FAST AND ROBUST HOMOGRAPHY ESTIMATION BY ADAPTIVE GRADUATED NON-CONVEXITY

Gaku Nakano and Takashi Shibata

Central Research Labs, NEC Corporation, Kawasaki, Japan
g-nakano@cq.jp.nec.com, t-shibata@hw.jp.nec.com

ABSTRACT

This paper proposes a novel fast and robust homography estimation by adaptively controlling the threshold of graduated non-convexity (GNC). Based on the fact that GNC is a variant of deterministic annealing, we provide a new method for updating the inlier threshold at each GNC iteration by utilizing the statistical properties of residuals of potential inliers. Contrary to RANSAC, our approach gives the same unique parameter for a single input due to without random sampling. Moreover, computational time increases linearly against outlier ratio changes, whereas RANSAC increases exponentially. Synthetic data evaluation shows that the proposed method is more robust and faster than RANSAC for highly contaminated data containing more than 80% outliers. Additionally, we demonstrate that our method works on severe real images that the state-of-the-art RANSAC method fails.

Index Terms— Homography, Robust Estimation, Graduated Non-Convexity, RANSAC

1. INTRODUCTION

Homography estimation between two images of a 3D plane is a fundamental problem in computer vision. A typical way to find a homography is to use a set of point pairs by matching feature points on two images. Those point correspondences generally consist of two groups: inliers that satisfy a homography equation and outliers that wrongly matched points. It is not a rare case where outlier ratio exceeds 50%, therefore, homography estimation needs to be robust against outliers for practical applications.

Robust parameter estimation from noisy data contaminated by outliers has been extensively studied over the last few decades. Typical methods are random sample consensus (RANSAC) [1], M-estimator [2], and iteratively reweighted least squares (IRLS) [3].

RANSAC has been virtually the standard method for problems of geometric parameter estimation, *e.g.*, line fitting and multi-view geometry. The main reason is that its algorithm simpleness and capability on highly contaminated data consist of more than 50% outliers. Despite being widely used, RANSAC has several unsolved drawbacks [4] caused by random sampling and a predefined threshold. To deal with those issues, numerous variants have been proposed for accelerating computational time [5, 6], improving solution stability [7–9], and determining threshold automatically [10–12]. Some of these methods can be merged into a single framework. Since introducing more techniques generally requires handling more parameters, integration have to be done carefully so that the simpleness of RANSAC scheme would not be lost. Actually, USAC [13], which is an unified implementation of RANSAC variants, fails to compute homography on images with a significant viewpoint change [12].

IRLS has been widely used for conducting regularized least squares [14, 15]. Since IRLS and M-estimator are mathematically equivalent as proved by Black and Rangarajan [3], IRLS-based methods are also sensitive to the threshold. However, it is heuristically determined depending on the problem. One way to robustify IRLS is to incorporate graduated non-convexity (GNC) [16], which smooths a non-convex objective function by gradually decreasing a large threshold. In existing methods of IRLS with GNC, the threshold is updated by a simple rule, *i.e.* just multiplying a predefined scale factor at each iteration. On the other hand, how to determine the decreasing factor has not been discussed well yet.

Recently, deep neural networks have been applied to geometric problems [17–19]; however, the conventional approaches based on optimization theory still have advantages against those deep-based methods in terms of theoretical aspects and computational efficiency.

This paper proposes an IRLS method with a novel adaptive GNC for fast and robust homography estimation. The proposed IRLS-GNC method has two advantages against RANSAC: a unique solution for each trial, and a moderate increase of computational time on outlier ratio. First, we point out that minimizing an objective function of IRLS is equivalent to jointly finding the best homography and maximizing the number of inliers under a given threshold. Then, a new adaptive update rule is derived for determining the GNC threshold by integrating the classical constant decreasing with the proposed method based on residual analysis of potential inliers. By conducting a synthetic and a real data experiment, we show that IRLS with the proposed GNC method outperforms RANSAC and USAC.

The final published version is available at https://doi.org/10.1109/ICIP.2019.8803337.

2. PROPOSED METHOD

This section begins with interpreting the IRLS formulation for homography estimation. Then, we provide a method for adaptively determining the threshold at each GNC iteration. Finally, we describe how to obtain the best result after GNC is stopped. The whole procedure is outlined in Algorithm 1.

2.1. Interpretation of IRLS on homography estimation

IRLS is an optimization method for solving regularized least squares problems by iteratively updating a weight $w_i \in [0, 1]$ of an i-th residual r_i. IRLS for estimating an unknown parameter vector $\boldsymbol{\theta}$ can be written of the form

$$\min_{\boldsymbol{\theta}, \mathbf{w}} E(\boldsymbol{\theta}, \mathbf{w}) = \sum_{i=1}^{n} w_i \, r_i^2 + \lambda \, \Phi(w_i), \tag{1}$$

where $\mathbf{w} = [w_1, \ldots, w_n]^\mathsf{T}$, λ is a nonnegative scalar that controls the strength of $\Phi(w_i)$, and $\Phi(w_i)$ is a function that behaves $\Phi(w_i) \to 0$ as $w_i \to 1$ and $\Phi(w_i) \to 1$ as $w_i \to 0$.

For minimizing Eq. (1), IRLS conducts an alternate optimization over $\boldsymbol{\theta}$ and $\mathbf{w}$, instead of solving $\partial E / \partial \boldsymbol{\theta} = 0$ and $\partial E / \partial w_i = 0$, simultaneously. First, we solve $\partial E / \partial \boldsymbol{\theta} = 0$ w.r.t. $\boldsymbol{\theta}$ by fixing w_i. Then, substituting $\boldsymbol{\theta}$ into $\partial E / \partial w_i = 0$, we update w_i. Iterating the above procedure until convergence, we finally obtain the solution of $\boldsymbol{\theta}$.

The first term, $\sum w_i \, r_i^2$, is called data term that represents the sum of weighted squared residuals. Thus, minimizing $\sum w_i \, r_i^2$ w.r.t. $\boldsymbol{\theta}$ is equivalent to finding a homography parameter that fits to point correspondences having a high weight. The second term, $\lambda \sum \Phi(w_i)$, is known as regularization term, which prevents the first term not to be overly minimized.

Since a weight w_i is given by its corresponding residual r_i with λ and $\partial \Phi / \partial w_i$, λ can be interpreted as the threshold that determines whether the i-th point pair $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$ is an inlier or an outlier, similarly to RANSAC. Addition to that, considering the weight range $0 \leq w_i \leq 1$, we can see that a weight w_i represents the inlier probability of the i-th point.

From this, minimizing $E(\boldsymbol{\theta}, \mathbf{w})$ with a certain λ can be interpreted that simultaneously finding the best homography transformation and maximizing the number of inliers.

2.2. Adaptive GNC

The key to obtain a good solution to $E(\boldsymbol{\theta}, \mathbf{w})$ of Eq. (1) is to choose an adequate value of λ. The simplest way is to use a constant value like M-estimators; however, such methods often get trapped in an undesirable local minimum because $E(\boldsymbol{\theta}, \mathbf{w})$ is generally non-convex. To avoid the local minimum issue, Blake and Zisserman [16] proposed graduated non-convexity (GNC) written by

$$\lim_{\lambda_{\max} \to \lambda_{\min}} \min_{\boldsymbol{\theta}, \mathbf{w}} E(\boldsymbol{\theta}, \mathbf{w}). \tag{2}$$

As shown in Eq. (2), GNC is a kind of deterministic annealing [20] starting with a very large $\lambda_{\max}$ to make $E(\boldsymbol{\theta}, \mathbf{w})$ smooth and close to convex. Then, λ is gradually decreased until reaching to the minimum value $\lambda_{\min}$.

At each IRLS iteration, we have potential inliers whose weight is greater than 0.5. On the assumption of least squares, residuals of the potential inliers follow a normal distribution of mean μ and standard deviation σ. Based on that, we can determine λ as the confidence interval, $\lambda = \mu + \beta\sigma$, so that low potential inliers having large residuals will be regarded as outliers at the next iteration. In short, a point having $|r_i| < \mu + \beta\sigma$ should be an inlier and a point having $|r_i| > \mu + \beta\sigma$ should turn to be an outlier.

The convergence speed with the above update rule is sometimes slow due to very small difference between the old and the new thresholds. Specifically, it happens in either case of high outlier ratio or large β or close to $\lambda_{\min}$. To accelerate this, we introduce two techniques.

The first technique is to use a constant decreasing, $c\lambda$, as in the conventional GNC methods. Taking the smaller of $\mu + \beta\sigma$ and $c\lambda$, we can obtain the better threshold that surely proceeds the convergence. The second one is to prepare the minimum update value δ in case of that the update difference by the first technique is too small. For example, if the difference is less than 0.5 pixels when using a geometric cost function, the result of the next iteration is predicted to be almost same to that of the current iteration.

To summarize, the proposed adaptive GNC can be written as follows:

$$\lambda_{\text{temp}} = \min(c\lambda, \; \mu + \beta\sigma), \tag{3}$$

$$\lambda_{\text{new}} = \begin{cases} \lambda_{\text{temp}} - \delta & \text{if } \lambda_{\text{temp}} - \lambda < \delta \\ \lambda_{\text{temp}} & \text{else} \end{cases} \tag{4}$$

2.3. How to choose the best result

After performing GNC with the new update rule described in Section 2.2, the iterative procedure finally stops at $\lambda_{\min}$. Note that $\lambda_{\min}$ is not the best threshold that maximizes the number of inliers since reaching to $\lambda_{\min}$ is merely a stopping criterion for GNC.

Inspired by Litman *et al*. [12], we assume that the best inlier ratio is not sensitive to the perturbation of a threshold λ. Since λ decreases monotonically and smoothly according to Eqs. (3) and (4), if the above assumption does not hold, the inlier ratio has several peaks against different thresholds. Therefore, we determine the best threshold λ_{best} that gives the best inlier ratio by checking the slope of the threshold–inlier ratio curve.

At an iteration, we can estimate the current inlier ratio p by the cumulative average of the weight vector $\mathbf{w}$:

$$p = \frac{1}{n} \sum_{i=1}^{n} w_i. \tag{5}$$

Algorithm 1 Homography Estimation by Adaptive GNC

Input: 2D point correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$ $(1 \le i \le n)$
Output: Homography $\mathtt{H}$, weight $\mathbf{w}$, threshold λ

$\mathtt{H} \leftarrow \mathtt{I}$, $\quad \mathbf{w} \leftarrow [1, \ldots, 1]^\mathsf{T}$, $\quad \lambda \leftarrow \lambda_{\max}$, $\quad p \leftarrow 1$, $\quad \Delta p_{\min} \leftarrow \infty$
while $\lambda \ge \lambda_{\min}$
  $\mathbf{r} \leftarrow \mathtt{computeResiduals}(\mathtt{H},\ \mathbf{x},\ \mathbf{x}')$
  $\mathbf{w} \leftarrow \mathtt{calcWeight}(\lambda,\ \mathbf{r})$
  $\mathtt{H} \leftarrow \mathtt{optimizeHomography}(\mathtt{H},\ \mathbf{x},\ \mathbf{x}',\ \mathbf{w})$
  $[\mu,\ \sigma] \leftarrow \mathtt{calcMeanStd}(\ \mathtt{sqrt}(\mathbf{r}(\mathbf{w} > 0.5).\hat{}2)\ )$
  $\lambda_{\text{temp}} \leftarrow \min(c\lambda,\ \mu + \beta\sigma)$
  if $\lambda_{\text{temp}} - \lambda < \delta$
    $\lambda_{\text{new}} \leftarrow \lambda_{\text{temp}} - \delta$
  else
    $\lambda_{\text{new}} \leftarrow \lambda_{\text{temp}}$
  end
  $p_{\text{new}} \leftarrow 1/n \sum w_i$ ▷ Inlier ratio estimation
  $\Delta p \leftarrow |p_{\text{new}} - p| / |\lambda_{\text{new}} - \lambda|$ ▷ Slope of λ–p curve
  if $\Delta p < \Delta p_{\min}$ ▷ Store tentative best
    $\mathtt{H}_{\text{best}} \leftarrow \mathtt{H}$, $\quad \mathbf{w}_{\text{best}} \leftarrow \mathbf{w}$, $\quad \lambda_{\text{best}} \leftarrow \lambda_{\text{new}}$, $\quad \Delta p_{\min} \leftarrow \Delta p$
  end
  $\lambda \leftarrow \lambda_{\text{new}}$, $\quad p \leftarrow p_{\text{new}}$
end
return $\mathtt{H}_{\text{best}}$, $\quad \mathbf{w}_{\text{best}}$, $\quad \lambda_{\text{best}}$

The slope between the current and the next iteration is given by

$$\Delta p = \frac{|p_{\text{new}} - p|}{|\lambda_{\text{new}} - \lambda|}, \tag{6}$$

where p_{new} and λ_{new} are an inlier ratio and a threshold at the next iteration, respectively. The current parameters are stored as the tentative best if Δp is smaller than the smallest slope $\Delta p_{\min}$, then $\Delta p_{\min}$ is replaced with Δp. Consequently, we can obtain the best parameters after the convergence.

See Fig. 1, which visualizes the proposed method described in Sections 2.2 and 2.3. In this example, the best threshold λ_{best} is determined as 26 pixels at the 68th iteration when the smallest slope is observed. After that, inliers are overly trimmed out due to too small thresholds.

3. EXPERIMENT

In this section, we report two experimental results on synthetic data and real data. First, by the synthetic data evaluation, we mainly discuss the performance of the proposed approach applied with different cost functions. Then, we choose promising combinations and compare with the existing RANSAC-based methods in real data evaluation.

We configured the GNC parameters by preliminary experiments that $c = 0.95$, $\beta = 2$, $\lambda_{\max} = 10^4$, $\lambda_{\min} = 1$, and $\delta = 0.5$. Also, we used the truncated L2 function $\Phi(w_i) = w_i - 1$ by considering the similarity with RANSAC. All experiments were conducted on MATLAB and Core i7-6700.

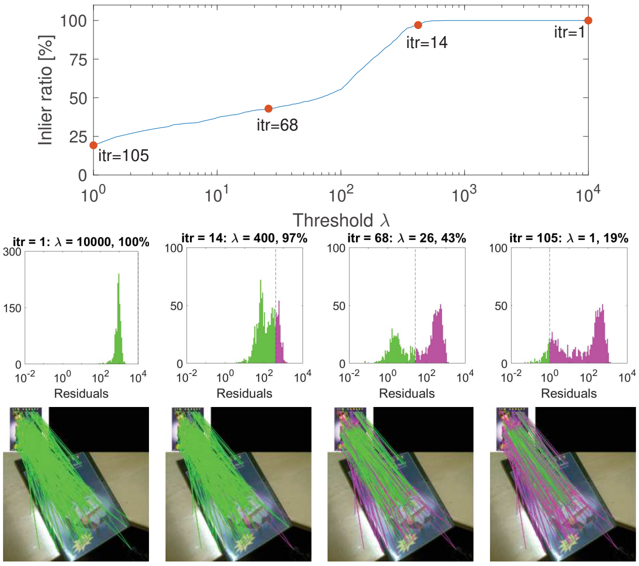

Fig. 1. *Top*: The threshold–inlier ratio curve. *Middle*: Residual histograms split into inliers (green) and outliers (purple) by λ (blue dotted line). *Bottom*: Illustrations of inlier/outlier pairs corresponding to the middle figures.

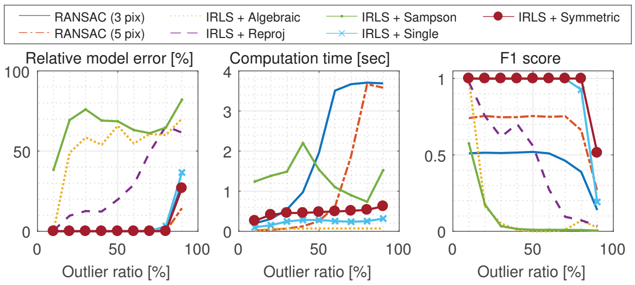

Fig. 2. Results of the synthetic data evaluation.

3.1. Synthetic data evaluation

This section describes a quantitative evaluation on the performance of the proposed method by comparing to the naive RANSAC[1] in synthetic data experiments. We implemented five variations of IRLS with the adaptive GNC corresponding to five cost functions [21]: algebraic, Sampson, reprojection, single transfer, and symmetric transfer errors. RANSAC was set to have 10^4 maximum iterations and 3– or 5–pixel threshold. We randomly generated a homography transformation induced by 1000 point correspondences with Gaussian noise of zero mean and 2.0 standard deviation. Then, we varied outlier ratio from 10% to 90% to measure the following criteria: the relative homography error, computational time, and F1 score[2] of predicted inliers.

[1]Implemented by `http://www.peterkovesi.com/matlabfns`
[2]Harmonic mean of the precision and the recall rates of predicted inliers.

(a) Graf (155, 27%) (b) Bark (341, 77%) (c) DayNight (929, 12%) (d) Cap. Reg. (994, 12%)

Fig. 3. Real image dataset with the number of the point correspondences and the inlier ratio.

Table 1. Quantitative results on the real images shown in Fig. 3.

Image	IRLS+Single			IRLS+Symmetric			RANSAC			USAC		
	error	time	F1	error	time	F1	error	time	F1	error	time	F1
(a)	0.52	90	0.78	2.47	235	0.74	2.55	1495	0.73	2.04	11	0.52
(b)	0.03	18	0.92	0.02	166	0.92	0.02	4	0.91	0.02	5	0.92
(c)	0.001	185	0.99	1.36	604	0.91	30.3	3518	0.70	NaN	9	0
(d)	92.8	230	0.18	1.23	657	0.98	11.8	3670	0.29	NaN	10	0

Figure 2 shows the results of average values over 100 independent trials for each outlier ratio. Among the five IRLS methods, those using the single transfer error and the symmetric transfer error gave accurate results in terms of parameter estimation, which is comparable with RANSAC while maintaining a moderate increase of runtime against the outlier ratio change. In addition, the F1 score indicates that the two IRLS methods are able to obtain more inliers than RANSAC correctly and stably.

Interestingly, the five proposed methods have different tendencies. Using the two non-geometric cost functions, the algebraic and the Sampson errors, result in bad convergence even for low contaminated data, such as $< 30\%$ outliers. On the other hand, although the reprojection error is geometrically meaningful, its optimization is not as stable as the single and the symmetric transfer error minimization. The cause of the instability can be considered that the reprojection error minimization was trapped into local minima due to many unknown variables. These synthetic data experiments reveal that selecting a proper cost function is crucial for IRLS-based methods to achieve good performance.

3.2. Real data evaluation

According to Section 3.1, we selected two proposed methods, IRLS+Single and IRLS+Symmetric, to be evaluated in this experiment. Then, we compared these methods with RANSAC and USAC [13]. RANSAC and USAC were configured to have 10^4 maximum iterations and a 3–pixel threshold, and other parameters were set by default values. We picked four image pairs from publicly available datasets [13, 22], as shown in Fig. 3. Similarly to the synthetic data experiments, we measured the same three criteria to evaluate the existing and the proposed methods.

Table 1 summarizes quantitative results for each image pair. Note that the computational time is denoted in milliseconds. Among the four methods, the proposed IRLS+Symmetric only works for all four images. Despite that USAC utilizes many techniques to complement the weakness of RANSAC, it fails on the two images, Figs. 3(c) and 3(d). From this, we can say that RANSAC variant integration becomes unstable unless their parameters are carefully tuned depending on input data. Strictly speaking, the computational time evaluation is not a fair comparison because USAC is written in C++ and the others in MATLAB. However, as described in the previous section, the computational time by IRLS+Single and IRLS+Symmetric increases moderately against the increase of outlier ratio. Therefore, the proposed IRLS-based methods are expected to be sufficiently fast for real-time applications even with highly contaminated data if implemented on C++. Other three methods are better than IRLS+Symmetric in some aspects; however, we can conclude that IRLS+Symmetric is superior to them by considering the trade-off issues.

4. CONCLUSIONS

In this paper, we have presented a fast and robust homography estimation using an IRLS with an adaptive GNC. We have validated by experiments that the proposed adaptive GNC with symmetric transfer error minimization outperforms RANSAC and USAC on challenging dataset, where outlier ratio is more than 80%. Although this paper addresses only on homography estimation, our proposed IRLS approach could be applied on other parameter estimation problems in computer vision as well, such as plane fitting, PnP problem, and fundamental matrix estimation.

5. REFERENCES

[1] Martin A Fischler and Robert C Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[2] Zhengyou Zhang, "Parameter estimation techniques: A tutorial with application to conic fitting," *Image and vision Computing*, vol. 15, no. 1, pp. 59–76, 1997.

[3] Michael J Black and Anand Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *International journal of computer vision*, vol. 19, no. 1, pp. 57–91, 1996.

[4] Sunglok Choi, Taemin Kim, and Wonpil Yu, "Performance evaluation of ransac family," *Journal of Computer Vision*, vol. 24, no. 3, pp. 271–300, 1997.

[5] D.R. Myatt, P.H.S. Torr, S.J. Nasuto, J.M. Bishop, and R. Craddock, "Napsac: High noise, high dimensional robust estimation - it's in the bag," in *Proceedings of the British Machine Vision Conference*. 2002, pp. 44.1–44.10, BMVA Press, doi:10.5244/C.16.44.

[6] Ondrej Chum and Jiri Matas, "Matching with prosac-progressive sample consensus," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 220–226.

[7] Jean-Charles Bazin, Hongdong Li, In So Kweon, Cédric Demonceaux, Pascal Vasseur, and Katsushi Ikeuchi, "A branch-and-bound approach to correspondence and grouping problems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 7, pp. 1565–1576, 2013.

[8] Tat-Jun Chin, Pulak Purkait, Anders Eriksson, and David Suter, "Efficient globally optimal consensus maximisation with tree search," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2413–2421.

[9] Dylan Campbell, Lars Petersson, Laurent Kneip, and Hongdong Li, "Globally-optimal inlier set maximisation for simultaneous camera pose and feature correspondence," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1–10.

[10] Rahul Raguram and Jan-Michael Frahm, "Recon: Scale-adaptive robust estimation via residual consensus," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1299–1306.

[11] Jongmoo Choi and Gérard Medioni, "Starsac: Stable random sample consensus for parameter estimation," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 675–682.

[12] Roee Litman, Simon Korman, Alexander Bronstein, and Shai Avidan, "Inverting ransac: Global model detection via inlier rate estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5243–5251.

[13] Rahul Raguram, Ondrej Chum, Marc Pollefeys, Jiri Matas, and Jan-Michael Frahm, "Usac: a universal framework for random sample consensus," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 2022–2038, 2013.

[14] Anat Levin, Rob Fergus, Frédo Durand, and William T Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM transactions on graphics (TOG)*, vol. 26, no. 3, pp. 70, 2007.

[15] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun, "Fast global registration," in *European Conference on Computer Vision*. Springer, 2016, pp. 766–782.

[16] Andrew Blake and Andrew Zisserman, *Visual reconstruction*, MIT press, 1987.

[17] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich, "Deep image homography estimation," *arXiv preprint arXiv:1606.03798*, 2016.

[18] Benjamin Ummenhofer, Huizhong Zhou, Jonas Uhrig, Nikolaus Mayer, Eddy Ilg, Alexey Dosovitskiy, and Thomas Brox, "Demon: Depth and motion network for learning monocular stereo," in *IEEE Conference on computer vision and pattern recognition (CVPR)*, 2017, vol. 5, p. 6.

[19] Ignacio Rocco, Mircea Cimpoi, Relja Arandjelović, Akihiko Torii, Tomas Pajdla, and Josef Sivic, "Neighbourhood consensus networks," in *Advances in Neural Information Processing Systems*, 2018, pp. 1658–1669.

[20] Davi Geiger and Federico Girosi, "Parallel and deterministic algorithms from mrfs: Surface reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 5, pp. 401–412, May 1991.

[21] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[22] Gehua Yang, Charles V Stewart, Michal Sofka, and Chia-Ling Tsai, "Registration of challenging image pairs: Initialization, estimation, and decision," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 11, 2007.