Minimal Solutions for Panoramic Stitching Given Gravity Prior - Supplementary Material

Yaqing Ding[1], Daniel Barath[2], Zuzana Kukelova[3]
[1] School of Computer Science and Engineering, Nanjing University of Science and Technology
[2] Computer Vision and Geometry Group, Department of Computer Science, ETH Zürich
[3] Visual Recognition Group, Faculty of Electrical Engineering, Czech Technical University in Prague
dingyaqing@njust.edu.cn

Abstract

In this supplementary document, we provide additional details for the varying focal length and distortion solvers. Sec. 1 of this document specifies derivation details and properties of the solvers. In Sec. 2, we show more experimental results for the varying focal length and distortion solvers. Also, results on undistorted images are reported.

1. Details on Solvers

1.1. Varying Focal Length Solver - H2$\mathbf{f_{12}}$(G)

H2f_{12}(G) solver assumes two images with unknown and different focal lengths. This case occurs, *e.g.*, when the images are taken with a zooming camera. This is a 3-DOF problem with respect to unknowns $\{s, f_1, f_2\}$, and we need at least 1.5 point correspondences to solve it. In practice, we still need to sample two points, but only 3 out of 4 linearly independent equations provided by two point correspondences (Eq. (6) from the main paper) are used to solve the problem. The 4^{th} equation can be used to eliminate geometrically infeasible solutions.

In this case, one point correspondence results in three equations (Eq. (6) from the main paper) of the form

$$\boldsymbol{a_1} \cdot [s^2fw, s^2f, s^2w, s^2, sfw, sf, sw, s, fw, f, w, 1]^\top = 0, \quad (1)$$
$$\boldsymbol{a_2} \cdot [s^2fw, s^2f, s^2w, s^2, sfw, sf, sw, s, fw, f, w, 1]^\top = 0, \quad (2)$$
$$\boldsymbol{a_3} \cdot [s^2f, s^2, sf, s, f, 1]^\top = 0, \quad (3)$$

with $w = \frac{1}{f_2}$ and $f = f_1$. Note, that only two out of these three equations are linearly independent, due to the rank-2 skew symmetric matrix $[\mathbf{p}^u_{2i}]_\times$ in Eq.(6) in the main paper.

Equation (3) is of degree 3 and does not contain the unknown w. Therefore, two point correspondences give us two equations of the form (3) in two unknowns $\{s, f_1\}$. These equations can be used to solve for $\{s, f_1\}$. In this case, f_1 can be easily eliminated, leading to a quartic equation in s, which can be solved in closed-form. This is done by rewriting two equations (3) as

$$\mathbf{C}(s) \begin{bmatrix} f_1 & 1 \end{bmatrix}^\top = 0 \quad (4)$$

where $\mathbf{C}(s)$ is a 2×2 coefficient matrix, and the entries of $\mathbf{C}$ are polynomials of degree 2 in the hidden variable s. Since (4) has non-trivial solutions, matrix $\mathbf{C}(s)$ must be rank-deficient. Thus, $\det(\mathbf{C}(s)) = 0$, which is a univariate polynomial equation in s of degree 4. Once s is computed, f_1 can be extracted from the null space of matrix $\mathbf{C}(s)$. Then substituting solutions for $\{s, f_1\}$ into (1) yields up to 4 possible solutions to $w = \frac{1}{f_2}$.

1.2. Varying Focal Length and Radial Distortion - H3λ_{12}f$_{12}$(G)

The final case that we consider is the most general one, in which the focal lengths and distortion parameters of the two cameras are different. This situation occurs, *e.g.*, when the images are taken using a zooming wide-angle camera. For this problem, there are a total of five unknowns $\{s, f_1, f_2, \lambda_1, \lambda_2\}$. The minimal case is 3 point correspondences, and it is similar to the 2-point solvers (H2λf(G), H2f_{12}(G)). We only need a single equation from the constraints implied by the last point correspondence.

In this case, similarly to the H2λf(G) solver, two out of the three equations coming from Eq.(6) in the main paper, are of degree 6 and one is of degree 4. However, now we have equations in five unknowns. Again only two from these three equations are linearly independent. Moreover, the equation of degree four, *i.e.*, the equation corresponding to the last row of the matrix $[\mathbf{p}^u_{2i}]_\times$, has a simpler form and contains only three unknowns $\{s, f_1, \lambda_1\}$ as follows:

$$\boldsymbol{b}[s^2f_1\lambda_1, s^2f_1, s^2, sf_1\lambda_1, sf_1, s, f_1\lambda_1, f_1, 1]^\top = 0, \quad (5)$$

where $\boldsymbol{b}$ is a coefficient vector. To simplify the solver, we use only this simpler equation (5) from the third point corre-

spondence, *i.e.* equation of degree four in three unknowns. In this way, we obtain 5 polynomial equations in 5 unknowns. However, three of them contain only three unknowns $\{s, f_1, \lambda_1\}$.

Therefore, we can first use the three equations of form (5) to solve for $\{s, f_1, \lambda_1\}$. We solve these equations using the hidden variable technique [3]. The basic idea of this technique is to consider one of variables as a hidden variable and then compute the resultant. We treat s as the hidden variable, *i.e.*, we hide it into the coefficient matrix. The three equations (5) coming from the three correspondences can be expressed in terms of the monomials $\{f_1, \lambda_1, f_1, 1\}$ as

$$\mathbf{C}(s)[f_1\lambda_1 \; f_1 \; 1]^\top = \mathbf{0}, \tag{6}$$

where $\mathbf{C}$ is of size 3×3 with entries that are polynomials in the hidden variable s. In this case, $\mathbf{C}(s)$ has the form

$$\mathbf{C}(s) = \begin{bmatrix} C_1^{[2]}(s) & C_2^{[2]}(s) & C_3^{[2]}(s) \\ C_4^{[2]}(s) & C_5^{[2]}(s) & C_6^{[2]}(s) \\ C_7^{[2]}(s) & C_8^{[2]}(s) & C_9^{[2]}(s) \end{bmatrix}, \tag{7}$$

where the upper index $^{[\cdot]}$ denotes the degree of the respective polynomial $C_i(s)$. Since (6) has a non-trivial solution, matrix $\mathbf{C}(s)$ must be rank-deficient. Thus, $\det(\mathbf{C}(s)) = 0$, which is an univariate sextic equation in s that can be solved using Sturm sequences. Once s is known, f_1, λ_1 can be extracted from the null space of the matrix $\mathbf{C}(s)$. Finally, f_2 and λ_2 can be extracted from the remaining two equations.

1.3. Special case

If the y-axis of the camera is considered to be physically aligned with the gravity direction, we have $\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{I}$. In this case, the homography matrix $\mathbf{G}$ can be simplified as

$$\mathbf{G} = \widetilde{\mathbf{K}}_2 \mathbf{R}_y \widetilde{\mathbf{K}}_1^{-1}. \tag{8}$$

With this formulation, all the problems result in solving one quadratic equation.

1.4. Computational Complexity

The complexity and run-time of a single estimation of the new varying focal length and radial distortion solvers as well as three state-of-the-art solvers for varying f and λ are reported in the following table

Similar to the main paper we only show the major steps performed by each solver. The number in the cells, *e.g.* 6×8, denotes the matrix size to which the G-J elimination or Eigen-decomposition is applied. The number in the fourth column denotes the degree of the univariate polynomial that needs to be solved.

The properties of all these five solvers, e.g. the number of solutions, DOF, reference, etc., are listed in Table 1

Solver	G-J	Eigen	Poly	Time (μs)
H3$f_{1,2}$	-	7×7	-	7
H5$\lambda_{1,2}$	16×21	5×5	-	12
H6$\lambda_{1,2}$	6×8	-	2	14
H2$f_{1,2}$(G)	-	-	4	5
H3$\lambda_{1,2}f_{1,2}$(G)	-	-	6	5

2. Additional Experiments

Synthetic evaluation. The synthetic data were generated similar to the main paper. The only difference is that the focal length of the second camera was set to 500 pixels. To show the benefits of the varying focal length (distortion) solvers, we use the best fixed focal length (distortion) solvers as the baseline. Fig. 1a reports the focal length (left column) and rotation errors (right) for the solvers assuming zero distortion. Fig. 1b reports the relative focal length (left) and absolute radial distortion error (right). We can see that the varying focal length (distortion) solvers perform significantly better than the fixed focal length (distortion) solvers. In addition, our new solvers outperform the SOTA for most of the cases.

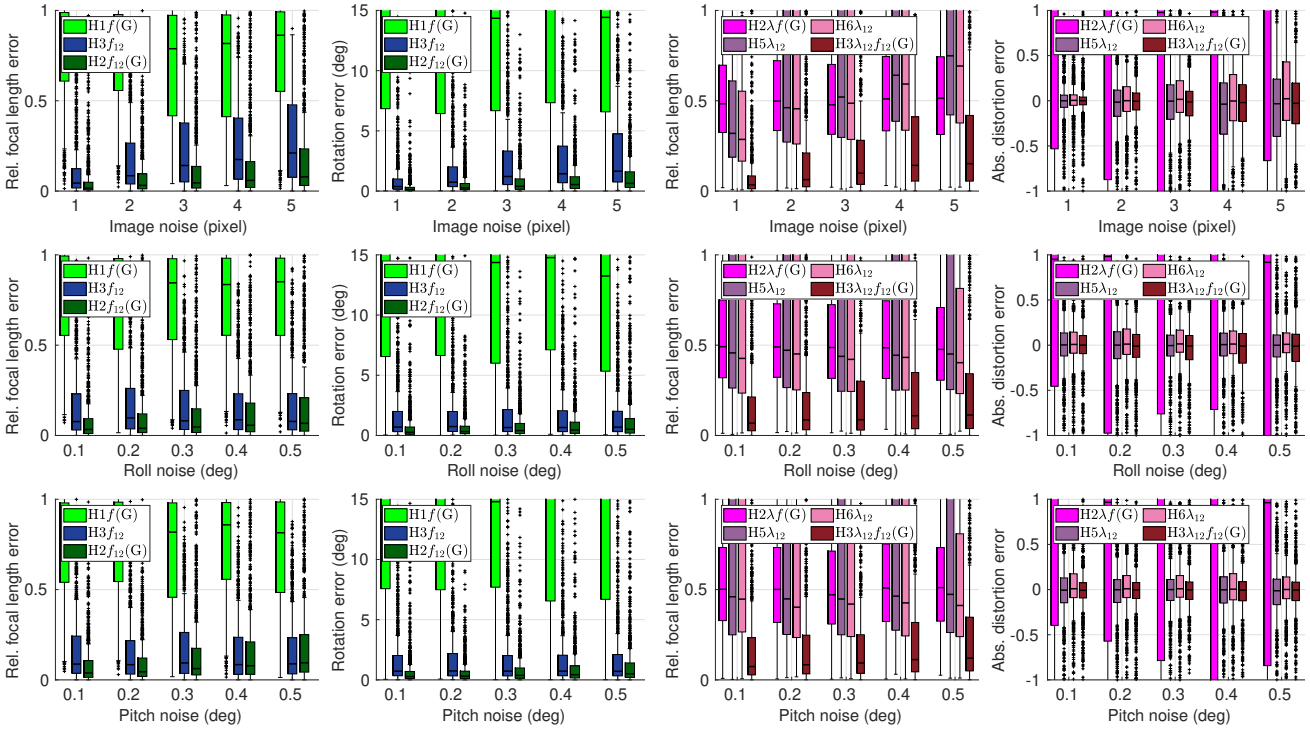
Undistorted images. Since in the SUN360 dataset the ground truth radial distortion parameters are known, we used them to undistort the images. Fig. 2a shows the cumulative distribution functions of the average run-times (in seconds), re-projection errors (in pixels), and relative focal length errors on $579,800$ image pairs. We do not show the results of methods estimating the radial distortion. The proposed H1f(G) solver is the fastest while leading to the most accurate homographies and focal lengths.

Varying radial distortion and focal length solvers. In general, finding or capturing stitching data with varying focal length and radial distortion is challenging. This stems from the fact that stitching is usually done by a single camera rotating around a 3D axis with a fixed focal length and radial distortion. Still, there are application fields where such solvers can be useful in practice, *e.g.*, calibrating zooming PTZ cameras or rotating zooming surveillance cameras. We, however, have not found publicly available datasets for such problems. Therefore, we ran the varying focal length and radial distortion solvers on the SUN360 and the captured phone datasets, where both the focal length and radial distortion are fixed.

In Fig. 2b, the cumulative distribution functions of all solvers, including the ones assuming varying radial distortion or focal length, on the distorted images from the SUN360 dataset and, also, on the captured phone dataset are shown. The additional solvers that we include in these experiments and were excluded from the main paper

	H3$f_{1,2}$	H5λ_{12}	H6λ_{12}	H2f_{12}(G)	H3$\lambda_{12}f_{12}$(G)
reference	[2]	[4]	[4]		
Different focal lengths	✓	✓	✓	✓	✓
Radial distortion		✓	✓		✓
Different distortions		✓	✓		✓
Number of points	3	5	6	2	3
Number of solutions	7	5	2	4(2)	6(2)
Gravity prior				✓	✓
Pure $\mathbf{R}$	✓			✓	✓
DOF	5	10	10	3	5

Table 1. The properties of the proposed gravity-based (gray) and state-of-the-art solvers. The number of solutions in brackets refers to a special case when the y-axis of the camera is considered to be physically aligned with the gravity direction.



(a) No radial distortion.

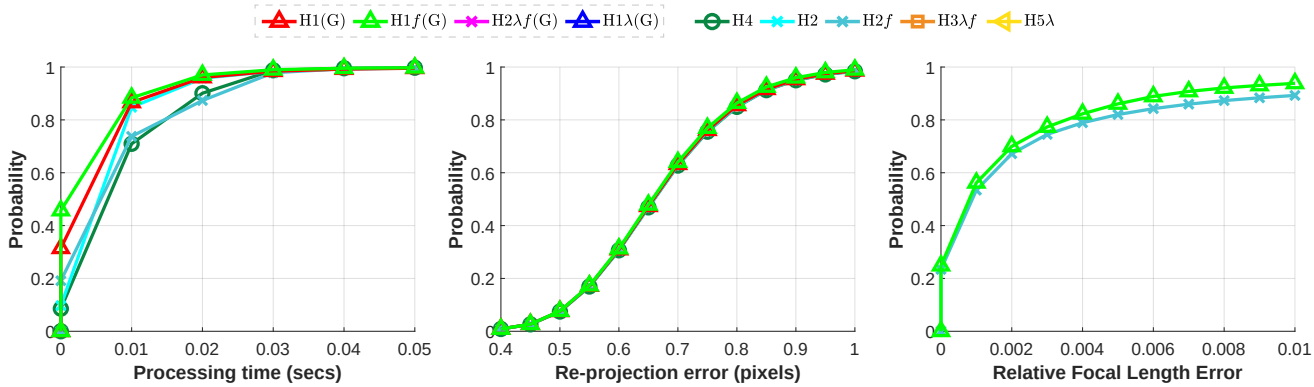
(b) Radial distortion.

Figure 1. **(a)** Boxplot of the relative focal length error and rotation error for the zero distortion case. **Left column**: Relative focal length error. **Right column**: Rotation error. **(b)** Boxplot of the relative focal length error and absolute distortion error for the distortion solvers. **Left column**: Relative focal length error. **Right column**: Absolute distortion error. From top to bottom: increased image measurement noise, increased roll noise and 2 pixel standard deviation image measurement noise, increased pitch noise with constant 2 pixel standard deviation image measurement noise.

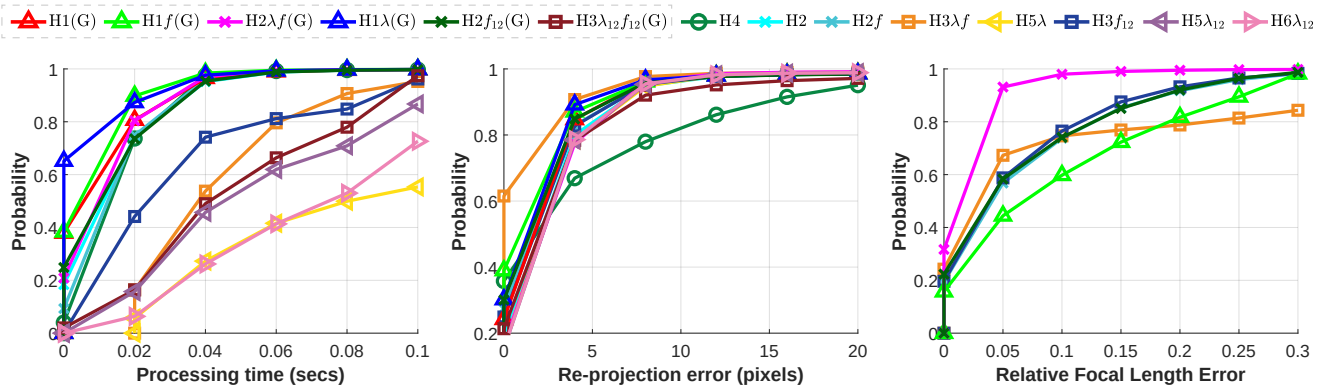
are H3$f_{1,2}$ [2], H5$\lambda_{1,2}$ [4], H6$\lambda_{1,2}$ [4] and the proposed H2$f_{1,2}$(G) and H3$\lambda_{1,2}f_{1,2}$(G) algorithms.

On the SUN360 dataset, the added solvers do not change the conclusions from the main paper – they perform reasonably well but none of them are the most accurate nor the fastest ones. This is to be expected since the SUN360 dataset contains images with the same radial distortion and focal length. As expected due to the large sample size, H5$\lambda_{1,2}$ and H6$\lambda_{1,2}$ are among the slowest methods when used for fitting to minimal samples.

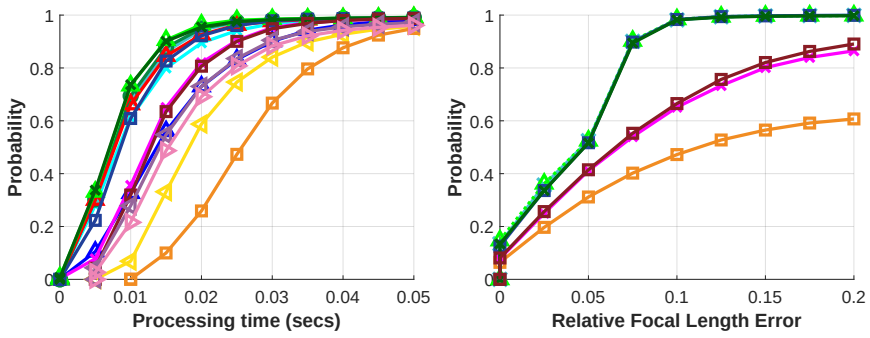
On the captured phone dataset, the proposed H2$f_{1,2}$(G) solver performs remarkably efficiently and accurately. Together with the proposed H1f(G) algorithm, they are the fastest and most accurate methods for image stitching in the semi-calibrated case, *i.e.*, with unknown focal length.



(a) Undistorted images. Solvers not estimating radial distortion.



(b) Distorted images. All solvers.



(c) Phone dataset. All solvers.

Figure 2. The cumulative distribution functions of the processing times (in seconds), average re-projection errors (in pixels) and relative focal length errors of GC-RANSAC [1] when combined with different minimal solvers. In the top two rows, the values are calculated from a total of $579,800$ image pairs from the SUN360 dataset. In the bottom row, the values are calculated from a total of $7,770$ image pairs from the captured phone dataset. The confidence was set to 0.99 and the inlier threshold to 3 px. Being accurate is interpreted as a curve close to the top-left corner.

3. Example Panoramic Stitching Results

Example panoramic stitching results on sequences from the SUN360 and Smartphone datasets are shown in Fig. 3. The image rotations are estimated by the proposed algorithm. Bundle adjustment then minimized the re-projection error throughout the entire sequence. The images are warped together considering spherical views, a Graph-Cut-based seam finder [5] is applied, and a multiband blender blends the images together.

(a) Example result on a sequence from the SUN360 dataset [6].

(b) Example result on a sequence from the captured Smartphone dataset.

Figure 3. Example stitching results on sequences from the SUN360 and Smartphone datasets. The image rotations are estimated by the proposed algorithm. Bundle adjustment then minimized the re-projection error throughout the entire sequence. The images are warped together considering spherical views, a Graph-Cut-based seam finder [5] is applied, and a multiband blender blends the images together.

References

[1] Daniel Barath and Jiří Matas. Graph-cut RANSAC. In *Computer Vision and Pattern Recognition (CVPR)*, 2018. 4

[2] Matthew Brown, Richard I Hartley, and David Nistér. Minimal solutions for panoramic stitching. In *Computer Vision and Pattern Recognition (CVPR)*, 2007. 3

[3] David A Cox, John Little, and Donal O'shea. *Using algebraic geometry*. Springer Science & Business Media, 2006. 2

[4] Zuzana Kukelova, Jan Heller, Martin Bujnak, and Tomas Pajdla. Radial distortion homography. In *Computer Vision and Pattern Recognition (CVPR)*, 2015. 3

[5] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 30(6):1068–1080, 2008. 4, 5

[6] Jianxiong Xiao, Krista A Ehinger, Aude Oliva, and Antonio Torralba. Recognizing scene viewpoint using panoramic place representation. In *Computer Vision and Pattern Recognition (CVPR)*, 2012. 5