Computer Vision and Image Understanding **78,** 1–7 (2000)
doi:10.1006/cviu.1999.0833, available online at http://www.idealibrary.com on IDEAL®

INTRODUCTION

Robust Computer Vision: An Interdisciplinary Challenge

Peter Meer, Guest Editor

Electrical and Computer Engineering Department, Rutgers University, 94 Brett Road, Piscataway, New Jersey 08854-8058
E-mail: meer@caip.rutgers.edu

Charles V. Stewart, Guest Editor

Computer Science Department, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180-3590
E-mail: stewart@cs.rpi.edu

and

David E. Tyler, Guest Editor

Statistics Department, Rutgers University, 110 Frelinghuysen Road, Piscataway, New Jersey 08854-8018
E-mail: dtyler@caip.rutgers.edu

This special issue is dedicated to examining the use of techniques from robust statistics in solving computer vision problems. It represents a milestone of recent progress within a subarea of our field that is nearly as old as the field itself, but has seen rapid growth over the past decade. Our Introduction considers the meaning of robustness in computer vision, summarizes the papers, and outlines the relationship between techniques in computer vision and statistics as a means of highlighting future directions. It complements the available reviews on this topic [12, 13].

DEFINING ROBUST ESTIMATION IN COMPUTER VISION

The ultimate goal of computer vision is to make possible systems that can autonomously interpret the visual environment under almost any operating conditions, that is, to reproduce the amazing performance of human visual perception. The elusiveness of this goal can be seen from attempts to define robustness in the context of computer-based image understanding. Different definitions emerge at different levels of the hierarchy of techniques often associated with solutions to computer vision problems.

1077-3142/00 $35.00

At the top level, a robust vision system should be able to recognize new objects based exclusively on previous examples from the same class of functionality. Thus, a piece of furniture should be identified independent of its style, a car independent of its maker, etc. Such cognitive components, however, are yet to be developed for general purpose vision systems.

Changes in the visual environment due to nonrigid motion of the objects, alteration of viewpoint, or both are a primary concern in building object descriptions at the intermediate level of the vision hierarchy. Complex visual events arise for which robust interpretation requires separating the external causes from the intrinsic properties in the appearance of each object. This task, roughly equivalent to perceptual constancies in human visual perception, is currently an active research area in computer vision. Several papers in this special issue address this topic [1, 3, 6, 8].

Robust statistical methods were first adopted in computer vision to improve the performance of feature extraction algorithms at the bottom level of the vision hierarchy. These methods tolerate (in various degrees) the presence of data points that do not obey the assumed model. Such points are typically called "outliers." The definition of robustness in this context often is focused on the notion of the *breakdown point*: the smallest fraction of outliers in a data set that can cause an estimator to produce arbitrarily bad results. The breakdown point, as defined in statistics, is a worst-case measure. A zero breakdown point only means that there exists one potential configuration for which the estimator will fail.

The concept of breakdown also implicitly assumes that the absolute majority of the data, i.e., more than half of the data points, obeys the model whose parameters are sought. This assumption clearly is not met in many vision problems. For example, multiple instances of a class of features—e.g., lines or independently moving objects—are often present in the data. To delineate these features, a single global process will not suffice. Visual data is usually more complex than the data typically analyzed in statistics, and so often a straightforward application of robust statistical techniques does not work.

Some of the failings of robust statistical methods are addressed, at least in part, by robust techniques developed in the computer vision literature. Arguably, the most successful general vision algorithm is the Hough transform, a 38-year-old patent to detect the trace of charged subatomic particles in a bubble chamber [9]. The success of the Hough transform is due to the ability to generate reliable hypotheses about all the significant instances of the class of features sought. From this viewpoint the Hough transform is more "robust" in practice than estimators imported from statistics. Various aspects of the Hough transform are still actively being investigated, and two papers in this issue [4, 7] address such problems.

The RANSAC method [10], although developed 20 years ago in the vision community, has only recently been widely recognized. This recognition is partially due to its noted similarity to the high breakdown point least median of squares (LMedS) estimator [11] introduced into statistics a few years later and adopted by the vision community. RANSAC now often replaces LMedS in vision algorithms since, having two tuning parameters (the cardinality of the minimally acceptable set of data points and the amount of noise allowed to corrupt the model), it can be better adapted to complex data analysis situations. Both RANSAC and LMedS suffer from the same sensitivity problems, and improving the numerical behavior of RANSAC is addressed in [8].

PAPERS IN THE SPECIAL ISSUE

The eight papers selected for the *Robust Statistical Techniques in Image Understanding* special issue of *Computer Vision and Image Understanding* offer a representative sample of

the state of the art in robust computer vision. Several common trends can be recognized in these papers. The concept of robust analysis is extended to semantically more meaningful problems (higher levels of the visual hierarchy) in [1, 6] and to more realistic operating conditions in [5]. A second trend is to combine robust estimators with other techniques to obtain algorithms with increased autonomy and better performance, overcoming some of the shortcomings outlined above. The combinations include the maximum likelihood paradigm with RANSAC [8], the minimum description length principle with high breakdown point estimation [6], and fusion of information from different domains to create, in effect, a generalized M-estimator [5].

Robust behavior is often achieved in vision applications in spite of employing estimators with theoretically zero breakdown points. The M-estimators of regression are the best example; their zero breakdown point captures the possibility of failure in the presence of a severe leverage point. Attributing a nonzero breakdown point to M-estimators of regression is a misconception that shows up occasionally in vision papers, partly due to an incorrect statement in [12]. Only the more elaborate generalized M-estimators of regression (GM-estimators) and the multivariate M-estimators have a positive breakdown point. The satisfactory performance of vision algorithms containing M-estimator-based computational modules only underlines the worst-case nature of the breakdown point concept. Since image data is almost always bounded, the conditions for severe leveraging may not be present. This subject is addressed in [2].

Another issue concerning the zero breakdown point as it applies to the redescending M-estimators that are commonly used in vision algorithms should be also emphasized. Breakdown for redescending M-estimators, which can have multiple solutions, means that at least one of the solutions is bad and not that all of the solutions fail. So it is still possible that a redescending M-estimator produces a good solution even though theoretically it breaks down.

Here are some of the highlights of the papers in the special issue.

- The paper by Black *et al.* [1] illustrates the role robust estimation can play in developing techniques to explain complex changes in appearance. The authors employ mixture models to nonlinearly combine global parametric motion, global illumination changes, local specular highlights, and iconic motion. To recover the model parameters a robust expectation maximization (EM) algorithm based on t-distributions instead of normal distributions is proposed. The technique is validated with a variety of image data sets and combinations of models.
- The paper by Ben-Ezra *et al.* [2] shows that often global dominant motion models may be estimated without the use of complex, expensive search algorithms. In spite of having a zero breakdown point, M-estimators in general, and the minimization of an L1 error norm in particular, may suffice. The latter has the advantage of not requiring a reliable prior scale estimate. The authors use linear programming to efficiently minimize the L1 error norm and show that it can yield a performance comparable with the high breakdown point least median of squares. Experiments are based on image registration, transformation estimation, and mosaicing.
- The paper by Demirdjian and Horaud [3] investigates robust estimation of projective motion and detection of moving objects from a moving, uncalibrated stereo rig. To minimize a geometrically meaningful criterion, the high breakdown point RANSAC technique is combined in an iterative procedure with a reweighted least-squares method. Independent motion is detected by clustering tracked feature point correspondences that are inconsistent

with the dominant projective motion (outliers). Robust estimation of the epipolar geometry is used in feature tracking.

• The paper by Kiryati and Bruckstein [4] applies the Hough transform technique to estimating the best fitting line in heteroscedastic data, i.e., when the noise (error) distribution varies at each individual point. A robust voting kernel is derived from the Mahalanobis distance between a point and the line. A coarse-to-fine grid search is used to reduce the amount of computation. The resolution of the line detection procedure can be tuned by the voting kernel. The technique is illustrated with synthetic and real data.

• The paper by Lai [5] addresses the problem of image matching under varying illumination, using intensity constraints, an affine motion model, and low-order polynomial functions for additive and multiplicative intensity changes. An M-estimator problem formulation is enhanced with a prior weighting function that favors corresponding locations with similar intensity gradients. The method employs a data sampling technique to increase computational efficiency and a hierarchical nearest-neighbor search to initialize the M-estimator-based alignment estimation. Results are demonstrated in the context of printed circuit board inspection.

• The paper by Leonardis and Bischof [6] extends robustness to eigenimage recognition techniques by introducing a random sampling search that is traditional only in regression problems. Starting from random subsets of points in the eigenspace, robust representations are found by iteratively refining the subsets using a least-squares technique at each step. The minimum description length (MDL) principle is then employed to select the best hypothesis for the corrupted input image. The experimental results show tolerance of occlusions, cluttered backgrounds, and salt-and-pepper noise.

• The paper by Matas *et al.* [7] presents an improvement of the randomized Hough transform. The voting procedure is stopped once the probability that a bin's count could be due to a chance distribution drops below a threshold. The points associated with the delineated line are then removed from the accumulator and the procedure is restarted with the remaining data. A thorough set of experiments demonstrates the performance of the algorithm.

• The paper by Torr and Zisserman [8] proposes a novel paradigm for high breakdown estimation in computer vision by replacing the inlier count of RANSAC methods with weighted voting based on an M-estimator. The advantages of such an approach are discussed in detail and motivated in a rigorous maximum likelihood estimation framework. The chosen application domain is the estimation of the geometry of two uncalibrated cameras as described by models of increasing complexity; homography, fundamental matrix, and quadratic transformation. The issue of different parametrizations of these models is also discussed, and the theory is supported by a large number of experiments with real data.

FUTURE DIRECTIONS OF RESEARCH

An important question arises from the above discussion and from review of the papers in this issue. The most popular robust vision techniques (the Hough transform and RANSAC) were developed by the vision community. Similarly, the robust estimators imported from statistics are often only simple computational modules in complex vision algorithms. Furthermore, there is a significant difference in the nature of the data processed in statistics and computer vision. In light of this, can the robust paradigm as developed in statistics be helpful for progress in image understanding? That is, should vision researchers attempt

to build a robust estimation toolbox exclusively from the standpoint of vision tasks and without regard to work in statistics?

Whether one answers this question positively or negatively, it is still important to correctly formulate statistical problems within vision in order to have a basic understanding of the employed methods and thus to apply the proper technique. In support of this statement the two concepts that are particularly important to statistical problems in vision should be discussed: the definition of the residual and the knowledge of scale.

All of the robust methods discussed above, including the Hough transform, RANSAC, M-estimators, and LMedS, are residual-based. The first step in applying any of these methods is to have the proper definition of the residuals, i.e., the deviations from an estimated structure. Unlike classical statistics, in computer vision applications the residual cannot be viewed as simply due to a stochastic error term within a well-defined model; instead, it should be defined based on its geometric meaning. The definition affects the invariance properties of the parameter estimates under transformations of the input and therefore the optimality of the employed estimation process.

A classic example is the relationship between the Hough transform in computer vision and linear regression in statistics. In computer vision, the Euclidean distance between a point and a candidate line is generally more important, whereas in linear regression it is the vertical listance, i.e., the residual in the direction of the "response." Thus, the proper analogy of the Hough transform to regression techniques in statistics is not to classical linear regression but to orthogonal regression. The differences between linear and orthogonal regression can be quite large, especially for lower signal-to-noise ratios, a typical situation in feature extraction.

Once the appropriate definition of a residual is established, the issue of scale has to be addressed. Methods from computer vision (Hough transform and RANSAC) treat scale as known or as a tuning constant, while general robust statistical methods such as LMedS and M-estimators treat scale as unknown and something to be estimated. Viewed in this way, the computer vision estimators are less general than those from statistics. In effect, vision methods sacrifice generality to be able to handle the complexities of the data.

The similarities and differences between the methods are revealed by considering all of them under the unified banner of M-estimators with auxiliary scale. Suppose there are n observations which are associated with n residuals r_i, $i = 1, \ldots, n$. To estimate the model underlying the measurements, the expression

$$\sum_{i=1}^{n} \rho\left(\frac{r_i}{s}\right) \tag{1}$$

has to be minimized. In (1) s is a measure of scale and $\rho(\cdot)$ is a function that is symmetric, nonnegative, and nondecreasing on $\mathcal{R}^+$. The scale may be known a priori or estimated from the data. In the latter case it may be computed off-line using a scale statistic, or it may be estimated simultaneously with the fit. Fits minimizing the criterion (1) are called M-estimates with auxiliary scale. What distinguishes the various M-estimators with auxiliary scale is the choice of the ρ-function and the choice of the scale statistics.

The robust vision methods, i.e., the Hough transform and RANSAC, use the same special case of the criterion (1), namely a scale s determined a priori and

$$\rho(x) = \begin{cases} 1 & \text{if } |x| \geq 1 \\ 0 & \text{if } |x| < 1. \end{cases} \tag{2}$$

In the Hough transform the fits correspond to local minima of (2). Several instances of the same model can be recovered by identifying these minima in the voting space. In RANSAC, the fit should correspond to the global minimum of (2), but multiple fits can be obtained by repeating the voting process. It is important to stress, however, that the objective function is the same for both the Hough transform and RANSAC. They only differ in the computational procedure. The clustering used in the Hough transform and the probabilistic resampling in RANSAC have no theoretical bearing on the estimation.

The least median of squares estimator can also be formulated as minimizing the criterion (1) with ρ defined as in (2), but with the scale term estimated simultaneously from the data as

$$s^2 = \min_i \operatorname{med} r_i^2, \tag{3}$$

where the minimum is taken over all possible fits. This formulation shows not only the similarities but also the differences between RANSAC and LMedS: RANSAC uses an a priori scale whereas LMedS estimates scale. In vision application, this fundamental distinction is often obscured by stressing the nearly identical nature of the resampling technique used by RANSAC and LMedS for minimization. By emphasizing a computational issue instead of the principle behind the estimation, the potential for developing more advanced methods better suited to image understanding applications can be missed.

One improvement in both RANSAC and the Hough transform is obtained by understanding that using the jump ρ-function (2) can lead to local instability in the fits. This issue has been recognized in both the computer vision and the statistics literature. In particular, both RANSAC and LMedS suggest a second reweighting step to increase stability, i.e., to improve the local robustness behavior. Further improvements are obtained by incorporating continuous ρ functions directly into the original objective functions. This has been done for the Hough transform, see for example [4], as well as for RANSAC, where the proposed MLESAC estimator [8] is in fact an M-estimator with a known scale and a nonstandard search technique. These modifications of the underlying objective functions resulted in significant performance improvement.

The problem of scale is difficult to solve in computer vision. Prior knowledge of scale is often not available, and scale estimates are a function of both noise and modeling error, which are hard to discriminate. On the other hand, outright outlier rejection is very important in vision applications. This argues for the use of M-estimators that have a redescending influence function which in turn need accurate scale estimates. The combination of a large amount of clutter, unknown scale, and the requirement of stable estimates provides a substantial challenge for estimation in computer vision. A large body of recent research in statistics, however, addresses related issues. The proposed methods include simultaneous M-estimators of scale, S-estimators, MM-estimators, and constrained M-estimators, e.g., [14]. These methods should be adapted to problems in computer vision since robust statistics has still a lot to offer.

We can conclude that vision research has reached a level of maturity where it is ready for a more formal systematic treatment of its statistical problems. Similarly, the problems raised by image understanding tasks are complex enough to challenge statisticians. A closer collaboration between the two communities can be only advantageous for everybody involved.

Finally, we thank the authors for putting up with several revisions and *Computer Vision and Image Understanding* for hosting this special issue.

REFERENCES

1. M. J. Black, D. J. Fleet, and Y. Yacoob, Robustly estimating changes in image appearance, *Comput. Vision Image Understand.* **78**, 2000, 8–31.
2. M. Ben-Ezra, S. Peleg, and M. Werman, Real-time motion analysis with linear programming, *Comput. Vision Image Understand.* **78**, 2000, 32–52.
3. D. Demirdjian and R. Horaud, Motion–egomotion discrimination and motion segmentation from image-pair streams, *Comput. Vision Image Understand.* **78**, 2000, 53–68.
4. N. Kiryati and A. M. Bruckstein, Heteroscedastic Hough transform (HtHT): An efficient method for robust line fitting in the 'errors in the variables' problem, *Comput. Vision Image Understand.* **78**, 2000, 69–83.
5. S.-H. Lai, Robust image matching under partial occlusion and spatially varying illumination change, *Comput. Vision Image Understand.* **78**, 2000, 84–98.
6. A. Leonardis and H. Bischof, Robust recognition using eigenimages, *Comput. Vision Image Understand.* **78**, 2000, 99–118.
7. J. Matas, C. Galambos, and J. Kittler, Robust detection of lines using progressive probabilistic Hough transform, *Comput. Vision Image Understand.* **78**, 2000, 119–137.
8. P. H. S. Torr and A. Zisserman, MLESAC: A new robust estimator with application to estimating image geometry, *Comput. Vision Image Understand.* **78**, 2000, 138–156.
9. P. V. C. Hough, *Method and Means for Recognizing Complex Patterns*, United States Patent 3069654, December 18, 1962.
10. R. C. Bolles and M. A. Fischler, A RANSAC-based approach to model fitting and its application to finding cylinders in range data, in *Proc. 6th International Joint. Conf. on Artificial Intelligence, Vancouver, Canada, August 1981*, pp. 637–643.
11. P. J. Rousseeuw, Least median of squares regression, *J. Amer. Statist. Assoc.* **79**, 1984, 871–880.
12. P. Meer, D. Mintz, D. Y. Kim, and A. Rosenfeld, Robust regression methods in computer vision: A review, *Internat. J. Comput. Vision* **6**, 1991, 59–70.
13. C. V. Stewart, Robust parameter estimation in computer vision, *SIAM Rev.* **41**, 1999, 513–537.
14. B. Mendes and D. Tyler, Constrained M-estimation for regression, in *Robust Statistics, Data Analysis, and Computer Intensive Methods: In Honor of Peter Huber's 60th Birthday* (H. Rieder, Ed.), Lecture Notes in Statistics, Vol. 109, pp. 299–320, Springer-Verlag, New York, 1996.