

# Fast Segmentation of Multiple Motions

Reza Hoseinnezhad<sup>1</sup>, Ba-Ngu Vo<sup>1</sup>, David Suter<sup>2</sup>

1: The University of Melbourne, Australia

2: The University of Adelaide, Australia

**Abstract:** A guided sampling method for robust segmentation of multiple motions is introduced. It is substantially faster than random sampling as it effectively makes use of the spatial proximity of the points belonging to each motion. A fast high-breakdown robust estimator called Guided-LKS (GLKS) is devised using the guided search to minimize the  $k$ -th order statistics of squared distances. A number of experiments on homography estimation problems are presented. They involve up to eight motions and benchmark the performance of GLKS estimator in comparison to a number of state of the art robust estimators. The results show that while GLKS performs similar to other estimators in terms of segmentation accuracy, it significantly outperforms them in terms of computation time. The fast convergence and high breakdown point of GLKS make this estimator an outstanding choice for real-time estimation and segmentation of multiple motions.

**Keywords:** robust estimation, motion segmentation, fundamental matrix, homography

## 1. Introduction

Solving parametric motion segmentation problems involves robust estimation in a 4D parameter space for homography and a 7D or 8D parameter space for fundamental matrix models. In the presence of multiple motions, a small ratio of data points belong to each single motion. Therefore, *high breakdown* robust estimators need to be employed. Since the introduction of RANSAC [1], several high breakdown robust estimators have been specially designed to solve computer vision problems. Some recent examples include MSSE [2], ASSC [3], MLESAC [4], pbM-estimator [5], HBM [6] and FLKS [7]. All such estimators include three main steps: optimization, segmentation and refinement. The *optimization* step involves searching the parameter space to find an initial estimate that optimizes the objective function of the estimator. In the *segmentation* step, an inlier-outlier dichotomy is extracted using the parameter estimate returned by optimization. The parameter estimate is then *refined* in the final step by fitting the model to the extracted inliers using a maximum likelihood estimator *e.g.* total least-squares.

Most high breakdown estimators use *random sampling* for the optimization search. Random sampling is a random search scheme in the *sample space* for an acceptable  $p$ -

tuple (also called elemental subset or random sample) that results in a parameter estimate with an optimal (or close to optimal) objective function value. A  $p$ -tuple is a subset of  $p$  data samples ( $p$  denotes the dimension of parameter space) that defines a full rank system of equations from which a *model candidate* can be computed. If the number of  $p$ -tuples randomly selected is  $N$ , then with a probability of:

$$P_{\text{success}} = 1 - [1 - (1 - \epsilon')^p]^N \quad (1)$$

at least one of them is a *good*  $p$ -tuple (i.e. all its samples belong to the inlier structure), where  $\epsilon'$  is the ratio of outliers. Thus, for a given success probability  $P_{\text{success}}$ , at least:

$$N = \left\lceil \frac{\log(1 - P_{\text{success}})}{\log(1 - (1 - \epsilon')^p)} \right\rceil \quad (2)$$

$p$ -tuples should be randomly examined where  $\lceil \cdot \rceil$  means rounding up to the next integer. When the ratio of inliers to each motion is small, the number of random  $p$ -tuples given by (2) is substantial and the computational load of segmentation is too high for real-time (or near real-time) applications [6, 8, 9]. For instance, in case of having eight homography motions with almost equal number of matching points and 10% mismatches, at least 28,748 random 4-tuples are required to ensure 99% success probability. The computational cost is higher in case of general motions modeled by eight fundamental matrices, requiring at least 20,191,937 random 7-tuples.

In this paper, we introduce a new a guided sampling approach that makes use of the spatial proximity of the points belonging to each motion. This method can be used to optimize the cost function of many estimators, and it requires far less random  $p$ -tuples (compared to random sampling) in case of parametric segmentation of multiple motions. Using our optimization search to minimize the  $k$ -th order statistics of squared distances, we develop a guided estimator called Guided Least  $k$ -th order Statistics in this paper (GLKS for short). Our experiments show that in terms of estimation error, GLKS performs similar to state of the art high breakdown robust estimators, but it runs substantially faster when segmentation of several motions is involved.

## 2. Related Work

Tordoff and Murray [10] have tackled the computational cost issues of random sampling by devising a guided sam-

pling technique in which the  $p$ -tuples are randomly chosen using a Monte-Carlo method according to the matching probabilities. In another effort, the “preemptive RANSAC” suggested by Nister [11] preemptively prioritizes a large number of random  $p$ -tuples and depending on the available time, the cost function is computed and compared for a portion of them. However, in the examples presented by Tordoff and Murray [10] and Nister [11] to benchmark the performance of their methods, the number of  $p$ -tuples required to obtain accurate fits is still close to the theoretical number given by Eq. (2).

An alternative approach to random sampling is an iterative search such as the gradient-based search over a Grassman manifold in the  $p$ -dimensional space of parameter vectors implemented within pbM-estimator [5], and an iterative search based on reweighted least-squares used in High Breakdown M-estimator (HBM) [6, 9]. Bab-Hadiashar and Hoseinnezhad [7] also suggested an iterative search routine in their Fast Least  $k$ -th order Statistics (FLKS) estimator to accelerate random sampling. In cases involving multiple motions, the iterative search within both pbM and HBM and FLKS need to be repeated for a number of times each starting from a random  $p$ -tuple. The number of repetitions needs to be tuned offline by trial-and-error.

A number of modified versions of RANSAC have also appeared in the computer vision literature. Some recent examples include LO-RANSAC [12], randomized RANSAC [13] and BEEM [14]. These methods follow RANSAC in that they assume the availability of a scale (or error threshold) to determine the inliers for each hypothetical fit. Such a threshold is not always available. For instance, in motion estimation, the scale of noise mainly depends on the point matching techniques utilized to find corresponding points. In this paper, we are interested only in estimators that work without any knowledge of the scale of measurement noise. Indeed, in many such estimators, the scale is either estimated within the estimation procedure or can be separately calculated upon its completion.

Our guided sampling method requires a small number of random samples – far less than the theoretical number given by Eq. (2). The search process does not need the scale of noise or any other threshold to determine inliers. Indeed, a theoretical stopping criterion is formulated and used to determine the minimum required number of  $p$ -tuples for a given success probability.

### 3. Guided sampling for motion segmentation

With sophisticated feature matching techniques such as SIFT [15], the ratio of mismatches in parametric motion segmentation is usually small. In robust statistical terms, the ratio of *gross* outliers are small, but *pseudo* outliers (points belonging to structures other than the target) constitute a large portion of data if several moving objects exist. The main reason for the large number of random samples

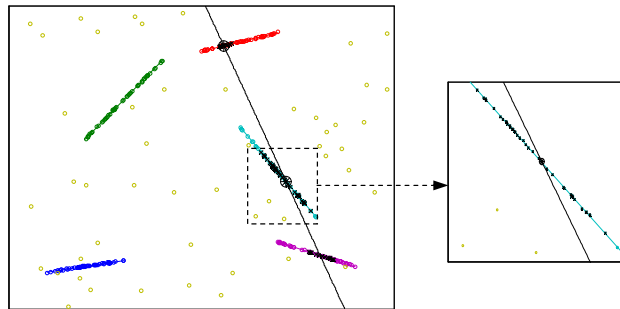


Figure 1: A 2D representation of a fit given by data points belonging to different structures: The fit passes through several structures.

is that in Eq. (2) many  $p$ -tuples include points from different structures and are simply discarded because they are not *good* samples. We believe that there are information in those  $p$ -tuples that can be utilized to effectively direct the search toward the optimum point of the cost function.

The fit given by a so-called bad sample usually passes *through* the structures which contribute at least one point to the  $p$ -tuple. Fig. 1 shows a 2D case. Using the fit given by the two points marked by  $\otimes$ , an inlier-outlier dichotomy specifies a number of points as inliers to the fit (the black points in Fig. 1). Those inliers belong to different structures, however, in motion segmentation, the points belonging to each structure are spatially close to each other and can be separately clustered from the others. This statement can be violated if there are objects occluding each other, but the number of simultaneously occluding objects is always less than the total number of motions.

It is noteworthy that should the scale of noise is very large, the points grouped in the largest cluster may be closely spaced and the effect of noise may be amplified which may lead to poor results of resampling from the cluster. As it is shown in Fig. 1 (see the magnified section in the right side of the figure) and through our simulation results involving homography estimation, resampling from the largest cluster usually results in an acceptable  $p$ -tuple.

Our suggested guided sampling is shown as a schematic diagram in Fig. 2. It comprises two levels of nested random sampling. First,  $n_1$  random  $p$ -tuples are chosen. Each  $p$ -tuple corresponds to a parameter estimate for which an inlier-outlier dichotomy is computed. The inliers are then clustered into separate groups of data points and the largest group is found. If the group contains sufficient data points, then in an inner level of random sampling,  $n_2$  random  $p$ -tuples are selected from the points in the group. The total number of random samples will be no more than  $N = n_1 n_2$ . For a given  $n_1$  and  $n_2$ , the probability of having at least one “good”  $p$ -tuple is computed as follows.

Each of the  $n_1$  random  $p$ -tuples in the first round of sampling contains  $p$  correct matches (though belonging to different motions) with a probability of  $(1 - \epsilon)^p$  where  $\epsilon$  is the

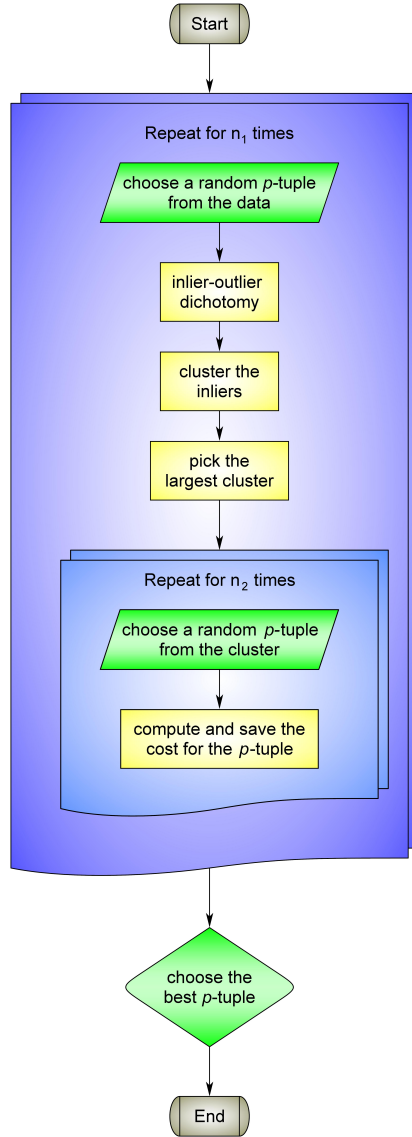


Figure 2: A schematic diagram for the Guided LKS algorithm.

ratio of mismatches.<sup>1</sup> Such a  $p$ -tuple corresponds to a fit and some inliers. The inliers are clustered into groups and the largest group is selected for a second round of random sampling. If at most  $n_{occ}$  objects can be simultaneously occluding each other, and in the worst scenario they contain an equal number of matching points in the data (the smallest possible ratio of inliers), each of the  $n_2$  random samples selected in the second round will be a good sample (with all points belonging to one motion) with a probability of  $(1/n_{occ})^p$ . Therefore, the probability of having at least one

<sup>1</sup>It is important to note that  $\epsilon$  is the ratio of *gross* outliers and in the presence of several motions, it is far less than  $\epsilon'$  in Eqs. (1) and (2) which equals the sum of gross and pseudo outlier ratios. In the presence of  $n_{obj}$  motions with an equal number of matching points,  $\epsilon' = \epsilon + (1 - \epsilon)(1 - 1/n_{obj})$ . For instance, if  $\epsilon = 5\%$  of data are mismatches and there are  $n_{obj} = 8$  motions, we have  $\epsilon' = 88.125\%$ .

good sample in the second round is given by:

$$P_2 = 1 - \left[ 1 - \left( \frac{1}{n_{occ}} \right)^p \right]^{n_2}. \quad (3)$$

The overall success probability is then given by:

$$P_{\text{success}} = 1 - [1 - (1 - \epsilon)^p P_2]^{n_1} \\ = 1 - \left[ 1 - (1 - \epsilon)^p \left\{ 1 - \left[ 1 - \left( \frac{1}{n_{occ}} \right)^p \right]^{n_2} \right\} \right]^{n_1} \quad (4)$$

and for a given  $n_2$  and success probability, the minimum number of random samples required to be examined throughout the search is:

$$N = n_2 n_1 \\ = n_2 \left\lceil \frac{\log(1 - P_{\text{success}})}{\log(1 - (1 - \epsilon)^p \{ 1 - [1 - (\frac{1}{n_{occ}})^p]^{n_2} \})} \right\rceil. \quad (5)$$

In order to demonstrate the reduction of number of samples compared to the traditional random sampling, we have plotted the number of samples given by (5) versus the ratio of mismatches, compared with the number of random samples required by RANSAC-based methods given by (2) for several number of motions. Figures 3(a) and 3(b) show the plots for homography ( $p = 4$ ) and fundamental matrix ( $p = 7$ ) estimation. In both cases, it is assumed that no more than two moving objects can occlude each other ( $n_{occ} = 2$ ). Noting the logarithmic scale on vertical axes, the plots show drastic reductions in the number of random samples, especially in cases where numerous motions are involved.

**Remark 1.** The above results are independent of the clustering technique used in the method, and any of the modern clustering methods can be employed. In our simulations we have used the Mean Shift method [16]. The mean shift algorithm is a nonparametric clustering technique which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters.

**Remark 2.** As it was mentioned before, only a *sufficiently large* group of points are subjected to a second round of sampling and if no such group results from clustering, the optimization procedure simply skips to process the next  $p$ -tuple of the first round of random sampling. In practice, a minimum size for the structure is assumed available (denoted by  $k_{\min}$ ) and a group is considered sufficiently large if it contains at least  $k_{\min}$  data points.

**Remark 3.** In deriving the formulas (4) and (5), we have assumed that for any of the  $n_1$  random samples that contains no mismatches, clustering of the inliers to the fit returned by the  $p$ -tuple will result in at least one sufficiently large group. It is noted that for some random samples that include *marginal* points on the edge of structures, this condition may not be satisfied, but we believe that such

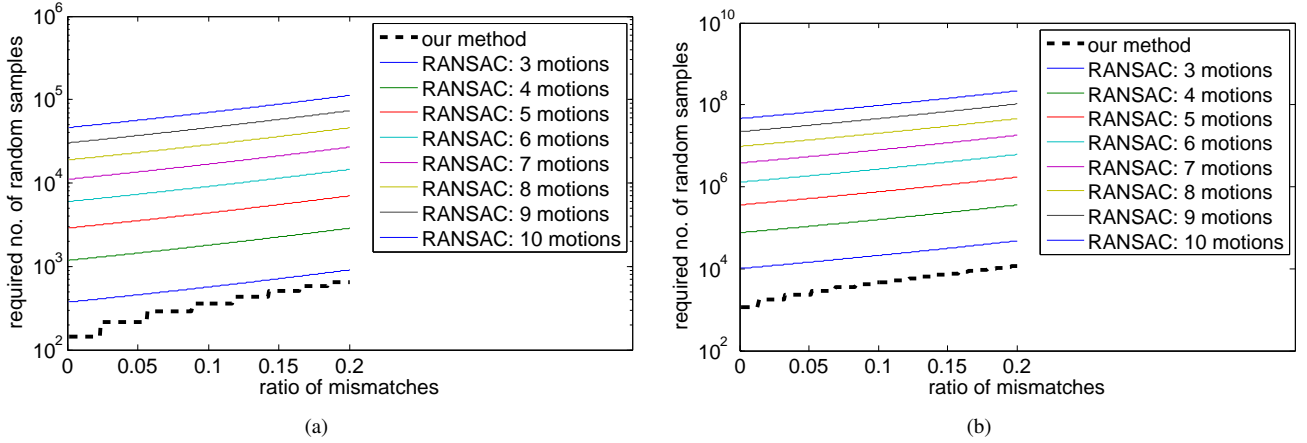


Figure 3: Comparison of the number of random samples required by our guided search compared to theoretical number of random samples as used by a RANSAC-type estimator for cases involving estimation of various numbers of (a) homography motions (b) fundamental matrix motions.

$$d'_i = \frac{|h_{11}x_i + h_{12}y_i + h_{13} - h_{31}x'_i - h_{32}y'_i - h_{33}|}{\sqrt{(h_{11} - h_{31}x'_i)^2 + (h_{12} - h_{32}y'_i)^2 + (h_{33} + h_{31}x_i + h_{32}y_i)^2}} \quad (7)$$

$$d''_i = \frac{|h_{21}x_i + h_{22}y_i + h_{23} - h_{31}x'_i - h_{32}y'_i - h_{33}|}{\sqrt{(h_{21} - h_{31}x'_i)^2 + (h_{22} - h_{32}y'_i)^2 + (h_{33} + h_{31}x_i + h_{32}y_i)^2}}. \quad (8)$$

marginal points constitute a small fraction of points. Indeed, the inaccuracy can be compensated for by treating such points like mismatches via increasing the ratio  $\epsilon$  in Eq. (5). It is important to note that in practice, the actual ratio of mismatches is small (using a modern matching technique such as SIFT [15]) and despite the above mentioned addition, a relatively small  $\epsilon$  will be practically sound in most applications.

#### 4. Simulation Results

To realize a high breakdown robust estimation scheme, we have used our guiding sampling to optimize the  $k$ -th order statistics of the squared distances (also called “squared residuals” in robust statistics literature). We call the complete estimator as Guided Least  $k$ -th order Statistics or GLKS for short. In a number of experiments, we have compared the performance of GLKS with MSSE [2], Adaptive Scale Sample Consensus (ASSC) estimator [3] and High-Breakdown M-estimator (HBM) [6]. For a given homography matrix  $H = [h_{ij}]$ , the following two constraints relate each pair of corresponding points  $\{(x_i, y_i), (x'_i, y'_i)\}$ :

$$\begin{aligned} h_{11}x_i + h_{12}y_i + h_{13} - h_{31}x'_i - h_{32}y'_i - h_{33} &= 0 \\ h_{21}x_i + h_{22}y_i + h_{23} - h_{31}x'_i - h_{32}y'_i - h_{33} &= 0. \end{aligned} \quad (6)$$

If the measurements  $[x_i \ y_i \ x'_i \ y'_i]^\top$  are corrupted with noise, first order approximations to the perpendicular distances of the point  $[x_i \ y_i \ x'_i \ y'_i]^\top$  from the two 6D manifolds

defined by Eq. (6) – similar to the Sampson distance [17] – are given by Eqs. (7) and (8) and we use the distance:

$$d_i = \sqrt{(d_i'^2 + d_i''^2)/2} \quad (9)$$

for segmentation. When the measurement noise is i.i.d. and distributed according to  $\mathcal{N}(0, \sigma^2)$ , the squared distances  $\{d_i^2\}$  are distributed according to  $\sigma^2\chi_2^2(\cdot)$  and 99.4% of the distances are smaller than  $4\sigma$ .

To obtain an inlier-outlier dichotomy for a given parameter estimate (the segmentation step within GLKS) we have employed the specific routine suggested by Bab-Hadiashar and Suter within the Modified Selective Scale Estimator (MSSE) [2]. The procedure is described as follows. Having a set of distances  $d_1, \dots, d_n$ , the distances are sorted in ascending order and the following unbiased scale estimates are computed:

$$\hat{\sigma}_k = \left[ \sum_{i=1}^k d_{i:n}^2 / (k - p) \right]^{1/2}; \quad k = k_{\min}, \dots, n \quad (10)$$

where  $k_{\min}$  is a known lower bound for the number of matches in a single motion and  $d_{i:n}$  denotes the  $i$ -th sorted distance. Then the smallest index  $k'$  is found for which  $d_{(k'+1):n} > 4\hat{\sigma}_{k'}$ . Finally, an inlier-outlier dichotomy is obtained by labeling all the points associated with distances  $\{d_{1:n}, \dots, d_{k':n}\}$  as inliers and the rest as outliers. This segmentation method has been chosen because it has been shown to have the best performance in terms of consistency [18] and finite sample bias [19].

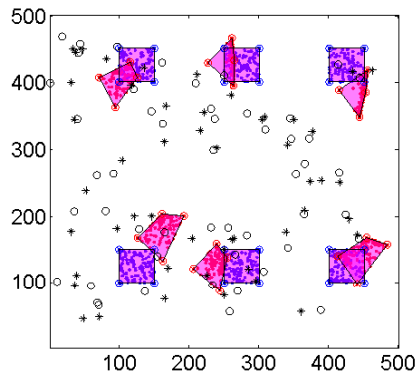


Figure 4: A sample of synthetic data generated in the experiment. The blue points in the squares move to red points. The black points marked with '\*' and 'o' symbols represent the mismatches in images 1 and 2, respectively.

Our experiments involve segmentation of 4 to 8 homography motions. Each experiment includes 100 runs, each involving randomly generated homography motions for 4 to 8 objects which appear as squares in image 1 and evolve to a generally irregular quadrilateral in image 2. For each motion, 100 correct matches are found and the pixel coordinates are synthetically corrupted with noise. Those points as well as a total of 50 random mismatches are taken as measurements. A sample of the measurements generated in an experiment involving 6 motions is shown in Fig. 4.

In each experiment, the measurements are utilized by a robust motion segmentation method and two quantities are recorded: the processing time taken by the method for segmentation of the first motion, and the number of matches computed for that motion. The number of samples for each method and number of motions are averaged over the 100 runs, and the results shown in Fig. 5. It is observed that in terms of segmentation accuracy, the performance of GLKS is similar to other methods examined in the experiments.

Figure 6 shows the average computation times for each method plotted versus the number of motions involved. It is observed that the computation time of GLKS (the number of required random samples) does not vary with the number of motions.<sup>2</sup> The computation time of other estimators substantially increase with the number of motions because the outlier ratio  $\epsilon'$  in Eq. (2) increases. Quantitatively, the computation time of GLKS is observed to be 10 to 1000 times faster than other methods. The computation advantage gained via using GLKS is especially substantial when numerous motions exist.

## 5. Conclusions

A guided sampling method is introduced that is substantially faster than random sampling. In its two-level nested

<sup>2</sup>Of course, similar to other methods, there is an application-dependent limit on the number of motions that can be segmented by GLKS.

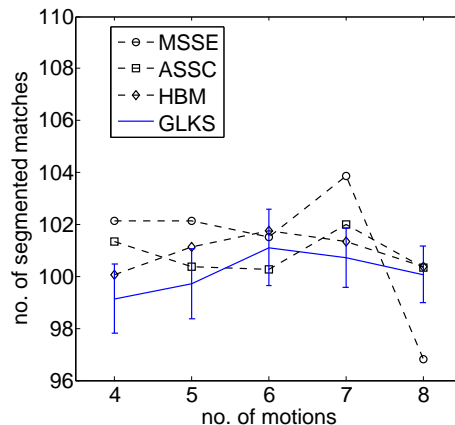


Figure 5: Average number of segmented inliers to the first estimated motion returned by the examined techniques. For GLKS, the error bars are also shown.

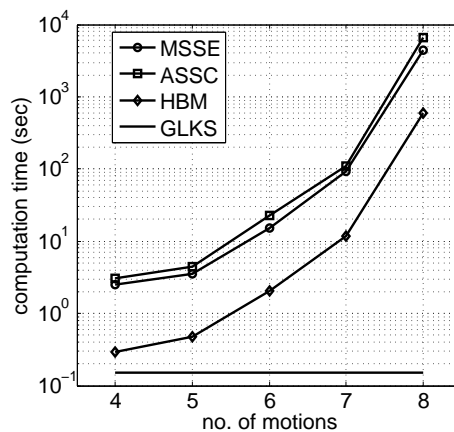


Figure 6: Average computation times required to segment the first motion.

search scheme, our method exploits the spatial proximity of the points belonging to each motion. A new high-breakdown robust estimator, called GLKS, uses our guided search to minimize the  $k$ -th order statistics of squared distances. Our homography estimation experiments involving synthetic data which involve up to eight motions show that while GLKS performs similar to other estimators in terms of segmentation accuracy, it is substantially faster. The advantage of using GLKS is more significant when numerous motions are to be estimated. The fast convergence and high breakdown point of GLKS make this estimator an outstanding choice for real-time estimation and segmentation of multiple motions.

## Acknowledgement

This research was supported by Australian Research Council through the Discovery grant number DP0880553.

## References

- [1] M.A. Fischler and R.C. Bolles: *Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography*, Comm. Assoc. Comp. Mach., 24(6):381–395, 1981.
- [2] A. Bab-Hadiashar and D. Suter: *Robust segmentation of visual data using ranked unbiased scale estimator*, ROBOT-ICA, 17:649–660, 1999.
- [3] H. Wang and D. Suter: *Robust adaptive-scale parametric model estimation for computer vision*, IEEE Trans. PAMI, 26(11):1459–1474, 2004.
- [4] P.H.S. Torr and A. Zisserman: *MLESAC: A new robust estimator with application to estimating image geometry*, Computer Vision and Image Understanding, 78:138–156, 2000.
- [5] Raghav Subbarao and Peter Meer: *Subspace estimation using Projection based M-estimators over Grassman manifolds*, ECCV'06, pages 301–312, Graz, Austria, May 2006. Springer.
- [6] R. Hoseinnezhad and A. Bab-Hadiashar: *A novel high breakdown m-estimator for visual data segmentation*, ICCV, Rio de Janeiro, Brazil, October 2007. IEEE.
- [7] Alireza Bab-Hadiashar and Reza Hoseinnezhad: *Bridging parameter and data spaces for fast robust estimation in computer vision*, In Proceedings of Digital Image Computing: Techniques and Applications (DICTA'08), pages 1–8, Canberra, ACT, Australia, 2008.
- [8] Reyhaneh Hesami, Alireza Bab-Hadiashar, and Reza Hoseinnezhad: *A novel hierarchical technique for range segmentation of large building exteriors*, In George Bebis, editor, Third International Symposium on Visual Computing (ISVC07), pages 75–85, Lake Tahoe, Nevada, USA, 26–28 November 2007. Springer-Verlag, LNCS 4842/2007.
- [9] Reza Hoseinnezhad and Alireza Bab-Hadiashar: *A novel high breakdown m-estimator for visual data segmentation*, In Digital Image Computing: Techniques and Applications (DICTA 2007), Adelaide, Australia, 3–5 December 2007. Australian Pattern Recognition Society, CSIRO Publishing.
- [10] B. Tordoff and D.W. Murray: *Guided sampling and consensus for motion estimation*, ECCV'02, pages 82–69, Copenhagen, Denmark, May 2002.
- [11] David Nister: *Preemptive ransac for live structure and motion estimation*, Machine Vision and Applications, 16(5):321–329, 2005.
- [12] Ondrej Chum, Jiri Matas, and Josef Kittler: *Locally optimized RANSAC*, In Bernd Michaelis and Gerald Krell, editors, Lecture Notes in Computer Science LNCS 2781, 25th DAGM Symposium, pages 236–243, Magdeburg, Germany, 2003. Springer.
- [13] O. Chum and J. Matas: *Optimal randomized RANSAC*, PAMI, 30(8):1472–1482, 2008.
- [14] L. Goshen and H. Shimshoni: *Balanced exploration and exploitation model search for efficient epipolar geometry estimation*, PAMI, 30(7):1230–1242, 2008.
- [15] David G. Lowe: *Distinctive image features from scale-invariant keypoints*, IJCV, 60(2):91–110, 2004.
- [16] Dorin Comaniciu and Peter Meer: *Mean shift: A robust approach toward feature space analysis*, PAMI, 24(5):603–619, 2002.
- [17] P.H.S. Torr and D.W. Murray: *Development and comparison of robust methods for estimating the fundamental matrix*, IJCV, 24(3):271–300, 1997.
- [18] Reza Hoseinnezhad and Alireza Bab-Hadiashar: *Consistency of robust estimators in multi-structural visual data segmentation*, Pattern Recognition, 40:3677–3690, May 2007.
- [19] Reza Hoseinnezhad, Alireza Bab-Hadiashar, and David Suter: *Finite sample bias of robust scale estimators in computer vision problems*, In George Bebis et al., editor, Lecture Notes on Computer Science (LNCS), No. 4291 (International Symposium on Visual Computing - ISVC'06), pages 445–454, Lake Tahoe, Nevada, USA, November 2006. Springer.