Nonparametric smoothing of interferometric height maps using "confidence" values

Jochen Restle[1], Michael Hissmann[1*], Fred A. Hamprecht[2]

[1] Robert Bosch GmbH, FV/PLF2, Postfach 300240, 70442 Stuttgart, Germany,

Tel.: +49(711)811-37402, Fax: +49(711)811-37245, e-mail: {Jochen.Restle, Michael.Hissmann}@de.bosch.com

[2] IWR, University of Heidelberg, 69120 Heidelberg, Germany,

Tel.: +49(6221)54 88 00, Fax: +49(6221)54 88 50, e-mail: Fred.Hamprecht@iwr.uni-heidelberg.de

[*] Corresponding author.

Nonparametric smoothing of interferometric height maps using "confidence" values

Abstract. We use an extension of normalized convolution to smooth height maps from interferometry using "confidence" values. The latter are often used for dichotomous good/bad decisions only, with all bad data being discarded. To minimize loss of information, we weight each pixel individually by the inverse of its expected variance. The relation between supplied confidence values and empirical variances is found by regression. The width of the smoothing kernel—as small as possible to prevent loss of spatial resolution, as large as necessary to average out noise—is adjusted locally so as to yield a smoothed image with a prespecified uncertainty that is homogeneous throughout. In our experimental investigations using metrological data from a white light interferometric sensor, the variable-width mask leads to images with somewhat lower absolute deviation from an average image than the fixed-width masks we compare to.

Keywords: confidence, normalized convolution, weighted regression, white light interferometry

Introduction

Many metrological techniques, such as white light interferometry [1, 2] or phase measuring triangulation [3], provide not only a height value for each point, but also a measure of the associated uncertainty. The latter is often called a "confidence value"; no unique definition exists and these values are not to be confounded with statistical confidence intervals. In white light interferometry, for instance, the intensity curve at each pixel is recorded while the object of investigation is scanned through the range in which interferences with the reference beam appear. The height of the object at a point is determined by locating the scan position provoking the largest interference. In general, the stronger the contrast of the interference, the better the reliability at a given point. As a consequence, the amplitude of the interference at its maximum can be taken as a confidence value.

In this paper, we discuss how a confidence map can provide guidance in identifying outliers in the height map and help averaging over these with the smallest possible loss of information. We emphasize cases in which the confidence map is strongly non-homogeneous, as in our interferometric experiments where the confidence can vary by an order of magnitude from pixel to pixel.

Established approaches

The probably best-known method to reduce noise in height maps and images is to smooth them with a Gaussian filter mask or a median filter, if more robustness is required. As the transfer function of smoothing masks drops to zero at high wave numbers, high frequencies are suppressed and spatial resolution is lost (see [4] for discussion). The additional knowledge provided by confidence values is not accounted for.

A straightforward approach making use of confidence information is to set a threshold and discard all height data with insufficient confidence values. The resultant height map contains holes that render most analyses (like extraction of position, inclination or shape parameters) difficult. The result of follow-up image processing may be very sensitive to the threshold, and often the choice of threshold is a tradeoff between discarding "good" data and retaining outliers, as in Fig. 1b which features a continuum of confidence values from pure to noise to precise measurements.

Much information is lost in these simple approaches—the former completely disregards the confidence data while the second makes use of a hard threshold that does not do the data justice.

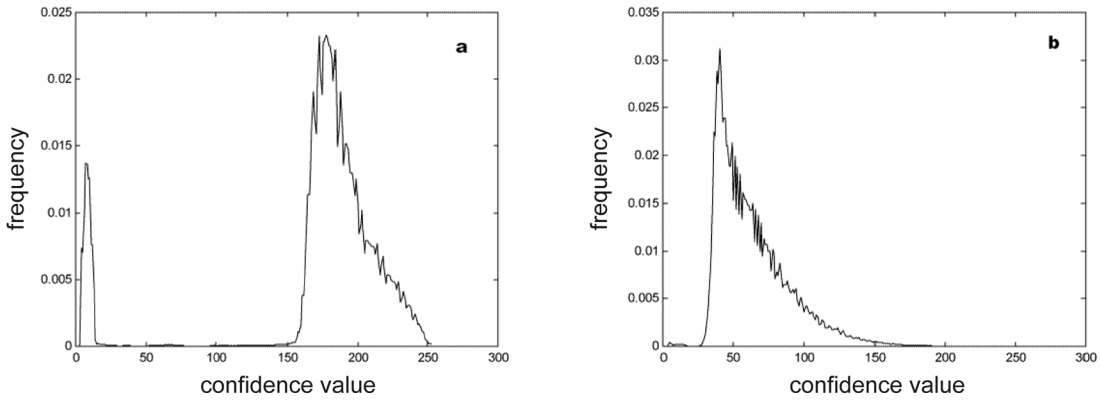


Fig. 1. Distribution of confidence values. **a** shows a benign case: a surface determined with high reliability that can be distinguished clearly from the noisy background. **b** features a typical measurement with low reliability and a continuum of confidence values pertaining to pure noise at one end of the spectrum and "good" data on the other.

Relation to adaptive smoothing methods

We wish to smooth only to the extent that is absolutely required. Since this extent varies locally with the reliability of our data, we need an adaptive system. All of the following are possible candidates provided that the intrinsic information these approaches extract from the image itself is replaced by the extrinsic information provided by the confidence map.

In [5], the size of a Gaussian mask was adapted using the local variance as intrinsic information. The same kind of information was extracted from interferometric data and used in a multiscale wavelet-based denoising and defect detection algorithm that kept type II errors constrained [6]. The edge-preserving properties of variable-width filters are investigated in the context of bilateral filtering [7, 8].

Heuristic measures based on local derivatives are used [9, 10, 11] to change weights in a fixed size filter mask. Filter coefficients are set according to an analysis of the local noise characteristics. Related works use rank-order filter masks [12, 13]. Here, MAD values (eq. 4) are used to tune the parameters and evaluate the performance of a weighted median filter.

In the field of regularization-based approaches, edges have long played an important role. Information on edges has been incorporated in the energy functional either explicitly as line processes [14, 15] or implicitly and generalized as in [16]. Similarly, the local noise variance has been investigated as a smoothing term in [17, 18].

In all of these adaptive approaches, confidence information could be used instead of measures based on local derivatives and local variance.

Smoothing images with variance-weighted variable-width kernels

We wish to denoise height maps derived from interferometric measurements. Instead of adapting a smoothing filter according to local height map features or discarding all data with confidence below a fixed threshold, we proceed as follows:

We weight each height value with its uncertainty that we derive from confidence values given by the imaging system. At the same time, the noise in the height data is accounted for by averaging over a local neighborhood, the size of which should depend on the uncertainty, as in kernel smoothing (see [19, 20]). These two aspects are detailed in the next sections.

Relating "confidence" values to variances

There are many definitions of "confidence" values, and the exact relation to standard deviation or variance may be difficult to derive, or even completely unknown. In our specific metrological system, the confidence value is given by the maximum difference between adjacent maxima and minima [21]. However, the commercial systems used in an industrial environment are mostly based on unpublished proprietary algorithms, and this is the problem we wish to address. Accordingly, we treat the sensor as a "black box", and find an empirical relation between confidence and standard deviation by performing repeated measurements on the same object, and regressing the empirical standard deviation on the found confidence values.

We recorded 25 height maps and confidence maps and took good care to avoid any displacement of the object between individual scans. For each pixel we calculated the standard deviation of the height values and plotted these against the mean of the confidence values. The empirical standard deviations are biased because we use a finite scan range leading to finite empirical deviations even though the proper deviations are infinite in pure noise areas.

The resulting scatter plot (Fig. 2) reveals two characteristic features: firstly, for high confidence values, the deviation of height values reaches a saturation, showing that there is a minimum residual error of the

measurement that cannot be overcome. We have made no distinction between relative errors (coming from the interferometric post-processing) and systematic errors (coming from a depth-shift of the object between measurements, caused by play in the scanning mechanics).

Secondly, the empirical standard deviation grows dramatically for small confidence values. Depending on the algorithm, the confidence may still take finite values even if there is no object in the scan range and the record consists of noise only. In this case, the empirical standard deviation has a negative bias due to the finite scan range.

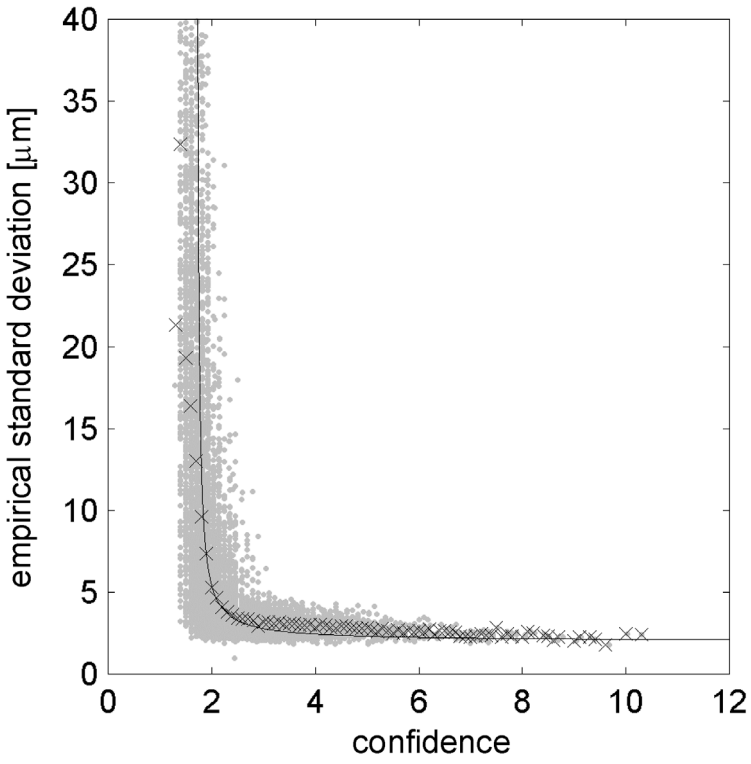


Fig. 2. The scatter plot shows the robust empirical standard deviation (eq. 4) of height values determined from 25 scans against the confidence values for all data with a mean gray value of the camera image of 80. The line shows a nonlinear regression.

Fig. 2 shows the robust empirical standard deviation (light gray dots) for each pixel as a function of confidence value with an average gray value of the camera image of 80. The black marks show the mean value of all observations with a 0.1 binning of the confidence values.

The plot suggests modeling the relationship with a hyperbolic function and parameters for pole location c_{pole}, and residual standard deviation, σ_{opt}.

We first estimated the position of the pole and the residual σ_{opt}; the pole depends on the intensity I at a pixel and we found $c_{pole} = 0.94 + 0.0096*I$. The asymptotic standard deviation σ_{opt} gives the residual empirical standard deviation that remains greater zero even for very large confidence values.

Given these estimates we approximate the observations using

$$\sigma(c,I) = \begin{cases} \frac{1}{c - c_{pole}(I)} + \sigma_{opt} & for\ c > c_{pole} \\ \infty & elsewhere \end{cases} \quad (1)$$

Smoothing with variable-width kernels

At this point, we have a height map with estimated variances for each point that can be used in a nonparametric regression of the object surface. In an image processing context this approach is known as "normalized convolution" [22]. We choose the symmetric Nadaraya-Watson kernel smoother [23, 24] and are confronted with the problem of a proper choice for the kernel bandwidth. Using a weighting kernel with square support of side length n (n odd), we obtain the smoothed height map $g'_{x,y}$ as

$$g'_{x,y} = \frac{\sum_{k_x=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{k_y=-\frac{n-1}{2}}^{\frac{n-1}{2}} g_{x+k_x,y+k_y} \frac{f^h_{k_x,k_y}}{\sigma^2(c_{x+k_x,y+k_y})}}{\sum_{k_x=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{k_y=-\frac{n-1}{2}}^{\frac{n-1}{2}} \frac{f^h_{k_x,k_y}}{\sigma^2(c_{x+k_x,y+k_y})}} \quad (2)$$

where $f^h_{k_x,k_y}$ are the coefficients of the smoothing mask with bandwidth h. Considering individual pixels as uncorrelated random variables (in our 25 scans, the deviations from the empirical mean are indeed not spatially correlated), the variance in the smoothed image becomes

$$\sigma'^2_{x,y} = \left(\sum_{k_x} \sum_{k_y} \frac{f^h_{k_x,k_y}}{\sigma^2(c_{x+k_x,y+k_y})} \right)^{-2} \sum_{k_x} \sum_{k_y} \left(\frac{f^h_{k_x,k_y}}{\sigma(c_{x+k_x,y+k_y})} \right)^2 \quad (3)$$

The bandwidth h of the kernel can now be increased iteratively, until the estimated variance $\sigma'^{2}_{x,y}$ reaches the target variance, for instance the asymptotic variance that is reached for very large confidence values, in our case $\sigma_{opt}^2 = 1.30^2\ \mu m^2$.

The result is a filter mask that is fine-tuned to the local uncertainty; if the expected noise is large, the height is averaged over a larger area; if the data seems reliable, the smoothing mask becomes narrow and blurring of the height map is avoided.

If, for instance, the data in the local neighborhood is extremely noisy with the exception of a single off-center pixel that has a large confidence and small predicted variance, the bandwidth h will grow until the "good" pixel receives a non-negligible weighting coefficient, at which point it will determine the height value at the central pixel because all pixels in the neighborhood but the "good" one will have a small weight due to their large variances.

If, on the other hand, the central pixel is "good", it will have a low variance from the start, the smoothing mask will remain tight and no blurring occurs.

If, finally, the local environment features constant confidence values, the width of the smoothing mask depends on the absolute confidence values: if these are good, there is no reason to smooth the data and provoke a decrease in spatial resolution. If the confidence values are bad, a large smoothing mask will be used because including more measurements lowers the variance of the predicted height. The cost is an increased bias which manifests itself as a loss of spatial resolution.

In our experiments, we have used a Gaussian kernel and a mask size n large enough such that the outer coefficients f are negligible.

Experimental results

We illustrate the method using the example of a black sheet of paper. The low luminance of the object led to a poor signal-to-noise ratio, resulting in a considerable number of pixels with small confidence (Fig. 1b). The object also features a fine-grained surface structure which lets us observe the decrease in spatial resolution. Fig. 3 shows the result when applying a 3x3 median, 3x3 Gaussian and a variable-width filter mask to the original data. The size of the variable-width mask was chosen to reach a target standard deviation of 1μm.

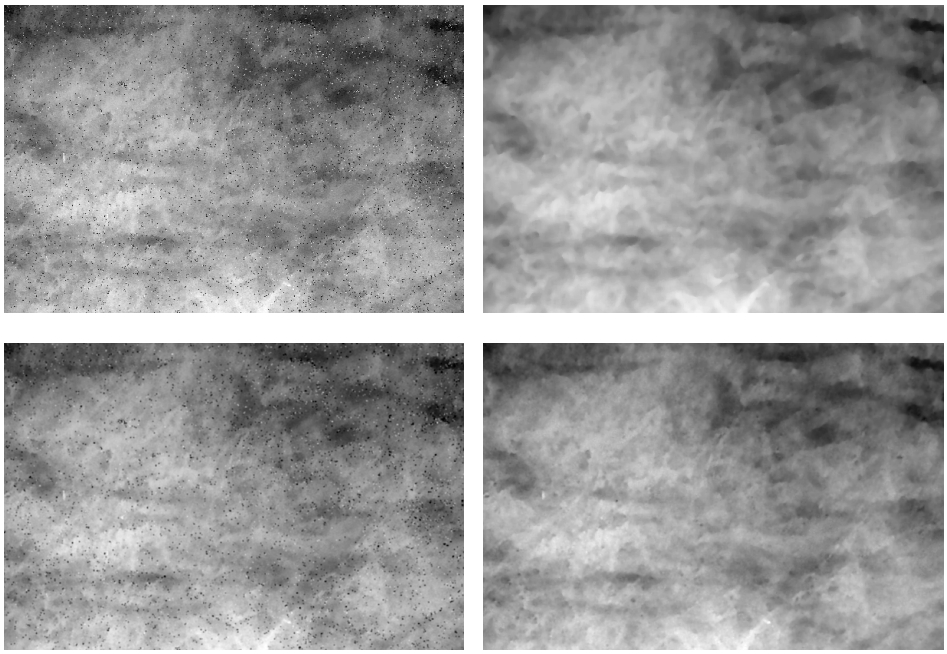


Fig. 3. Top left: the original height map as recorded by the sensor. Top right: original image smoothed with a 3x3 median filter. Bottom left: original image smoothed with a 3x3 Gaussian filter. Bottom right: the result of the described method.

A 3x3 Gaussian mask is clearly not large enough to eliminate outliers. The 3x3 median handles outliers much better, but visual inspection of the full-resolution (600x400) images seems to suggest that the median sacrifices more spatial resolution than the variable-width mask. The granular structures that persist in the output of the variable-width algorithm, on the other hand, seem to correspond to real features in the height map.

Still, all of these observations are subjective and a quantification thereof is non-trivial: we wish to retain fine structure and discard noise and their discrimination is far from obvious. In short, we do not have a "gold standard" to compare to. Our best approximation to the truth is a median image, where we take the median over 25 height values from the 25 scans at each individual pixel. We can then plot the absolute deviation of one of the processed images from the median image. To summarize the data, we would like to lump together the information for all those pixels with a common uncertainty. Our best approximation to the uncertainty of an individual pixel is a robust estimate of the spread of the 25 height values around their median. To measure this spread, we calculated the median absolute deviation *(MAD)* (cf. [25] for discussion):

$$MAD = 1.482\ med\ |\ g_i - g_{med}\ | \qquad (4)$$

at each individual pixel which gives, in words, the median of the absolute deviations of the 25 height maps from the median height map. The multiplicative constant assures that the same numerical value as for the standard deviation results if the data is normally distributed.

Fig. 4 shows the median of the absolute deviations of different processed images from the median image, broken down according to the estimated uncertainty of the reference height data. The left part illustrates that the loss of spatial resolution provoked by using larger median masks does manifest itself in the plot we propose. Especially in the leftmost part of the plot in which the empirical uncertainty as given by the MAD is very small (around 0.5μm), the greater absolute errors which result in a considerable loss of detail produced by the larger masks are visible.

The right part illustrates the absolute error obtained for different smoothing masks, again broken down according to the estimated uncertainty of the reference data. The 3x3 median clearly outperforms the 3x3 Gaussian for all pixels but the ones with smallest *MAD*. The variable-width mask, in turn, leads to somewhat lower errors than the median, but provides not as good an estimate as the unprocessed image in regions of very small *MAD*. We ascribe this to the fact that a large spread of empirical standard deviations is observed for a given low confidence value (Fig. 2). In particular, there are pixels with both a low confidence and empirical standard deviation. The former will lead to a smoothing that is not justified by the latter and will thus lead to an unnecessary loss of spatial information which provokes an increased absolute error. On the other hand, we found that in some cases, the empirical standard deviation describes the reliability of a measured value badly. If the intensity curve shows a parabolic behavior – this may occur at a defocused bright surface spot – and no interference phenomenon is detected, common algorithms erroneously put the height estimate at either the beginning or the end of the curve for all 25 measurements. This leads to a small empirical standard deviation although an infinite deviation would physically make sense.

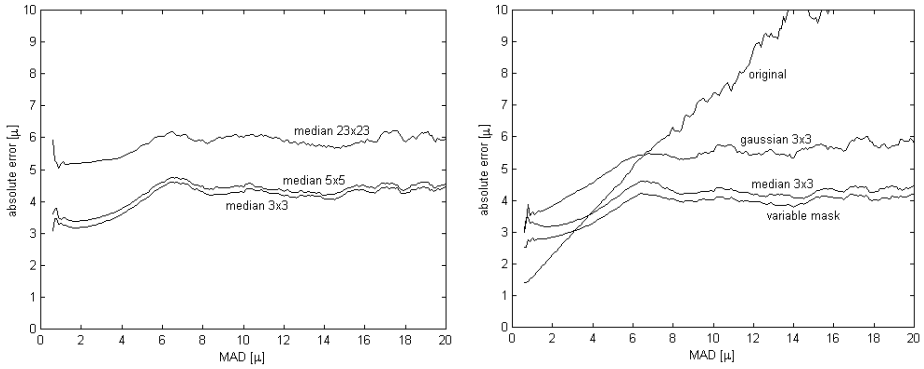
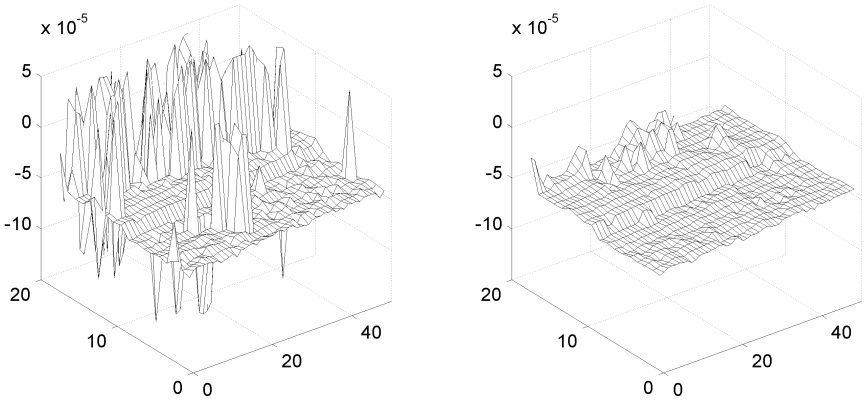


Fig. 4. The median of the absolute deviation of images smoothed with different methods from the reference image (see text), as a function of the estimated uncertainty of the reference image (see experimental results for details). Left: median absolute error of images that were smoothed using median masks of different sizes. Larger masks lead to greater loss of spatial resolution, resulting in larger absolute deviations. Right: median absolute error of an unsmoothed height map ("original") and of height maps smoothed with a 3x3 Gaussian, 3x3 median, and a variable-width Gaussian mask.



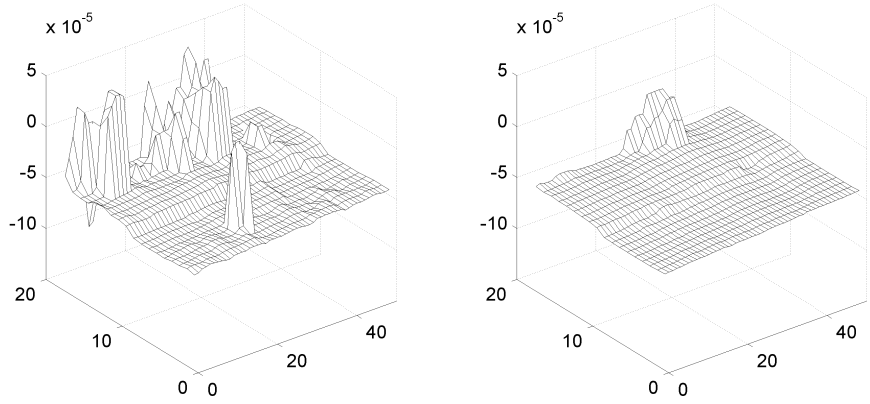


Fig. 5: Detail of a metallic surface with a 10μm step added. Top left: the original height displayed as a surface map. Top right: the result of the described method. Bottom left: original image smoothed with a 3x3 median filter. Bottom right: original image smoothed with a 9x9 median filter.

Fig. 5 shows the surface of a metallic object which has been modified by adding a small step to the height map. All deviations from the plane were judged to be outliers. A 3x3 median filter does not suppress all outliers and even a 9x9 median filter leaves one region of outliers. Adaptive smoothing leaves some outliers comparable with the 9x9 median, but preserves the height step better.

Summary and outlook

We apply an extension of normalized convolution using "confidence" values associated to height maps obtained from interferometry. The information relating to the local reliability of an image or height map is used in two ways. Firstly, by weighting the data appropriately: uncertain data get less and reliable data more weight. Proper weighting is done with inverse variances, and an empirical relation of these to the supplied confidence values is found by regression (Fig. 2). Secondly, the local variance of the smoothed image is predicted and the smoothing mask increased to the minimum size required to attain a desired target uncertainty (eq. 3), thus minimizing the loss of spatial resolution.

The current implementation takes of the order of a second on 600x400 images on a commodity PC and is thus about ten times slower than a 3x3 median filter. This is irrelevant for our application (where measurements are slow), but excludes real-time applications.

In its current form, the method relies on repeated measurements on a sample object to establish the relation between the supplied confidence values and the empirical uncertainty for a certain class of surfaces. The variable-width mask performs slightly better than the fixed-size filter masks we compared to, but it depends on the quality of the confidence values. We realized that the measuring system used provides a confidence value which is intuitive but not mathematically well-founded and which, without transformation, is a poor approximation to the variance. An alternative approach to generate confidence values from interferograms is sketched in [26].

References

1. T. Dresel, G. Häusler and H. Venzke, "Three-dimensional sensing of rough surfaces by coherence radar", *Appl. Opt.* 31(7), 919-925 (1992).
2. P. J. Caber, "Interferometric profiler for rough surfaces", *Appl. Opt.* 32(19), 3438-3440 (1993).
3. K. G. Seitz, H. J. Tiziani and R. Litschel, "3D-Koordinatenmessung durch optische Triangulation", *Z. Feinwerktechn. Messtechn.* 94 (1986).
4. B. Jähne, *Digital Image Processing. Concepts, Algorithms, and Scientific Applications*, 5. ed., Springer Verlag, Heidelberg, 2001.
5. G. Deng and L. W. Cahill: "An adaptive Gaussian filter for noise reduction and edge detection", Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference, vol 3, 1615-1619, 1993.
6. R.-J. Recknagel, R. Kowarschik and G. Notni: "High-resolution defect detection and noise reduction using wavelet methods for surface measurement", *J. Opt. A: Pure Appl. Opt.* 2, 538-545 (2000).
7. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images", Proc. IEEE International Conference on Computer Vision, 839-846, 1998.
8. D. Barash, "A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation", *IEEE Trans. PAMI* 24(6), 844-847 (2002).
9. Ph. Saint-Marc, J.-S. Chen and G. Medioni, "Adaptive smoothing: a general tool for early vision", IEEE Trans. PAMI 13(6), 514-529 (1991).
10. Ch.-T. Chen and L.-G. Chen, "A self-adjusting weighted median filter for removing impulse noise in images", Proc., IEEE International Conference on Image Processing, vol. 1, 419-422, 1996.
11. G. Ramponi, "The rational filter for image smoothing", *IEEE Sig. Proc. Lett.* 3(3), 63-65 (1996).

12. A. Restrepo and A. C. Bovik, “Adaptive trimmed mean filters for image restoration”, *IEEE Trans. ASSP* 36(8), 1326-1337 (1988).
13. T. Chen and H. R. Wu, “Adaptive impulse detection using center-weighted median filters”, *IEEE Sig. Proc. Lett.*, 8(1) 1-3 (2001).
14. S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images”, *IEEE Trans. PAMI*, 6(6) 721-741 (1984).
15. J. L. Marroquin, *Surface Reconstruction Preserving Discontinuities*, MIT LIDS Technical Report no. 1402, 1984.
16. S. Z. Li, “On discontinuity-adaptive smoothness priors in computer vision”, *IEEE Trans. PAMI,* 17(6) 576-586 (1995).
17. N. P. Galatsanos, “Methods for choosing the regularization parameter and estimating noise variance in image restoration and their relation”, *IEEE Trans. Image Proc.,* 1(3), 322-336 (1992).
18. S. C. Park and M. G. Kang, “Noise-adaptive edge-preserving image restoration algorithm“, *Opt. Eng.,* 39(12), 3124-3137 (2000).
19. T. Hastie, R. Tibshirani and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, New York, 2001.
20. K. S. Riedel and A. Sidorenko, “Function estimation using data adaptive kernel smoothers – How much smoothing?”, *Computers in Physics* 8, 402-409 (1994).
21. G. Häusler, University of Erlangen, personal communication.
22. G. H. Granlund and H. Knutsson, *Signal processing for computer vision,* Kluwer Academic Publishers, Dordrecht, 1995.
23. E. A. Nadaraya, “On estimation regression”, *Theory Probab. Appl.* 10, 186-190 (1964).
24. G. S. Watson, “Smooth regression analysis”, *Sankhyā,* 26, 101-116 (1964).
25. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York, 1986.
26. J. Restle, *Optimierung der Weißlichtinterferometrie für Applikationen der industriellen Qualitätskontrolle,* Shaker, Aachen, 2003.