A nonparametric minimum entropy image deblurring algorithm

Cesario Vincenzo Angelino, Eric Debreuve, Michel Barlaud

A NONPARAMETRIC MINIMUM ENTROPY IMAGE DEBLURRING ALGORITHM

C.V. Angelino, E. Debreuve, M. Barlaud

Laboratory I3S - CNRS - University of Nice-Sophia Antipolis
2000, route des Lucioles - 06903 Sophia Antipolis, France
{angelino, debreuve, barlaud}@i3s.unice.fr

ABSTRACT

In this paper we address the image restoration problem in the variational framework. Classical approaches minimize the L^p norm of the residual and rely on parametric assumptions on the noise statistical model. We relax this parametric hypothesis and we formulate the problem on the basis of nonparametric density estimates. The proposed approach minimizes the residual differential entropy. Experimental results with non gaussian distributions show the interest of such a nonparametric approach. Images quality is evaluated by means of the PSNR measure and SSIM index, more adapted to the human visual system.

***Index Terms*—** deconvolution, variational methods, entropy, nonparametric estimation

1. INTRODUCTION

Image restoration attempts to reconstruct or recover an image that has been degraded by using a priori knowledge of the degradation phenomenon. We focus on variational methods, that have an important role in modern image research.

Most methods rely on the standard model $y = m * x_0 + n$, where degradations are modeled as being the result of convolution together with an additive noise term, so the expression image deconvolution (or deblurring) is used frequently to signify linear image restoration [1]. Here m represents a known space-invariant blur kernel (point spread function, PSF), x_0 is an ideal version of the observed image y and n is (usually Gaussian) noise.

The objective of restoration is to obtain an estimate $\hat{x}(u)$ as close as possible to the original image, by means of a certain criterion. Defining with $r = y - m * \hat{x}$ the residual image, a common approach to the deconvolution problem is to find a solution that minimizes a function $\varphi(\cdot)$ of the residual. If $\varphi(\cdot)$ is the square function, we obtain a least square (LS) solution of the problem. In classical statistics, Maximum Likelihood (ML) is the most commonly used method for parameter estimation. Its application to image restoration is based on the knowledge of the random properties of noise, so that its probability density function (pdf) is exactly known. In the case of additive Gaussian noise, the ML-method is equivalent to the LS method. As it is well known, LS estimation is sensitive to outliers, or deviations, from the assumed statistical model. In the literature other more robust estimators have been proposed, like M-estimators [2], involving non-quadratic and possibly non-convex energy functions. However, these methods rely on parametric assumptions on the noise statistics, which may be inappropriate in some applications due to the contribution of multiple error source, such as radiometric noise (Poisson), readout noise (Gaussian), quantization noise (Uniform) and "geometric" noise, the latter due to the non-exact knowledge of the PSF. Therefore density estimation using a nonparametric approach is a promising technique. We propose to minimize a functional of the residual distribution, in particular the differential entropy of the residual. We use entropy because it provides a measure of the dispersion of the residual, in particular low entropy implies that the random variable is confined to a small effective volume and high entropy indicates that the random variable is widely dispersed [3]. Moreover, entropy criterion is robust to the presence of outliers in the samples. Nonparametric methods and information measures have been recently used in the segmentation context [4, 5].

Experimental results with non gaussian distributions show the interest of such a nonparametric approach. The quality of restored images is evaluated by the largely used PSNR measure and also by means of the Structure Similarity (SSIM) measure [6], more adapted to the human visual system (HVS).

This paper is organized as follows. In section 2 the proposed algorithm is presented and in section 3 some experimental results are shown. Finally, discussion and future works are proposed in the last section.

2. ENERGY

Image deblurring is an inverse problem, that can be formulated as a functional minimization problem. Let Ω denote a rectangular domain in $\mathcal{R}^2$, on which the image function $x : u \in \Omega \to \mathcal{R}^d$ is defined, d being the image dimensionality. Ideally, the recovered image $\hat{x}$ satisfies

$$\hat{x} = \arg\min_{\mathrm{x}} \int_{\Omega} \Phi(\mathrm{y} - \mathrm{m} * \mathrm{x})\, \mathrm{d}u, \qquad (1)$$

where $\Phi(\cdot)$ is a metric representing data-fidelity. In the case of Gaussian noise, a quadratic function is used.
However, parametric assumptions on the underlying noise density function are not always suitable, due to the multiple source of noise. We define as energy to be minimized a continuous version of the Ahmad-Lin [7] entropy estimator ($H_{\mathrm{A-L}}(r)$), defined as:

$$\begin{aligned} E(x) &= |\Omega|\, H_{\mathrm{A-L}}(r) \\ &= -\int_{\Omega} \log(p_x(r(u)))\, \mathrm{d}u\, . \end{aligned} \qquad (2)$$

In order to solve the optimization problem $\arg\min_{\mathrm{x}} \mathrm{E(x)}$ a steepest descent method is used. The energy derivative has been analytically calculated and it is shown in section 2.1.

2.1. Derivative of E

The residual pdf is estimate by using a nonparametric continuous version Parzen estimator, with symmetric kernel $K(\cdot)$,

$$p_x(s) = \frac{1}{|\Omega|} \int_{\Omega} K(s - r(u))\, \mathrm{d}u\, . \qquad (3)$$

Note that $p_x(s)$ is the residual pdf associated to the current estimate image x. Therefore changes in x provides changes in $p_x(s)$, hence changes in the residual entropy (energy).
By taking the Gâteaux derivative of eq.(2) it can be shown (demonstration has been omitted for brevity) that the gradient of $E(x)$ at $v \in \Omega$ is equal to

$$\nabla E(x)(v) = \int_{\Omega} m(v - w)\, k(w)\, \mathrm{d}w, \qquad (4)$$

with

$$k(w) = \frac{\nabla p_x(r(w))}{p_x(r(w))} + \chi(w) \qquad (5)$$

and

$$\chi(w) = -\frac{1}{|\Omega|} \int_{\Omega} \frac{\nabla K(r(u) - r(w))}{p_x(r(u))}\, \mathrm{d}u\, . \qquad (6)$$

The first term in (5) is the normalized gradient of the residual pdf and it is proportional to the local mean-shift [8]:

$$\frac{\nabla p_x(X)}{p_x(X)} = \frac{d+2}{h^2} M_h(X), \qquad (7)$$

where

$$M_h(X) = \frac{1}{k} \sum_{X_i \in S_h(X)} (X_i - X) \qquad (8)$$

is the sample mean shift of the observations in the small region $S_h(X)$ centered at X ($S_h(X) = \{Y : \|Y - X\|^2 \leq h^2\}$). The integral in (6) is difficult to estimate, however if the Parzen kernel $K(\cdot)$ has a narrow bandwidth, only samples very close to the actual estimation point will contribute to the pdf. Under this assumption the residual pdf is approximatively $p_x(\alpha) \approx N(\alpha)/|\Omega|$, where $N(\alpha)$ is the measure of $\Omega_\alpha = \{u \in \Omega : r(u) = \alpha\}$. The set $\{\Omega_\alpha\}_{\alpha \in r(\Omega)}$ form a partition of Ω. Thus we have,

$$\begin{aligned} \chi(w) &\approx -\int_{r(\Omega)} \mathrm{d}\alpha \int_{\Omega_\alpha} \frac{\nabla K(\alpha - r(w))}{N(\alpha)}\, \mathrm{d}u \\ &= -\int_{r(\Omega)} \nabla K(\alpha - r(w))\, \mathrm{d}\alpha\, . \end{aligned} \qquad (9)$$

Since $\nabla K(\cdot)$ is an odd function, $\chi(w)$ is zero if $r(w)$ is such that the support of $\nabla K(\alpha - r(w))$ is contained by the support of $r(\cdot)$ and assumes nonzero values in a ring near the boundary of the latter. In grayscale images ($d = 1$), the integral (9) can be expressed in closed form and $\chi(w)$ has the same sign of the mean shift term but has a smaller magnitude. In the multidimensional case $\chi(w)$ conserves the same behavior, therefore it is possible to neglect it.
Thus the steepest descent algorithm is performed with the following evolution equation:

$$x^{(n+1)} = x^{(n)} - \nu \frac{d+2}{h^2} m \star M_h(X), \qquad (10)$$

where ν is the step size. The choice of h is explained in section 3.

2.2. Lower Bound

In this section we provide a lower bound (LB) to the energy in eq.(2), in order to check how our algorithm works on minimizing residual entropy (see Fig.1).
The residual can be viewed the sum of two random variables, namely, $R = N + \tilde{X}$. The first one is the noise, and the second one is the projection of the error by means of the operator $m(\cdot)$, i.e., $\tilde{x} = m * (x_0 - x)$.

Proposition 2.1. *The residual entropy $h(R)$ is lower bounded by the noise entropy $h(N)$.*

Proof. Let us consider the mutual information between R and $\tilde{X}$,

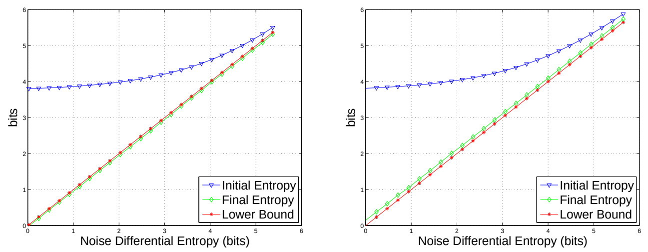
$$\begin{aligned} I(R; \tilde{X}) &= h(R) - h(R|\tilde{X}) \\ &= h(R) - h(N|\tilde{X})\, . \end{aligned}$$

Since the noise N is independent from $\tilde{X}$, $h(N|\tilde{X}) = h(N)$, and by the non negativity property of mutual information we obtain

$$h(R) \geq h(N)\, . \qquad (11)$$

□

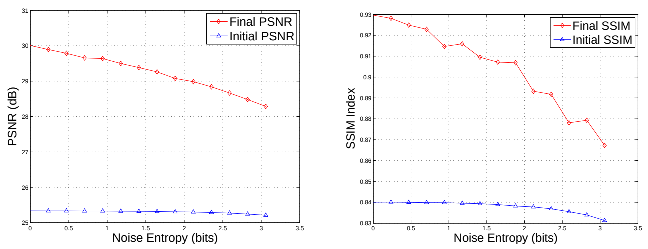
As it is well known, mutual information is a measure of the amount of information that one random variable contains



(a) Gaussian Noise (b) Uniform Noise

Fig. 1. Residual Entropy as function of noise entropy. Initial residual entropy (blue), Final residual entropy (green), Theoretical lower bound (red).

about another random variable [3]. The closer x is to the original image x_0, the less information on $\tilde{X}$ is carried by the residual. Therefore entropy minimization can be interpreted as the process which use the information carried by the residual to recover x_0, until there is no more information, i.e., the residual entropy reaches the lower bound.



(a) PSNR (b) SSIM

Fig. 2. Algorithm performances for uniform noise as function of noise entropy. (a) Initial (blue) and Final (red) PSNR; (b) Initial (blue) and Final (red) SSIM.

3. EXPERIMENTAL RESULTS

In this section, some results from the algorithm proposed are shown. In order to measure the performance of our algorithm we blurred the Lena image (512x512 pixel) by convolving it with a 13x13 Gaussian PSF with standard deviation $\sqrt{3}$, and adding noise with different distributions, such as Gaussian, Uniform, Gaussian mixture, Gaussian-Uniform mixture and different entropy magnitude. Residual entropy minimization is carried out via the gradient descent algorithm described in section 2.1. At each iteration the mean-shift kernel size h is proportional to the standard deviation of the residual, since this choice generally assures a good compromise between robustness and accuracy [9].
Fig.1 shows in blue the initial residual entropy in green the value attained when the algorithm converges and in red the theoretical LB. We considered Gaussian noise in Fig.1a and Uniform noise in Fig.1b. In the gaussian case the proposed algorithm achieves the lower bound of entropy. However, in the uniform case as well the final entropy is quite close to the LB with a maximum relative difference of 0.02%.
Fig.2 shows the PSNR and SSIM measures between the original image x_0 and the degraded image y (blue) and the restored image $\hat{x}$ (red) in function of the noise entropy for uniform distribution. PSNR measure and the related mean square error (MSE) are the simplest and most widely used quality metrics in image processing, with clear physical meanings, but they are not very well matched to perceived visual quality [6]. HVS is highly adapted for extracting structural information, and SSIM, which compares local patterns of pixel intensities, is more suitable for image quality assessment. Fig.(3) shows the restored images from different algorithms as Lucy-Richardson [10, 11] and Truncated SVD [12], with uniform noise (entropy 2 bits). The truncation parameter of TSVD is found with a generalized cross validation [12]. Our method has roughly the same PSNR of the TSVD restored image, however the latter has a more pronounced grain effect. This is well catched by the SSIM measure, for which our method is considered of higher quality. SSIM indicates better results also in the experiment of Fig.4, where a gaussian mixture noise has been used, even if the PSNR is lower than the one provided by the TSVD restoration.
A full set of comparisons is shown in Table1 for different noise distributions and entropy values. The proposed algorithm always outperforms the other techniques except that in the two aforementioned cases for the PSNR value.

Noise	Algorithm	PSNR (dB)	SSIM $\in [0, 1]$
Gaussian 1 bit	Lucy-Richardson	25.31	0.881
	Truncated SVD	29.59	0.908
	Proposed method	**29.65**	**0.922**
Gaussian 2 bits	Lucy-Richardson	25.29	0.880
	Truncated SVD	29.15	0.890
	Proposed method	**29.16**	**0.895**
Uniform 1 bit	Lucy-Richardson	25.31	0.881
	Truncated SVD	29.53	0.906
	Proposed method	**29.59**	**0.918**
Uniform 2 bits	Lucy-Richardson	25.28	0.878
	Truncated SVD	**29.04**	0.882
	Proposed method	29.03	**0.900**
Gaussian Mixture 1.85 bits	Lucy-Richardson	25.31	0.879
	Truncated SVD	**28.93**	0.876
	Proposed method	25.94	**0.899**
G + U Mixture 0.7 bits	Lucy-Richardson	25.32	0.881
	Truncated SVD	29.66	0.913
	Proposed method	**29.80**	**0.926**

Table 1. Quality measures comparison of Lena restored images with different algorithm.

(a) Degraded image

(b) Lucy-Richardson

(c) Truncated SVD [12]

(d) Proposed method

Fig. 3. Comparison of deconvoluted images. Uniform noise, entropy 2 bits. (a) The blurred noisy image (PSNR = 25.30[dB], SSIM = 0.837)

(a) Degraded image

(b) Lucy-Richardson

(c) Truncated SVD [12]

(d) Proposed method

Fig. 4. Comparison of deconvoluted images. Gaussian Mixture noise, entropy 1.85 bits. (a) The blurred noisy image (PSNR = 25.33[dB], SSIM = 0.837)

4. DISCUSSION

This paper presented a deconvolution method in the variational framework based on the residual entropy minimization. The simulations indicated robust performance for different noise distribution probabilities, showing in many cases slightly better results w.r.t. some popular deblurring techniques. Results are even more promising considering that, contrarily to what happens in other techniques like Truncated SVD, no regularization is applied. As future work, a possible regularization method is being taken into account that makes use of the Kullback-Leibler divergence between the residual distribution and the noise model, under the hypothesis that some a priori knowledge is available on the noise.
A further remarkable property of this algorithm is its possible extension to the case of multispectral images.

5. REFERENCES

[1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, second edition, 2002.

[2] M. J. Black and P. Anandan, “A framework for the robust estimation of optical flow,” 1993.

[3] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley-Interscience, 1991.

[4] J. Kim, J. W. Fisher, A. Yezzi, M. Çetin, and A. S. Willsky, “A nonparametric statistical method for image segmentation using information theory and curve evolution,” *IEEE TRANS. ON IMAGE PROC.*, vol. 14, pp. 1486–1502, 2005.

[5] A. Herbulot, S. Jehan-Besson, S. Duffner, M. Barlaud, and G. Aubert, “Segmentation of vectorial image features using shape gradients and information measures,” *Journal of Math Imaging Vis*, 2006.

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 13, NO.4, pp. 600–612, APRIL, 2004.

[7] I. A. Ahmad and P. Lin, “A nonparametric estimation of the entropy for absolutely continuous distributions,” *IEEE TRANSACTIONS ON INFORMATION THEORY*, 1976.

[8] K. Fukunaga and L. D. Hostetler, “The estimation of the gradient of a density function, with applications in pattern recognition,” *IEEE TRANSACTIONS ON INFORMATION THEORY*, vol. 21, NO.1, pp. 32–40, JANUARY, 1975.

[9] D. Comaniciu, “An algorithm for data-driven bandwidth selection,” *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 25, NO.2, pp. 281–288, FEBRUARY, 2003.

[10] W. H. Richardson, “Bayesian-based iterative method of image restoration,” *J. Opt. Soc. Am.*, vol. 62, 1972.

[11] L. B. Lucy, “An iterative technique for the rectification of observed distribution,” *The Astronomical Journal*, vol. 79, no. 6, pp. 745–754, 1974.

[12] P. C. Hansen, J. G. Nagy, and D. P. O’Leary, *Deblurring Images. Matrices, Spectra and Filtering*, SIAM, 2006.