poses for the images were obtained with the incremental motion estimation procedure proposed in section V. The pair-wise feature matchings used to estimate these camera poses were obtained

[2]Image sets can be found here: http://www.cs.unc.edu/~marc/
[3]http://www.di.ens.fr/pmvs/

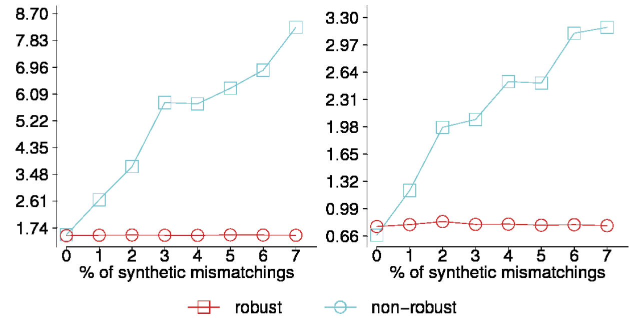

Fig. 4. Evolution of the reprojection error obtained with GEA for the data-sets *dubrovnik-135* (left) and *venice-89* (right) when we add in a 10% of the view pairs an increasing fraction of mismatchings. The figure compares the error obtained with, and without the robustification technique.

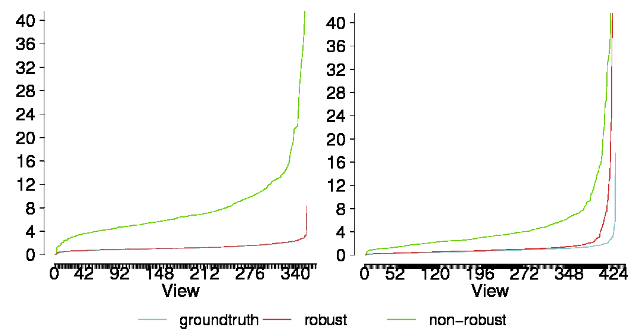

Fig. 5. Distribution of average reprojection error per view obtained with GEA for the data-sets *dubrovnik-356* (left) and *venice-427* (right), arranged from smallest to largest average error. The *groundtruth* configuration was produced by GEA using the original (outlier free) matchings in the data-sets. The *robust* and *non-robust* configurations were obtained by adding a large number of synthetic outliers to a 10% of the view pairs.

using the software package VisualSfM[4], which detects feature matchings using SIFT descriptors and checks their epipolar consistency with RANSAC.

The high quality of these reconstructions demonstrates that the motion estimation methods proposed obtain accurate camera poses, even when using input matchings which can contain outliers, as long as they are consistent with the epipolar geometry of the valid camera poses.

VII. Conclusions

We proposed a modification of the cost function expression to speed up the algebraic data reduction in structure-less BA methods. We have also discussed how to include a loss function in the cost of these methods, while keeping the computational advantages of the data reduction. In combination with the adequate image feature matching detection, the loss function can prevent the influence of mismatchings in the cost optimization, ensuring that we obtain accurate camera poses.

We have also introduced a structure-less motion estimation method, which can be used in SfM applications to initialize the camera poses efficiently. We demonstrated these results with experiments using both synthetic and real data, and hope these contributions can improve the computational efficiency of future SfM applications using structure-less BA.

[4]http://homes.cs.washington.edu/˜ccwu/vsfm/

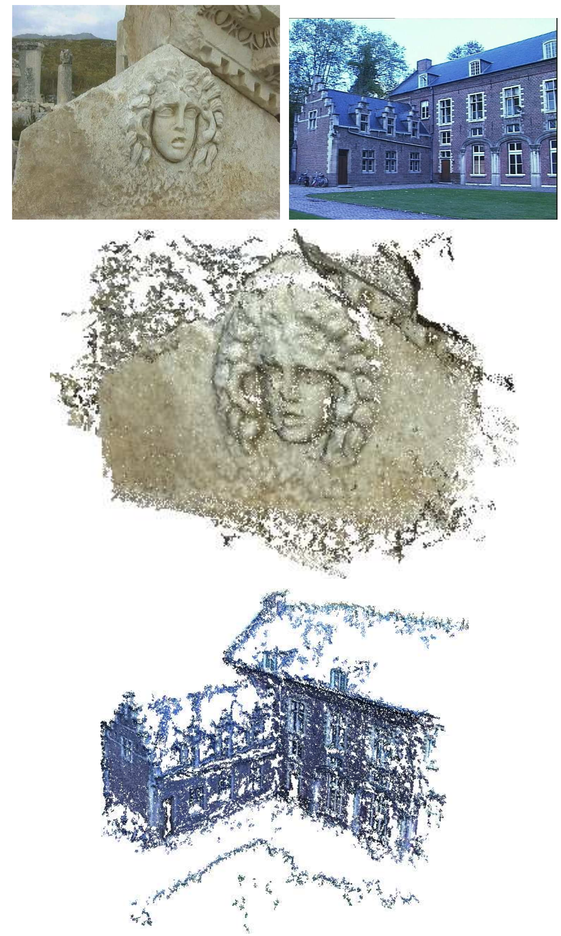

Fig. 6. **Top row:** sample images from the data-sets *medusa* and *leuven castle* (top row). **Middle and bottom:** dense reconstructions obtained for these data-sets using the PMVS2 software, and the camera poses provided by the incremental motion estimation procedure.

References

[1] Agarwal, S., Snavely, N., Seitz, S. M., and Szeliski, R. (2010). Bundle adjustment in the large. In *Proceedings of the 11th European Conference on Computer Vision*. Springer-Verlag.

[2] Byrod, M. and Astrom, K. (2010). Conjugate gradient bundle adjustment. In *Proceedings of the 11th European Conference on Computer Vision*. Springer-Verlag.

[3] Engels, C., Stewénius, H., and Nistér, D. (2006). Bundle adjustment rules. In *Photogrammetric Computer Vision*.

[4] Hartley, R., Aftab, K., and Trumpf, J. (2011). L_1 rotation averaging using the weiszfeld algorithm. In *Proceedings of the 24th Computer Vision and Pattern Recognition*.

[5] Hartley, R. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 edition.

[6] Hartley, R. I. (1998). Minimizing algebraic error in

geometric estimation problems. In *Proceedings of the 6th International Conference on Computer Vision*.

[7] Hartley, R. I., Gupta, R., and Chang, T. (1992). Stereo from uncalibrated cameras. In *Proceedings of the Computer Vision and Pattern Recognition*.

[8] Indelman, V., Roberts, R., Beall, C., and Dellaert, F. (2012). Incremental light bundle adjustment. In *Proceedings of the 23rd British Matchine Vision Conference*.

[9] Konolige, K. (2010). Sparse sparse bundle adjustment. In *Proceedings of the 21st British Matchine Vision Conference*.

[10] Rodríguez, A., López-de-Teruel, P., and Ruiz, A. (2011a). GEA optimization for live structureless motion estimation. In *Proceedings of ICCV'11 workshop on the 1st Live Dense Reconstruction from Moving Cameras*.

[11] Rodríguez, A. L. (2013). *Algebraic epipolar constraints for efficient multiview motion estimation*. PhD thesis, Universidad de Murcia, Spain.

[12] Rodríguez, A. L., López-de-Teruel, P., and Ruiz, A. (2011b). Reduced epipolar cost for accelerated incremental SfM. In *Proceedings of the 24th Computer Vision and Pattern Recognition*.

[13] Sim, K. and Hartley, R. (2006). Recovering camera motion using L_∞ minimization. In *Proceedings of the 19th Computer Vision and Pattern Recognition*.

[14] Snavely, N., Seitz, S. M., and Szeliski, R. (2006). Photo Tourism: exploring photo collections in 3D. In *Proceedings of the 33rd annual conference on Computer Graphics and Interactive Techniques*. ACM.

[15] Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W. (2000). Bundle adjustment - A modern synthesis. In *Proceedings of ICCV'99 workshop on Vision Algorithms*. Springer-Verlag.

[16] Wu, C., Agarwal, S., Curless, B., and Seitz, S. M. (2011). Multicore bundle adjustment. *Proceedings of the 24th Computer Vision and Pattern Recognition*.

[17] Zhang, Z. and Shan, Y. (2001). Incremental motion estimation through local bundle adjustment. Technical Report MSR-TR-01-54, Microsoft Research.