2013 IEEE International Conference on Robotics and Automation (ICRA)
Karlsruhe, Germany, May 6-10, 2013

ERODE: An Efficient and Robust Outlier Detector and its Application to Stereovisual Odometry

Francisco-Angel Moreno, José-Luis Blanco and Javier González-Jiménez

Abstract— **This paper presents ERODE, an efficient outlier detector with a quality similar to that of standard RANSAC but at a fraction of its computational cost. In contrast to RANSAC-based methods which follow a hypothesis-and-verify approach, ERODE employs instead the whole set of observations together with a robust kernel to perform robustified least-squares minimization. Our proposal has important practical applications among computer vision problems, which we demonstrate with stereovisual odometry experiments with both simulated and real data.**

I. INTRODUCTION

A common issue found in computer vision problems is the presence of noise and outliers that corrupt data and features gathered from images, representing a challenge to the accuracy and reliability of the intended applications. In general, however, the effects of random noise tend to cancel out when using least-squares estimation or other filtering techniques. In contrast, the presence of outliers (defined as *observations that appear to deviate markedly from other members of the sample in which it occurs* [1]) means a more serious problem and, in fact, may lead an application to completely inaccurate results. Hence the development of robust techniques for minimizing their impact has become essential.

In this sense, a wide range of robust estimators have been proposed in the technical literature, being the RANSAC method, originally published by Fischler and Bolles [2], probably the most widely employed in Computer Vision. RANSAC is a hypothesis-and-verify algorithm that iteratively generates a tentative solution from a randomly selected minimal subset of data and searches for consensus among the rest of the data, generating consensus sets (CS). The solution with highest support from all the hypotheses is taken as the final estimation of the model. Although being highly robust, RANSAC becomes computationally unfeasible when the percentage of outliers in the data increases significantly, since the number of hypotheses to be tested also grows, often preventing its usage in real time.

Hypothesis-and-verify methods are not the only solutions to properly cope with outliers. Methods that define robust distributions [3] to model the presence of errors in the data make also possible to estimate the camera motion from all the data, regardless of the presence of outliers. Furthermore, although they can achieve a good estimation of the camera motion without explicitly distinguishing between inliers and outliers, it is straightforward to use them to perform fast outlier detection and, after their removal, to refine the final estimation. To the best of our knowledge, this approach has not been explored and applied to visual odometry.

Following this approach, in this paper we propose an efficient and robust outlier detector (ERODE) aimed at rejecting outliers in the feature associations employed in stereovisual odometry. Unlike previous proposals, it neither follows a hypothesis-and-verify approach, nor relies on prior information about the points and achieves results comparable to those by RANSAC. Its fundamental advantage, in comparison to RANSAC, is a significant reduction in the processing time (about one order of magnitude), while still performing almost as good in rejecting outliers for a wide range of outliers ratios. The main weakness of our proposal is the need for a decent initial estimation for the unknown model, thus it cannot replace RANSAC in all applications but definitively is a much better alternative wherever such a gross estimate is available. In visual odometry, using the previous camera location as starting point for the new pose is good enough for our method, as will be demonstrated with experiments on both real and simulated datasets.

II. RELATED WORK

Several works have tried to outperform the accuracy or overcome the drawbacks of RANSAC in Computer Vision. Thus, Torr and Zisserman presented in [4] a pair of approaches called MSAC and MLESAC which introduce some modifications in the RANSAC method to reinforce its robustness, and applied them to estimate some multiple view relations between images related by rigid motions. Briefly, the former changes the way that RANSAC determines the quality of a certain CS from simply its cardinality (i.e. the number of inliers) to a measurement of how well the inliers fit the model. The latter goes one step further and uses the negative log-likelihood of the mixed distribution of the errors (including the outliers) as a score of the CS and simultaneously estimates the percentage of outliers present in the observation. Nevertheless, these approaches do not reduce the computational time of the RANSAC algorithm.

Focusing on the computational burden, a set of methods has been developed which adds a preemptive approach to RANSAC. Chum and Matas presented R-RANSAC [5], probably the first preemptive attempt to reduce RANSAC computational complexity, that decreases the amount of matches to test when evaluating a hypothesis by selecting a random subset of them. The same authors also presented PROSAC [6], that orders the putative matches according to their scores and selects the minimal consensus set from a subset of them to instantiate the hypotheses. With this modification, they report a decrease of up to two orders of magnitude in processing time.

978-1-4673-5642-8/13/$31.00 ©2013 IEEE

A similar approach that uses MLESAC as a basis, followed Tordoff and Murray when presented Guided-MLESAC [7], an improved version that reports a reduction of its computational time by an order of magnitude when applied to the estimation of camera motions. They employ the score of a match (in the paper, the normalized cross correlation of their grey-level patches) to derive its prior probability of being an inlier and to guide the selection of the features which form the initial minimal CS for a hypothesis. Hence, those matches with higher score are more probable to be selected to build an initial CS.

Finally, David Nister presented another preemptive RANSAC-based method [8] that estimates both structure and motion in multiple views in real time. This approach follows a *breadth-first* preemption scheme consisting of the generation of a set of hypothesis and their iterative scoring as introducing observations into the consensus set. In each iteration, only a decreasing number of the highest scored hypotheses are evaluated until the winner is found.

Apart from the series of works by Nister [9], [10] which relies on preemption, little importance has been given to the rest of schemes in vision-based motion estimation applications, even though they have shown to be sound improvements of RANSAC. Instead, standard RANSAC have been extensively employed [11], [12], [13], [14].

III. ROBUST STEREOVISUAL ODOMETRY

This section depicts the stereovisual odometry application and introduces the notation employed throughout the paper.

A. Stereovisual odometry

The term *stereovisual odometry* stands for the process of estimating the change in pose that a stereo camera undergoes between two consecutive time steps by exclusively employing the information from stereo images. Let $\mathbf{x_k} = \left[t_x^k, t_y^k, t_z^k, \alpha^k, \beta^k, \gamma^k\right]^T$ be a column vector representation of the change in pose between time steps k and $k+1$ so that t_x^k, t_y^k and t_z^k indicates the change in position (i.e. translation) of the cameras and α^k, β^k and γ^k stand for the Euler rotation angles (i.e. the orientation). In this work, $\mathbf{x_k}$ will sometimes be referred as the *model* and, for the sake of simplicity, the subscript regarding the time step will be dropped wherever it is unnecessary.

The general solution to the problem of estimating the model $\mathbf{x}$ with a stereo camera follows this procedure:

1) **Feature detection**. Use an interest point detector (Harris corner detector [15], SIFT [16], etc.) to extract features in a pair of images I_k grabbed with the stereo pair (the subscript stands for the k-th time step).
2) **Stereo matching**. Match the features in the left image with their correspondent in the right one, aided by the epipolar restriction and using some measurement of the similarity between the features (normalized cross-correlation, sum of absolutes differences, SIFT descriptors, etc.).
3) Repeat steps (1) and (2) for the pair of images I_{k+1}.
4) **Data association**. Search in images I_{k+1} for the correspondences of the matches in I_k within a certain window (of size $w \times w$) around the position of the features in I_k. An interest point will be considered correct only if it is found in all the four images. In this work we define an *individual observation* ($\bar{\mathbf{z}}_i$) as the set of the coordinates of a certain feature in I_{k+1}:

$$\bar{\mathbf{z}}_i = \left[u_l^{k+1}, v_l^{k+1}, u_r^{k+1}, v_r^{k+1}\right]$$

   with the subscripts l and r indicating the image (left or right, respectively) where the coordinates u, v (image column and row, respectively) correspond. The set of individual observations for a certain time step forms the *observation* ($\bar{\mathbf{z}} = \{\bar{\mathbf{z}}_i\}$).
5) Choose an initial estimation $\mathbf{x}^0$ of the full 6D motion of the stereo camera which, among with the observation, will constitute the input data to the process of computing the camera motion. Under the mild assumption of a small change in pose between frames, an all-zeros initial estimation $\left(\mathbf{x}^0 = [0, 0, 0, 0, 0, 0]\right)$ will be sufficient.
6) **Motion estimation**. Triangulate the matches in I_k into the 3D space and back-project them on I_{k+1} according to the current estimation of the motion. The function which relates the current model $\mathbf{x}$ to the coordinates of the features in I_{k+1} is known as the *prediction function* $\mathbf{p_z}(\mathbf{x})$ and such coordinates form the *prediction* ($\mathbf{p_z}(\mathbf{x}) = \mathbf{z} = \{\mathbf{z}_i\}$).
7) Iteratively improve the motion estimation $\mathbf{x}$ and repeat from step (5) until the mismatch between the predictions and observations converges to a minimum. Such mismatch is usually measured as the Mahalanobis distance between the predicted and the observed positions of the detected features in the images.

The iterative minimization of steps (5)–(6) can be achieved with a Maximum Likelihood Estimator (MLE) which, essentially, selects the model $\mathbf{x_k}^*$ for which the total probability of the observed data becomes the highest. If the data are only corrupted by zero-mean Gaussian noise, the MLE coincides with a nonlinear least-squares estimator which minimizes a cost function like:

$$F(\mathbf{x}) = \sum_i \frac{1}{2} \mathbf{\Delta z}_i^T \mathbf{W}_i \mathbf{\Delta z}_i \tag{1}$$

where $\mathbf{x}$ is the vector of model parameters, $\mathbf{\Delta z}_i = \mathbf{z}_i - \bar{\mathbf{z}}_i$ stands for the error between the prediction of the i-th feature and its observation, and $\mathbf{W}_i$ is an appropriate weighting matrix, proportional to the inverse covariance of the normally distributed noise. Assuming identical error distributions for all detected features in both x and y directions leads to the simplification $\mathbf{W}_i = \mathbf{I}, \forall i$.

B. Gauss-Newton nonlinear least-squares minimization

The aim of the minimization process is to find a suitable model $\mathbf{x}$ that produces the minimum combined error for all the observed features. In general, cost functions present

nonlinearities that prevent obtaining a closed-form solution, thus a truncated Taylor series expansion of the function is iteratively minimized instead, relinearizing the approximation around the current model estimation until convergence. At each iteration we look for a small change $\partial\mathbf{x}$ in the model for updating its estimation like $\mathbf{x}^{i+1} = \mathbf{x}^i + \partial\mathbf{x}$. It can be demonstrated [17] that the optimal step arises by solving:

$$\mathbf{H}\partial\mathbf{x} = -\mathbf{g} \tag{2}$$

Here, $\mathbf{H}$ stands for the Hessian matrix and $\mathbf{g}$ is the gradient vector of the cost function $F(\mathbf{x})$. However if we consider (1) to be the cost function and under the assumption that the prediction errors $\mathbf{\Delta z}_i$ are small, we find:

$$\mathbf{H} \approx \mathbf{J}^T\mathbf{W}\mathbf{J} \tag{3}$$

$$\mathbf{g} = \mathbf{J}^T\mathbf{W}\mathbf{\Delta z} \tag{4}$$

with $\mathbf{J} = \frac{d\mathbf{p_z}}{d\mathbf{x}}$ being the Jacobian of the prediction model and $\mathbf{\Delta z} = \{\mathbf{\Delta z}_i\}$, hence converting (2) into

$$\left(\mathbf{J}^T\mathbf{W}\mathbf{J}\right)\partial\mathbf{x} = -\mathbf{J}^T\mathbf{W}\mathbf{\Delta z} \tag{5}$$

which expresses the relation between the update $\partial\mathbf{x}$ of the estimated model in terms of the error $\mathbf{\Delta z}$ (often referred as the *residual*) and the prediction function $\mathbf{p_z}$.

C. Outlier rejection: the RANSAC estimator

Due to inaccuracies in the interest points detectors and mismatches in the data association process, the presence of outliers is inevitable and they should be removed from the input data in order to accurately estimate the camera motion. The most widely employed method to overcome this issue is RANSAC.

RANSAC performs by randomly selecting a minimal set of individual observations from the input data (three observations for stereovisual odometry) which leads the Gauss-Newton least-squares minimization process until an estimation of the model, called the hypothesis, is computed. Subsequently, the residuals of the rest of the data are computed subject to the hypothesized model and, according to a defined threshold, the data are classified in *inliers* or *outliers*, being the set of inliers referred as the consensus set (CS) and its cardinality as the support for the hypothesis.

This process is repeated a certain number of times and the hypothesis with the highest support is selected as the best estimation ($\mathbf{x_k}^*$) for the model.

The number of hypotheses (n_h) that RANSAC needs to explore in order to find an outlier-free consensus set with a certain level of confidence follows this expression:

$$n_h \geq \left\lceil \frac{log\left(1-q\right)}{log\left(1-\varepsilon^m\right)} \right\rceil \tag{6}$$

with ε being the estimated percentage of inliers in the data and q the probability of obtaining an outlier-free CS. For instance, with $m = 3$ as in stereovisual odometry, to achieve a 95% of probability of getting an outlier-free consensus set, the value of n_h spans from 2 to 2995 for an inlier percentage of 5% to 90%, respectively. However, these number of hypotheses have often been considered to be overoptimistic and, in practice, they must be increased to cope with degenerate configurations in the data. This issue renders the capability of RANSAC of dealing with data contaminated with a large ratio of outliers when used in real-time applications.

IV. OUR PROPOSAL: ERODE

In this work we propose a novel, efficient method for stereovisual odometry based on robust kernels to perform fast and reliable outlier rejection even under conditions of large ratio of outliers. The core of our approach is the usage of a robust radial distribution to model the errors in the data, including both the noise and the outliers.

Outliers produce errors in the data which do not follow a Gaussian distribution and, therefore, their presence unavoidably leads to unreliable results of the MLE due to the fact that the probability of finding gross errors under a Gaussian assumption is extremely low (the tails of a Gaussian rapidly diminish). Here is where a realistic model of the error distribution plays a crucial role for the MLE to be useful, since unmodeled outliers are sufficient to render standard least-squares estimators useless. Mixture of both Gaussians and uniform distributions, or more sophisticated ones as Cauchy or Huber distributions [18], are examples of models for both inliers and outliers (hence defined as *total distributions*). All of these distributions, among with the Gaussian itself, are also known as *radial distributions* and have negative log likelihood of the form:

$$F\left(\mathbf{x}\right) = \sum_i \frac{1}{2}\rho_i\left(\mathbf{\Delta z}_i^T\mathbf{W}_i\mathbf{\Delta z}_i\right) \tag{7}$$

where $\rho_i(s)$ can be any increasing function that fulfills $\rho_i(0) = 0$ and $\rho_i'(0) = 1$. Note that (1) is a particular case of this expression with $\rho_i(s) = s$ whilst more robust cost functions are sublinear in s, often tending to a constant value at ∞. It is important to highlight that the usage of robust error distributions would make unncessary to follow hypothesis-and-verify approaches as they are practically *inmune* to the presence of outliers. Nevertheless, although the influence of outliers in the cost function is almost negligible, in practice, better results are achieved if they are detected and removed.

In this work we consider to employ a pseudo-Huber distribution [19] to model the errors and to lead a robustified Gauss-Newton least-squares minimization process which will split the input data set in inliers and outliers. The negative log-likelihood of the probability density function of the pseudo-Huber distribution forms the cost function to be minimized

$$F_R(\mathbf{x}) = \sum_i \frac{1}{2}\left[2b^2\left(\sqrt{1+\left(\frac{s_i}{b^2}\right)}-1\right)\right] \tag{8}$$

with $s_i = \mathbf{\Delta z}_i^T W_i \mathbf{\Delta z}_i$ and b being a parameter which tunes the shape of the function.

Figure 1(a) shows the comparison between the cost functions of the standard least-squares approach and the pseudo-Huber version with $b = 2$. Note that, for the robustified

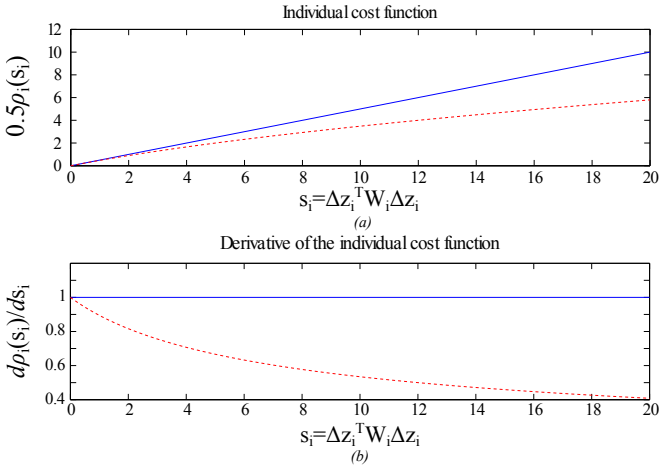


Fig. 1. (a) Cost functions and (b) the first derivative of ρ_i for a Gaussian (blue-solid) and a pseudo-Huber (red-dashed) distribution.

version, as the error increases (abscissa axis) the contribution of the observation to the cost function (ordinate axis) decreases compared to the standard least-squares case.

It is important to note that using a different cost function than (1) produces a modification in the above mentioned Gauss-Newton expressions. In concrete, we need to introduce the first derivative of ρ_i by weighting the gradient vector of the cost function with the vector $\bar{\rho}' = \{\rho_i'\}$ so that outliers contribute more slightly to it

$$\mathbf{g} = \bar{\rho}' \mathbf{J}^T \mathbf{W} \boldsymbol{\Delta}\mathbf{z} \tag{9}$$

thus leading to a modified version of equation (5):

$$\left(\mathbf{J}^T \mathbf{W} \mathbf{J}\right) \partial \mathbf{x} = -\bar{\rho}' \mathbf{J}^T \mathbf{W} \boldsymbol{\Delta}\mathbf{z} \tag{10}$$

where, for the pseudo-Huber distribution, we have

$$\rho_i' = \frac{d\rho_i}{ds} = \frac{1}{\sqrt{1 + \frac{s}{b^2}}} \tag{11}$$

The effects of this modification are shown in Figure 1(b) where it can be noted that the value of ρ_i' decreases as the residual grows, so that, when multiplied by the residual itself in (10), the contribution of large errors to the gradient vector is attenuated.

This approach implies that all the input data are considered in the minimization process but, on the other hand, there is no need to try different hypotheses of the model. With this method, the estimation process naturally converges towards the true solution and, after a few iterations, the outliers appear clearly visible in the vector of residuals so that we can remove them and, subsequently, refine the estimated solution to achieve higher accuracy.

A. Computational performance

Here we address the computational burden of both RANSAC and our proposed method when estimating the motion of the stereo camera. In this section, the superscripts ^{a} and ^{b} will be used to refer to the RANSAC method and ERODE, respectively.

Let M be the number of elements that form the input data for a certain time step, being m_{in} of them inliers. RANSAC explores n_h hypotheses and, for each one of them, it picks up m elements from the data and performs a complete minimization process which iteratively computes both the prediction and the Jacobian of the prediction function, spending p_p and p_J seconds per element, respectively. The number of iterations it performs until convergence will be denoted by n_i^a. Then, RANSAC evaluates the rest of the data ($M - m$ elements) against the estimated motion, spending just p_p seconds per element as the Jacobian is not computed here. After the n_h iterations, the largest CS is assumed to be the set of inliers and the computed solution considered the best possible. Finally, it is a common practice to subsequently start a new minimization process with only the set of inliers, and take the best solution so far as the initial estimation, in order to refine the final result. Let n_f^a be the number of iterations this refinement minimization would take. Thus, the computation time that RANSAC would spend to yield the final estimate of a particular motion of the camera would be

$$c^a = n_h \left[n_i^a m \left(p_p + p_J\right) + \left(M - m\right) p_p\right] + n_f^a m_{in} \left(p_p + p_J\right)$$

On the other hand, ERODE performs n_i^b iterations of the minimization process with the whole set of input data and, subsequently, the elements whose residual fall over a certain threshold are considered to be outliers. Finally, a refinement minimization process is started with only the inliers, reaching convergence in n_f iterations. Thus, the computational burden for this approach can be expressed as

$$c^b = n_i^b M \left(p_p + p_J\right) + n_f^b m_{in} \left(p_p + p_J\right)$$

In order to quantitatively evaluate the performance, let $p_p = p_J = 1$ be the computational time of each operation, $m = 3$, as stated for a stereovisual odometry application; let also the ratio of outliers (pct_o) to vary between 0.2 and 0.8 so that $m_{in} = (1 - pct_o)\, M$, and the rest of the parameters to take the realistic values $n_i^a = n_f^a = n_i^b = n_f^b = 4$ and $M = 300$. Figure 2 shows the theoretical computational time for both methods as the percentage of outliers grows.

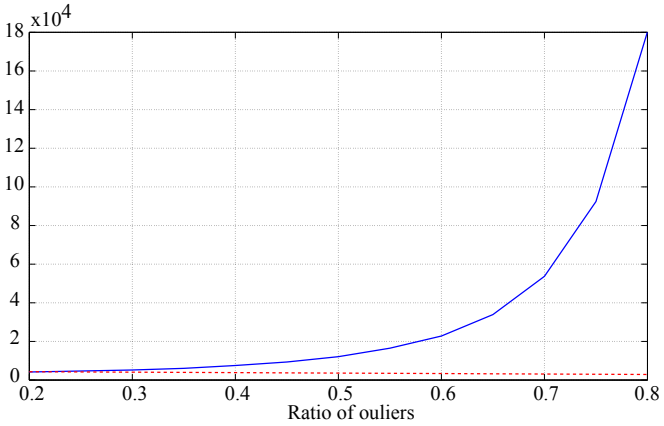


Fig. 2. Computational time for RANSAC (blue-solid) and ERODE (red-dashed) as the ratio of ouliers grows from 0.2 to 0.8.

V. EXPERIMENTS

This section presents two experiments which test the performance of our approach for computing stereovisual odometry. The first experiment simulates a typical scenario with a robot traversing an office-like environment while detecting interest points in stereo images while the second one uses the stereo video sequences published in the Karlsruhe outdoor dataset [20], gathered with a moving vehicle.

In these experiments we have performed a Gauss-Newton minimization process for both ERODE and RANSAC, regardless there exist other approaches (such as Horn's method [21]) that can be employed to estimate the change in pose of a stereo camera between two time steps. The usage of a maximum likelihood estimator here relies on the purpose of addressing the problem from a more generic approach that can be employed, for instance, for monocular cameras.

A. Simulated dataset

The usage of a simulated dataset to test our approach is motivated by both the availability of a ground truth and the capability of properly control the amount of outliers within the data.

This dataset have been created with the freely available Recursive World Toolkit[1] which provides a recursive language to define a 3D virtual world as a set of landmarks and to simulate a projective camera moving in it.

First, we set up a simple experiment where the camera just moved forward for about 15 cm, with no rotation. In order to cover different random sets of outliers, this experiment was repeated 100 times for each ratio of outliers. The error with respect to the ground truth was measured separately for the translation and the rotation where, in the latter, the Euler angles corresponding to the motion were converted to a 3-d *rotation vector* [22]:

$$[\alpha, \beta, \gamma] \rightarrow [w_1, w_2, w_3] \tag{12}$$

Thus, the two components of the error may be computed by the Euclidean distance between the real value and estimated one.

The average error (with 2σ confidence bars) for the translation and the rotation with respect to the outlier ratio for this experiment are shown in Figure 3. Note the similarity in the accuracy between RANSAC and our proposal.

Finally, a more complete experiment involving the full motion of the camera was also carried out. In this case, the stereo camera traversed a 50×90 meters virtual environment following a path of about 300-meters long. As the camera moved, the 3D landmarks were projected to the images at each time step, and their image coordinates (corrupted by zero-mean Gaussian noise with $\sigma = 0.5$ pixels) stored in a text file.

The presence of outliers was simulated by adding uniformly distributed noise in the interval $[-w/2, w/2]$ (with w being the size of the search window mentioned in III-A) to a randomly selected subset of points. The amount of outliers

[1] http://code.google.com/p/recursive-world-toolkit/

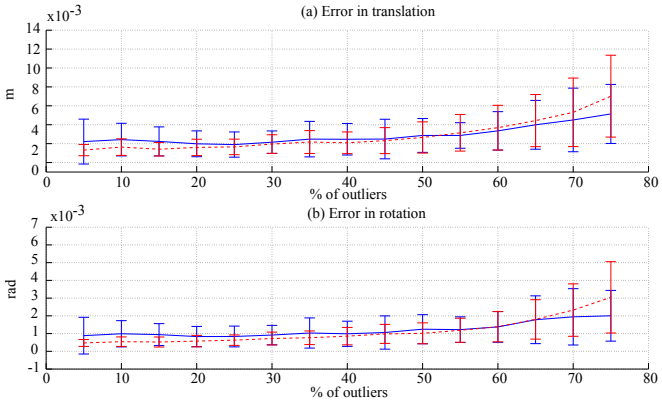


Fig. 3. Comparison of the errors in (a) translation and (b) rotation for standard RANSAC (blue-solid) and ERODE (red-dashed) with a forward movement with an outlier ratio ranging from 0.05 to 0.75. 2σ confidence bars are also displayed.

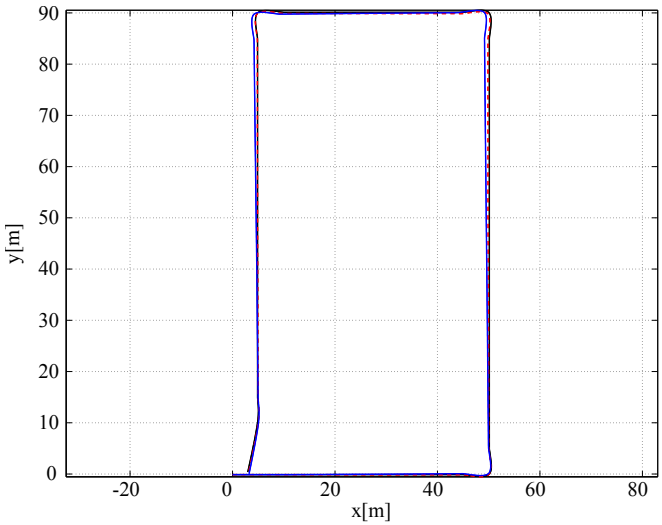


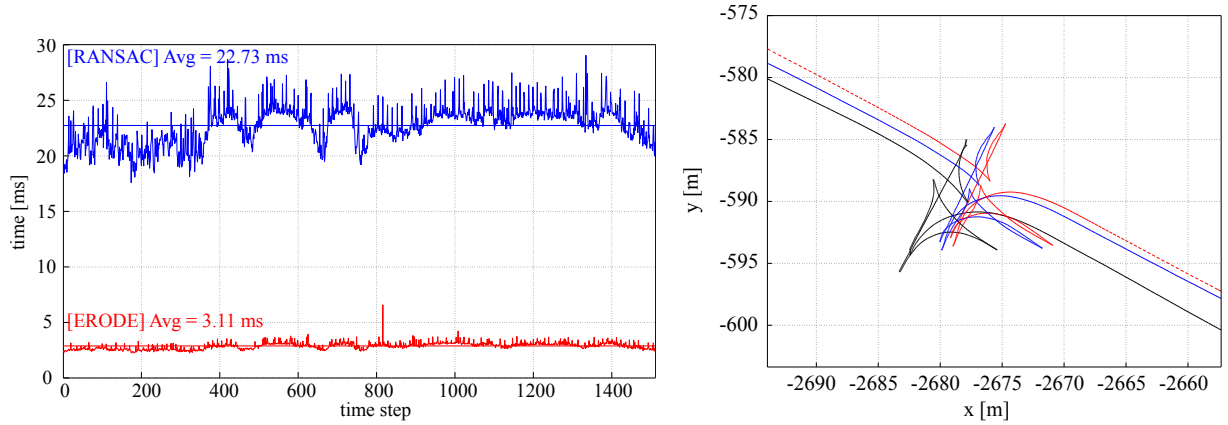
Fig. 4. Estimated paths with 50% of outliers for RANSAC (blue-solid) and ERODE (red-dashed). The ground truth is plotted in black.

at each time step was set according to a pre-defined ratio, which ranged from 0.05 to 0.75.

Figure 4 shows the estimated paths for RANSAC (blue-solid), ERODE (red-dashed) and the ground truth (black-dotted) for the case of 50% of outliers, respectively, which was taken as representative. As can be seen, ERODE yields similar results to RANSAC even when the ratio of outliers is high.

In order to evaluate the computational performance for this simulated dataset, we have measured the time spent by both RANSAC and ERODE methods as the ratio of outliers grows from 0.2 to 0.8, yielding Figure 6. Note as it fits the theoretical plot shown in Figure 2.

Finally, another indicative of the goodness of an outlier rejection method is the Receiver Operating Characteristic (ROC) curve (see Figure 7 for ERODE's ROC curve) and, in particular, the area under the curve (AUC). The ROC curve plots the sensitivity (ratio between the true positives and the total positives) against one minus the specificity (the fraction



(a) (b)

Fig. 5. (a) Computation time for each time step during the experiment when using RANSAC (blue-solid) and ERODE (red-dashed). The overall average times per time step are also shown as solid horizontal lines. (b) Camera paths as estimated with RANSAC (blue) and ERODE (red). The ground truth is plotted in black.

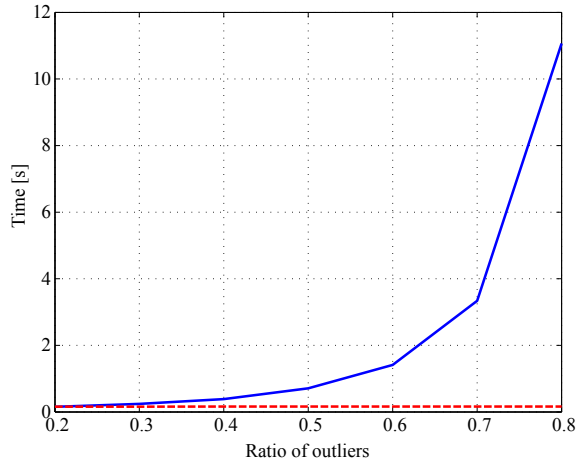


Fig. 6. Computational time for RANSAC (blue-solid) and ERODE (red-dashed) as the ratio of ouliers grows from 0.2 to 0.8 with the simulated dataset.

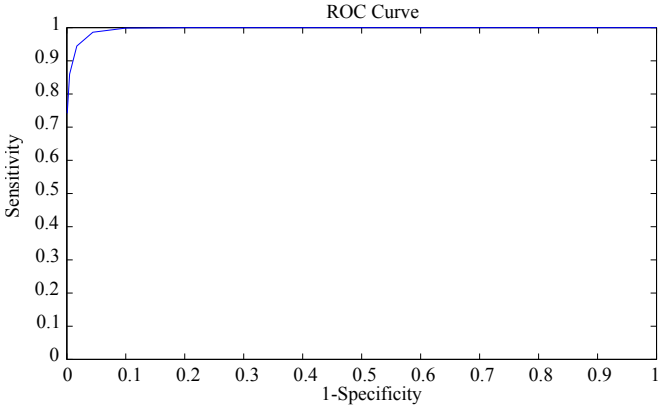


Fig. 7. ROC curve for ERODE in an experiment with 50% of outliers

of false positives out of the total negatives). In general, a method whose AUC value is over 0.96 is considered to have high discriminatory ability and, in our experiments, ERODE achieved an AUC value of 0.9957, getting considerably close to the ideal RANSAC's AUC value of 1.

B. Outdoor dataset

For a more realistic experiment, we implemented our algorithm inside the LibViso2[2] vision library by Andreas Geiger. This library offers solutions to the problem of estimating visual odometry for both monocular and stereo cameras, and also provides a collection of datasets to test its performance. The selected dataset for this experiment was taken with a pair of cameras mounted on a car following an about 270 meters trajectory within a city. The images had a size of 1344 $\times$ 391 pixels (see Figure 8 for an example) while the stereo pair presented a baseline of 0.572 meters.

Fig. 8. Example of the images in the outdoor dataset.

Since, in this experiment, the ratio of outliers within the data is not known beforehand, a conservative number of hypotheses for the RANSAC approach must be set in order to minimize the probability of not finding a outlier-free consensus set. In the original code the value for this parameter was set to ensure with 99% of probability that a proper CS is selected even in the presence of about 65% of outliers.

In order to compare the results and the time burden of our approach, we have measured the time that the original RANSAC-based code spent in the estimation of the

[2] http://www.cvlibs.net/software/libviso2.html

camera motion between two consecutive time steps and, subsequently, after replacing such code by our approach as described in section IV, we measured it again, yielding the results presented in Figure 5(a).

The estimated paths for both approaches are compared in Figure 5(b), showing their similar performance in terms of accuracy, while a comparative video illustrating the development of the experiment can be accessed on-line[3]. Please, note that ERODE considerably reduces the effect of outliers to the final result but such effect cannot completely be cancelled since they are still present in the minization process, hence the small deviation that ERODE incurrs with respect to RANSAC.

VI. CONCLUSIONS

RANSAC-based robust estimators and other hypothesis-and-verify approaches have been extensively employed in computer vision applications, and, in particular, in visual odometry solutions. Although proven to be highly robust, they suffer from a very high computational burden when the outlier ratio is significant. Here we propose a new approach based on robust functions which avoids the sampling nature of the above-mentioned methods an instead, employs all the input data to detect the outliers, and efficiently perform stereovisual odometry.

The results show similar accuracy than the RANSAC approach but reducing its computational cost in about one order of magnitude. Experiments in both simulated and real datasets support this contribution.

REFERENCES

[1] F. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, pp. 1–21, 1969.

[2] M. Fischler and R. 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.

[3] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustmenta modern synthesis," *Vision algorithms: theory and practice*, pp. 153–177, 2000.

[4] P. Torr and A. Zisserman, "Mlesac: A new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.

[5] O. Chum and J. Matas, "Randomized ransac with td, d test," in *Proc. British Machine Vision Conference*, vol. 2, 2002, pp. 448–457.

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

[7] B. Tordoff and D. Murray, "Guided-mlesac: Faster image transform estimation by using matching priors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1523–1535, 2005.

[8] D. Nistér, "Preemptive ransac for live structure and motion estimation," *Machine Vision and Applications*, vol. 16, no. 5, pp. 321–329, 2005.

[9] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1. IEEE, 2004, pp. I–652.

[10] ——, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol. 23, no. 1, pp. 3–20, 2006.

[11] K. Konolige, M. Agrawal, and J. Sola, "Large-scale visual odometry for rough terrain," *Robotics Research*, pp. 201–212, 2011.

[12] K. Ni and F. Dellaert, "Stereo tracking and three-point/one-point algorithms-a robust approach in visual odometry," in *Image Processing, 2006 IEEE International Conference on*. IEEE, 2006, pp. 2777–2780.

[13] M. Agrawal and K. Konolige, "Real-time localization in outdoor environments using stereo vision and inexpensive gps," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3. IEEE, 2006, pp. 1063–1068.

[14] K. Konolige, M. Agrawal, R. Bolles, C. Cowan, M. Fischler, and B. Gerkey, "Outdoor mapping and navigation using stereo vision," in *Experimental Robotics*. Springer, 2008, pp. 179–190.

[15] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of Alvey Vision Conference*, vol. 15, 1988.

[16] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[17] J.-A. Fernández-Madrigal and J.-L. Blanco, *Simultaneous Localization and Mapping for Mobile Robots: Introduction and Methods*. IGI Global, sep 2012, vol. (in press). [Online]. Available: http://www.igi-global.com/book/simultaneous-localization-mapping-mobile-robots/66380

[18] P. Huber, E. Ronchetti, and MyiLibrary, *Robust statistics*. Wiley Online Library, 1981, vol. 1.

[19] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[20] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving?" in *Computer Vision and Pattern Recognition (CVPR*, Providence, USA, June 2012.

[21] B. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A*, vol. 4, no. 4, pp. 629–642, 1987.

[22] J. Diebel, "Representing attitude: Euler angles, unit quaternions, and rotation vectors," *Matrix*, 2006.

[3] http://youtu.be/pcUhuM3pPOU