Maximum Entropy Principle in Image Restoration

Mihai-Alexandru PETROVICI, Cristian DAMIAN, Daniela COLTUC
University Politehnica of Bucharest, 060042, Romania
m.petrovici.ro@ieee.org

***Abstract*—Many imaging systems are faced with the problem of estimating a true image from a degraded dataset. In such systems, the image degradation is translated into a convolution with a Point Spread Function (PSF) and addition of noise. Often, the image recovery by inverse filtering is not possible because the PSF matrix is ill-conditioned. Maximum Entropy (MaxEnt) is an alternative method, which uses the entropy concept for estimating the true image. This paper presents MaxEnt method, starting with the historical references of the entropy concept and finalizing with its application in image restoration and reconstruction. The statistical model of MaxEnt for images is discussed and the connection of MaxEnt with the Bayesian inference is explained. MaxEnt is evaluated by using a modified version of Cornwell algorithm. Two cases are considered: images degraded by various PSF kernels in presence of additive noise and images resulted from incomplete datasets. The tests show PSNR gains ranging from 1 to 7dB for the degraded images and images reconstructed at 25dB from datasets with up to 80% missing pixels.**

***Index Terms*—image processing, image reconstruction, image representation, image restoration, image sampling.**

I. Introduction

The origin of the entropy concept is in thermodynamics. It is linked to the work of Clausius and the formulation of the second law of thermodynamics. Clausius used the entropy to express the loss of heat in thermodynamic systems. The entropy of Clausius is defined as [1]:

$$S = Q / T \tag{1}$$

where Q is the quantity of heat and T the temperature. The variation of S during a cyclic process represents the loss of heat. Clausius coined the term of entropy in 1865 and explained the preference for this word by its etymology. In Greek, *trope* means *transformation*.

Several years later, in 1877, Boltzmann developed a statistical evaluation of the entropy. He introduced the notion of probability in the definition of the entropy and expressed it as [1, 2]:

$$S_B(\Gamma) = k \log W_\Gamma \tag{2}$$

where $S_B(\Gamma)$ is the system entropy in the macrostate Γ, κ is Boltzmann's constant and W_Γ is proportional to the macrostate probability.

Boltzmann asserted the fact that the macrostates supervene the microstates. A microstate is defined by the states of the *N* particles constituting the system. Since many distinct microstates can correspond to a given macrostate, the probability of Γ is given by its number of microstates normalized by the number of all possible microstates. In Boltzmann's entropy, W is actually the number of microstates corresponding to Γ.

This number can be counted by using the formula for permutations [1,2]:

$$\Omega = N! / \prod_i N_i! \tag{3}$$

where N is the number of particles and N_i is the particles distribution. This particular calculus for W_Γ holds only for the equilibrium state where all the microstates complying with the state distribution have the same probability.

Gibbs generalized Boltzmann's entropy to:

$$S(p) = -k_B \sum_{i=1}^{m} p_i \log(p_i) \tag{4}$$

where p_i are the probabilities of the microstates corresponding to Γ. The expression (4) holds for any macrostate not only for the equilibrum one. Gibbs entropy equals Boltzmann's entropy for $p_i = ct$. In this latter case, the entropy reaches its maximum.

Based on Boltzmann's observation that a thermodynamical system evolves toward the equilibrium state, which is the most probable one, Gibbs introduced in 1902 the criterion of maximum entropy.

This criterion states that the macrostate of a system is determined by the probability distribution that maximizes the entropy, given some constraints [2].

The entropy has turned the interested of many scientists and researchers over the century and not only in thermodynamics or statistical mechanics.

In 1948, Claude Shannon considered a very different problem. While studying information transmission in noisy channels, he defined a measure for the uncertainty of a signal source. He gave the name of *entropy* to this measure, at the suggestion of the mathematician John von Neumann, who remarked the similarity with the formula of entropy in thermodynamics. Shannon's entropy is [3]:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{5}$$

where X is the signal source generating random variables with n outcomes $X = \{x_1 \dots x_n\}$ and probabilities $P(X) = \{p(x_1) \dots p(x_n)\}$.

In Shannon's Information Theory, the source with maximum entropy provides the maximum quantity of information.

After Shannon's theory publication, Jaynes established the connection between Information Theory and mechanical statistics (1957) [4]. He stated that Maximum Entropy

Digital Object Identifier 10.4316/AECE.2018.02010

(MaxEnt) is a general principle in setting up distributions from incomplete knowledge.

This was used for the first time in image restoration by Frieden [5]. Six years later, Gull and Daniell were publishing encouraging results on the restoration of radio astronomy images by using the same principle [6].

In the beginning of 90s, John Skilling gave a Bayesian interpretation to this method and stressed the fact that it is the only generally acceptable procedure for assigning prior probabilities [7].

By using the Bayesian frame, he also gave a solution for the reasonable choice of regularizing parameter α in MaxEnt objective function. Until then, this parameter was traditionally set such to have the χ^2 misfit of MAP solution approximately equal to the number of image pixels.

The Bayesian choice for α (always less or equal to the traditional choice) revealed a major drawback of maximum entropy image model, that of no prior information on pixel to pixel correlation. It has been introduced under the form of an Intrinsic Correlation Function (ICF) that convolved with the MaxEnt solution, gives the restored image.

The method evolved as an imaging tool along the years, especially for space imaging applications.

Since the use of a unique ICF was not satisfactory for the restoration of images containing both sharp and extended structures, Weir [8] proposed a multichannel maximum method (MEM).

In MEM, the restored image is a sum of a set of restored image channels, each of which with a different ICF. Heuristic weights are given to the contributing channels.

Bontekoe [9] took the idea further and introduced in 1994 the concept of Pyramid MEM, where the channels are a multi-resolution pyramid representation of the restored image.

Compared with MEM, the Pyramid MEM requires lower computing power. The method was dedicated to the construction of high-resolution images from the Infrared Astronomical Satellite (IRAS) survey data.

Pantin and Stark [10] replaced the pyramid representation by the wavelet transform, a tool that decomposes the signal in different frequency bands and naturally provides a multiresolution representation of images. In this frame, they established a link between the channel weight (until then heuristically determined) and the standard deviation of the noise in that channel and removed the problem of determining of α by preserving only the significant wavelet coefficients from a regularization. The method was tested on simulated data.

A recent application of these improvements is presented by Guan [11] for image restoration of the Hard X-ray Modulation Telescope (HXMT). A particularity of this application is the Poisson noise of the sensor.

Zao et al [12] are using the advantages of maximum entropy criterion for a robust watermarking algorithm.

The paper presents an overview of MaxEnt in image restoration. In giving some results, two scenarios are taken into account: images degraded by convolution with a Point Spread Function (PSF) and additive noise and images resulted from incomplete measurements.

In this latter case, the available pixels are obtained by random sampling, in presence of convolution with a Gaussian PSF and additive noise.

The paper is structured as follows: Section 2 is an overview of the MaxEnt theory. It starts with the statistical model of the image and continues with the models of the imaging system. Then Gull and Daniell's view is presented and, in the last part of the section, it is shown that MaxEnt is a special case of Bayesian inference. Section 3 describes Cornwell algorithm that is used in section 4 for giving some experimental results. Section 5 points our recommendations and conclusions.

II. IMAGE RESTORATION

There is a huge amount of literature describing methods, which concerns quality improvement of the image information content through restoration techniques [1-34].

The restoration methods are very strong related with the image formation model. A frequently used model is (Fig. 1):

$$d = A * x + n \tag{6}$$

where x and d are column vectors formed by column scanning of the true image and observed image, and A is the Toeplitz convolution matrix corresponding to the 2D PSF of the imaging system (denoted by h in Fig. 1).

The term n is a random vector representing measurement errors or additive observation noise.

In the case of stationary additive white Gaussian noise (AWGN), all the pixels have the same noise variance σ_n^2 [13].

If A was invertible, the deconvolution would be possible by inverse filtering of the data. This is the deterministic method. In real life, A is often ill-conditioned and the presence of noise prevents (6) to have a stable solution. In these situations, the deterministic method is replaced by statistical approaches.

The Bayesian inference and MaxEnt are such approaches.

The Bayesian inference assumes the image is a random field and obtains the solution as the probability distribution of the field. MaxEnt is also an inference principle but which is based on specific statistical model of the image.

In the next subsection, we discuss this model, which is a shift of paradigm comparing with what is usually understood by image statistical model.

A. The image statistical model in MaxEnt

Suppose that the image has the configuration of pixel $x = [x_1, x_2 ..., x_N]$. In conventional statistical methods, the image is viewed as the particular realization of a random field. Further, with an ergodicity hypothesis, the image distribution is estimated from its pixels.

In MaxEnt, the pixels x_i themselves are considered probabilities. This complies with two of Kolmogorov axioms i.e., x_i are non-negative and their sum has a physical meaning:

$$x_i \geq 0$$
$$\sum_i x_i = M \tag{7}$$

M is the image total luminance. The third axiom is satisfied if the image is normalized:

$$p_i = \frac{x_i}{M}$$
$$\sum_i p_i = 1 \tag{8}$$

In MaxEnt, the image normalization is not compulsory.

In the absence of normalization, we speak of a Positive Additive Distribution (PAD) instead of probability distribution.

Next we derive the image probability $p(x)$ by taking into account this model. The following rationale is used: let us consider an experiment where M luminance quanta are distributed randomly over the N pixels of the image.

Each luminance quantum has the same a priori chance of being in any pixel. The probability to have at the end of the experiment the configuration $x=[x_1,x_2,...,x_N]$ is proportional with the number Ω of ways to arrive to it:

$$p(x) \propto \Omega \tag{9}$$

Ω has the same combinatorial expression as in the case of Boltzmann's entropy (3)

$$\Omega = \frac{M!}{\prod_i x_i!} \tag{10}$$

where $M!$ is the number of permutations of all luminance quanta and $x_i!$ is the number of equivalent quanta reordering in i-th pixel. By applying the logarithm to both sides of eq.(10), one obtains:

$$ln(\Omega) = ln(M!) - \sum_i ln(x_i!) \tag{11}$$

Since both M and x_i are large, the factorials can be approximated by using Stirling's formula:

$$ln(\Omega) \approx Mln(M) - M - (\sum_i x_i ln(x_i) - \sum_i x_i)$$

$$ln(\Omega) \approx Mln(M) - M - \sum_i x_i ln(x_i) + M$$

$$ln(\Omega) \approx Mln(M) - \sum_i x_i ln(x_i)$$

$$ln(\Omega) \approx -\sum_i x_i ln(\frac{x_i}{M}) \tag{12}$$

By replacing the pixels luminance with the normalized values in (8), one obtains:

$$ln(\Omega) \approx -\sum_i M\, p_i ln(\frac{Mp_i}{M})$$

$$ln(\Omega) \approx -M \sum_i p_i lnp_i$$

$$ln(\Omega) \approx MH \tag{13}$$

where H is Shannon's entropy. This result shows that the image probability $p(x)$ is proportional to the exponential of the entropy:

$$p(x) \propto exp(MH) \tag{14}$$

The probability $p(x)$ has been derived with no prior information about the image content. Indeed, each of M quanta had the same chance to be assigned to anyone of the N pixels of the image.

J. Skilling showed that if some prior information is available under the form of an image model $m=[m_1,m_2,...,m_N]$, then the entropy has the following form:

$$S = \sum_i [x_i - m_i - x_i ln(\frac{x_i}{m_i})] \tag{15}$$

J. Skilling interpreted m_i as expectations of Poisson distribution for the number of quanta captured by the pixels in a fixed delay [14].

The image probability is now conditioned by the model m and is proportional with the entropy S:

$$p(x \mid m) \propto exp(S) \tag{16}$$

The rationale in deriving the image probability $p(x)$ in the absence of a model is the same as for Boltzmann's equilibrium state probability (α, Ω).

The model of the thermodynamic system is translated in the case of images by identifying the particles with photons and the state with the pixels distribution.

Without measurement constraints, the solution would evolve towards a flat image, which is the case of maximum entropy.

The model nested in the entropy S forces the solution to evolve towards the model distribution and not to the flat one. The measurement constraints prevent the image to reach the model but the solution will be close to this.

In order to distinguish between the use of S or H (with or without nested model), some authors use the name MaxEnt for the algorithms based on H and ME (Maximum relative Entropy) for those based on S [15].

It is said that MaxEnt assigns the values x_i and ME updates them by starting from m_i.

B. Image formation models

The image formation model plays a very important role in image restoration. Historically, in the restoration by entropy maximization there has been employed two models.

The classic version refers to a forward map imaging system (Fig. 1), where the true image x is hidden in the observed data d through a convolution with h, the PSF of the optic system (geometric blurring) [16], [17].

The blurred image is then corrupted by additive noise n to produce the observed data d.

Mathematically, the model in Fig. 1 is described by:

$$d = x * h + n \tag{17}$$

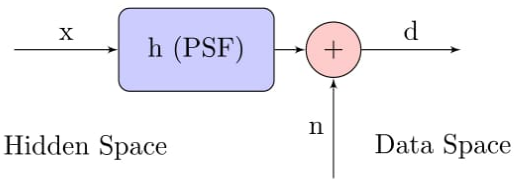

Figure 1. Forward map imaging system model

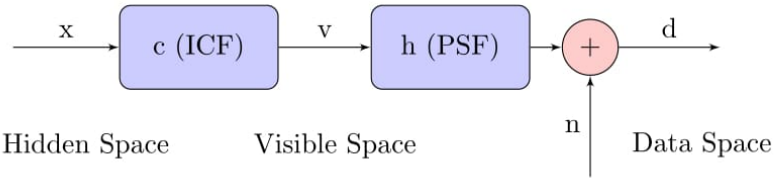

Figure 2. Forward map imaging system model with ICF

A subsequent model has been the forward map imaging with Intrinsic Correlation Function (ICF) [18]. The hidden image x is mapped to the visible space by the ICF c, is blurred with the system PSF and degraded with additive noise n to produce the observed data d (Fig. 2). The equation of ICF model is:

$$d = x * c * h + n = v * h + n \tag{18}$$

The ICF image model has the advantage of postulating pixel correlation that proved to be very helpful in image restoration [7].

C. *Image restoration by MaxEnt*

In their paper from 1978 [6], Gull and Daniell defined for the first time, the image restoration problem in terms of entropy maximization with constraints.

They interpreted the normalized image as a distribution of probabilities and expressed its entropy like Frieden [19] eq. (5). A set of measurements d_i of image pixels, taken in the presence of noise, represent the constraints.

The actual intensities x_i of the image pixels are obtained by maximizing the entropy with respect to d_i. Such problem is typically solved by Lagrange multiplier method.

An objective function that includes both the entropy and the constraints is maximized instead of maximizing solely the entropy. The ingestion of the constraints was done in the following way. Instead of imposing the perfect match of the solution with each constraints separately – this would generate false structures specific to noise and would proliferate the Lagrange multipliers, Gull and Daniell melted down them into a single expression:

$$\chi^2 = \frac{\sum_i (x_i - d_i)^2}{\sigma_n{}^2} \tag{19}$$

where σ_n is the standard deviations of the Gaussian noise that affects the measurements d_i. The statistics (19) has a χ^2 distribution. With this formalization, the objective function of MaxEnt is:

$$Q(\lambda) = H - \frac{\lambda}{2}\chi^2 \tag{20}$$

where λ is the Lagrange multiplier. The solution is an image estimate that maximizes Q(λ):

$$\hat{x}_{MaxEnt} = \arg\max_x \left(H - \frac{\lambda}{2}\chi^2\right) \tag{21}$$

For multiplicative noise, the fitting term χ^2 must be redefined. A typical solution is to use the I-divergence, also called generalized Kullback-Liebler divergence [33]:

$$I(d,i) = \sum_i d_i \log\frac{d_i}{x_i} - d_i + x_i \tag{22}$$

Another possibility is the functional

$$\sum_i \frac{d_i}{x_i} + \log x_i \tag{23}$$

proposed by Aubert and Aujol [34], based on MAP estimator of multiplicative Gamma noise (speckle in SAR images). A third option is to work with the logarithm of data. This operator transforms the multiplicative noise into an additive term and the fitting can be modeled like for additive noise, as mean square error:

$$\sum_i \left(\log d_i - \log x_i\right)^2 \tag{24}$$

A drawback is a bias in the restored image (the homomorphic methods do not preserve the image mean) that can however be corrected.

The equation (21) is solved by optimization algorithms.

The start point in the algorithm is a flat image i.e., the image of maximum entropy in the case of no constraints.

The choice λ of is not trivial. In adjusting λ, Gull and Daniell started from the observation that the minimum expected value of χ^2 is equal to the number of observations d_i. It corresponds to the situation when the solution matches the measurements in the presence of noise σ_n. This determined Gull and Daniell to adjust the value of λ such to have χ^2 equal to data size. It became the traditional way of

selecting λ in MaxEnt algorithms. Several years later, J. Skilling showed that there is no acceptable criterion for selecting λ only by looking at χ^2 and proposed a procedure in the Bayesian framework [7].
He redefined Q by using Lagrange multiplier α instead of λ:

$$Q = \alpha S - L \tag{25}$$

The term L reduce to $\chi^2/2$ for independent, Gaussian errors and α is a regularizing parameter that controls the competition between S and L: if is high, the entropy term dominates and the measurement α cannot move the reconstruction far from the flat image.

If α is low, there is little smoothing and the reconstruction will show wild oscillations as the errors in the measurements are interpreted as true signal.

The Bayesian choice of α has been not sufficient for a good reconstruction of the image: the derived α is too low and the restored image has oscillations.

J. Skilling explained the bad quality of the restored image by the flatness of the initial model and proposed to introduce correlation into the model in order to reduce the oscillations in the resulted image [7].

The new version of MaxEnt uses a correlated model, which is equivalent to consider an imaging system with ICF architecture (Fig. 2).

In the new version, the restoration mechanism is handled in the same way as for classic model.

The difference is that the objective function is now (25) and the simulated data are:

$$\hat{d} = \hat{x} * c * h \tag{26}$$

J. Skilling observed that the shape of the model blur c does not matter greatly and he arbitrarily restricted it to be a Gaussian. By the contrary, the width of c is crucial. It has to be about equal to the size of the correlation-length that is actually present in the image [7].

D. Relationship between Maximum Entropy and Bayesian Image Restoration

In Bayesian inference, the MAP estimate of the image x is:

$$\begin{aligned}\hat{x}_{MAP} &= \arg\max_x \ln(p(x \mid d)) \\ &= \arg\max_x \ln(p(d \mid x)\, p(x))\end{aligned} \tag{27}$$

where $p(d \mid x)$ is the likelihood of data and $p(x)$ is the prior that incorporates our knowledge about the expected image.

For Gaussian noise with zero mean and variance σ_n^2, the likelihood is:

$$p(d \mid x) = \frac{1}{(2\pi\sigma_n^2)^{\frac{N}{2}}} exp(-\frac{1}{2\sigma_n^2} \| d - Ax \|^2) \tag{28}$$

The exponential includes the statistics χ^2 defined in (19):

$$p(d \mid x) \propto exp(\frac{-\chi^2}{2}) \tag{29}$$

This equation shows that the likelihood incorporates the image formation model.

The expression of the prior depends on how the image is modelled. In subsection II.A, it was shown that, if the MaxEnt specific model is employed, then the prior is proportional with the entropy (14). By integrating (14) and (29) in (27), the posterior probability becomes:

$$\begin{aligned}p(x \mid d) &\propto exp(-\frac{\chi^2}{2}) exp(MH) \\ p(x \mid d) &\propto exp(MH - \frac{\chi^2}{2})\end{aligned} \tag{30}$$

The solution $\hat{x}_{MAP}$ is obtained by maximizing this probability, which is equivalent with maximizing the objective function:

$$Q(\alpha) = \alpha S - \frac{\chi^2}{2} \tag{31}$$

The important conclusion of this subsection is that MAP estimation by Bayesian inference is equivalent to MaxEnt if the image model in section II.A is used.

III. IMPLEMENTATION OF MAXENT

Classic implementations of MaxEnt restoration [6] was based on the entropy form in (15) and used a fixed point method in order to maximize $Q(\lambda)$ while having a fixed regularization parameter λ.

Thus x_i was iteratively updated with a relation of the following form:

$$\hat{x}_i \leftarrow m_i \exp[-\lambda \partial \chi^2 / \partial \hat{x}_i] \tag{32}$$

where $\lambda = 1/\alpha$. This method was attractive because successive iterates are automatically positive but the exponential introduced instabilities that were difficult to remedy [7].

A solution that bypasses the use of an exponential is the method of steepest ascents using [14]:

$$\hat{x}_i \leftarrow \hat{x}_i + p \partial Q / \partial \hat{x}_i \tag{33}$$

for a sufficiently small step value p. The big disadvantage of this method is that in the places where the data has a low value x_i tends to become negative and entropy is not defined for $\hat{x}_i < 0$.

Remedies for this problem are to reset the negative values to a small positive one, to use a heuristic form of entropy that is defined for those values like $\sum_i x_i \log(|x_i|)$ or to decrease p. In all cases, the algorithm concentrates too much on low values and convergence towards a maximum is slow.

The standard way of improving on the steepest ascents method is to use a variation of Newton's method. But this has the disadvantage that the Hessian matrix of the objective function has to be calculated.

The matrix contains N^2 elements so computing the Hessian for large images is infeasible.

A simple solution devised by Cornwell algorithm [20] is to neglect the non-diagonal elements of the Hessian. The approximated Hessian will be the following:

$$H = \nabla\nabla S - 2\lambda q I \tag{34}$$

where q is a scaling factor that should represent the sum of squares of PDF main lobe. The Hessian of the entropy is

diagonal and the elements along the diagonal are x_i^{-1} .
The update relation is:

$$\hat{x}_i \leftarrow \hat{x}_i + p(x_i^{-1} - 2\lambda q)^{-1} \partial Q / \partial \hat{x}_i \qquad (35)$$

If x_i is close to zero updates to it will be small, meaning that negative values are less likely and the algorithm will concentrate on significant values. The value of p can be chosen at each iteration by using the backtracking line search thus insuring that some progress is made with every iteration [21]. The optimization should stop when the distance between successive iterates of x is insignificant. The problem that remains is choosing an appropriate λ .

We chose to use a small λ at first, run the optimization algorithm then to increase λ by a factor k as long as χ^2 is above a threshold, usually N.

The method we used is summarized in Algorithm 1.

Data: $Q, m, k, tol, thresh$
Result: $\hat{x}$
$\hat{x} \leftarrow m$
$\lambda \leftarrow small$
while $\chi^2 > thresh$ **do**
  $v \leftarrow (\hat{x}^{-1} - 2\lambda q)^{-1} \partial Q / \partial \hat{x}$
  $p \leftarrow 1$
  while $Q(\hat{x} + pr) - Q(\hat{x}) < 1/2 v^T \partial Q / \partial \hat{x}$ **do**
    $p \leftarrow p/2$
  end
  $\hat{x} \leftarrow \hat{x} + pv$
  if $\|pv\| < tol$ **then**
    $\lambda \leftarrow k\lambda$
  end
end

Algorithm 1: Maximum Entropy deconvolution with simplified Newton method

IV. Experimental Results

For experiments, we have used a modified version of Cornwell algorithm. It was coded in the NRAO's Astronomical Image Processing System (classic AIPS) as VTESS. It is an algorithm that works well for many cases and its code is in the public domain. It is generally faster than CLEAN for larger images, the break-even point being around 1 million pixels.

Our modification consisted in removing the bisection in the calculation of λ . Instead, we have used the value 2 for *k* parameter in Algorithm 1.

A series of simulations are performed in order to demonstrate the capabilities of MaxEnt image restoration.

A tomographic phantom image is taken as a truth image. The tomographic image is blurred with a Gaussian kernel with standard deviation σ_{ker} and additive Gaussian noise is added at various PSNR levels. The model *m* is flat i.e., all m_i are equal to the average of the blurred image.

The effectiveness of the image restoration is measured by the PSNR gain defined as:

$$PSNR_{gain} = PSNR_{estimate} - PSNR_{data} \qquad (36)$$

Fig. 3 plots the PSNR gain and Fig. 5 illustrates the examples.

The presence of noise affects the image but to a lesser extent than for simpler deconvolution method like Wiener deconvolution. One can also observe that the residual noise is correlated suggesting that the algorithm denoises the image by discarding variations that are not consistent with the model PSF. If the noise is significant, the algorithm performs some denoising and no deburring effect can be observed. When the noise is small (PSNR > 40) a good deburring performance can be achieved. Some ringing is present around sharp edges, particularly if they transit from extreme values to average ones. This effect can be reduced by setting a higher threshold for χ^2 but not without some loss in resolution.

The algorithm is more robust to noise for larger PSFs because the input images are correlated.

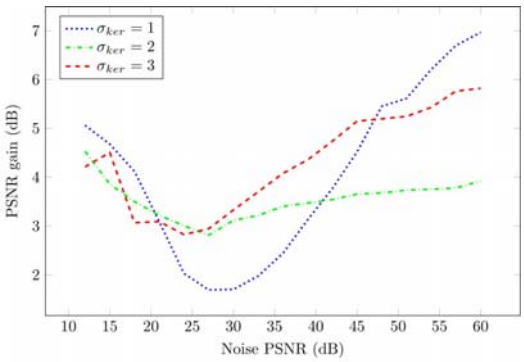

Figure 3. Experimental PSNR gain of the proposed algorithm for the phantom image. The input images were degraded with Gaussian kernels with σ_{ker} of 1, 2 and 3 and by adding noise.

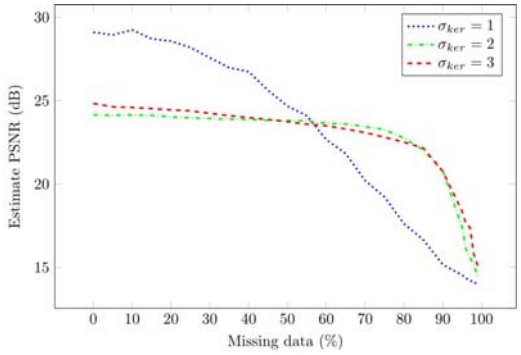

Figure 4. Restored image PSNR for the phantom image with missing data. The input images were degraded with Gaussian kernels with σ_{ker} of 1, 2 and 3 and by adding noise with a PSNR of 50dB.

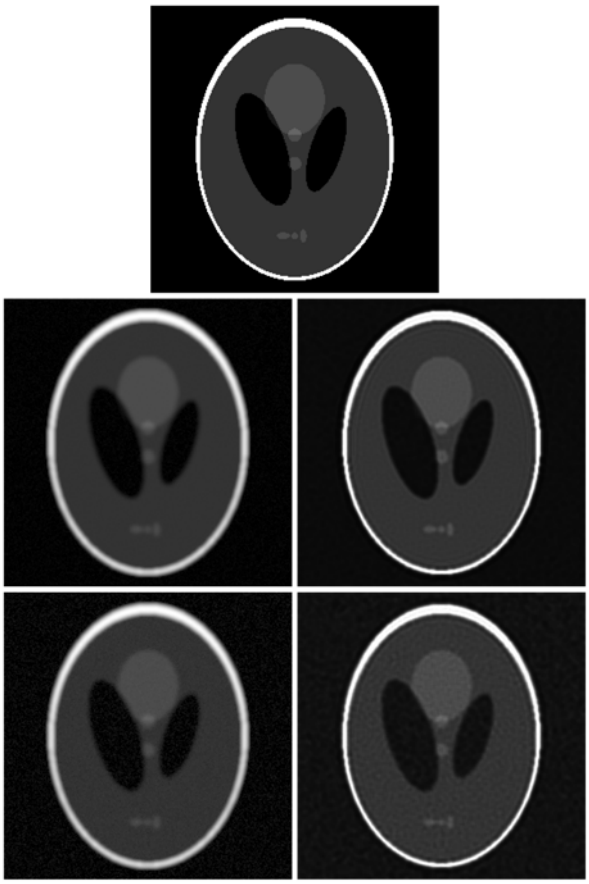

Figure 5. Top: phantom test image. Up-left: degraded phantom image with $\sigma_{\text{ker}} = 2$ and noise PSNR = 60 dB. Bottom-left: degraded phantom image with $\sigma_{\text{ker}} = 2$ and noise PSNR = 45 dB. Right: corresponding restored images. Up-right: The estimate PSNR is 24dB. Bottom-right: The estimate PSNR is 24dB.

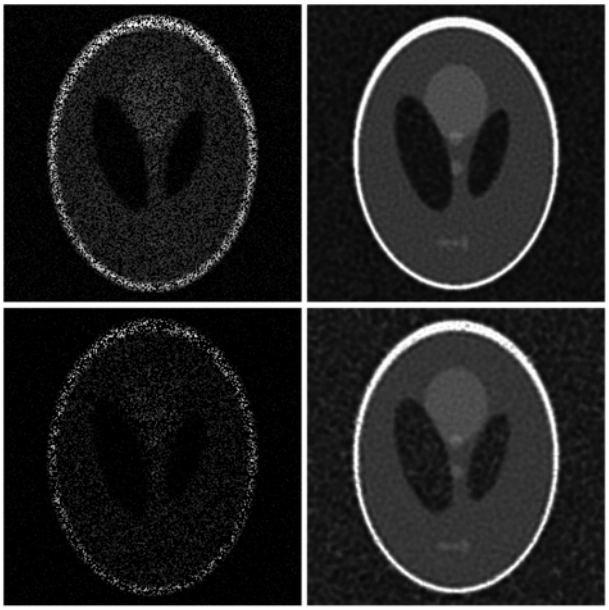

Figure 6. Left: degraded image with $\sigma_{\text{ker}} = 2$ and noise PSNR = 45 dB. Top-left: 50% of pixels missing. Bottom-left: 80% of pixels missing. Right: restored images. Top-right: Estimate PSNR is 24dB. Bottom-right: Estimate PSNR is 23dB.

The χ^2 constraint can be modified such that the variance associated with each pixel can be set individually. Consequently, interpolation of missing data can be done by setting the variance of the missing pixels to infinity.

This can be used to filter impulse noise after a separate detection of bad pixels as seen in Fig. 6. In our experiments the missing data was selected randomly. PSNR values of the estimate for different proportions of missing data are presented in Fig.4.

For larger PSF algorithm handles missing data well up to a point then the performance drops (at about 80 to 90 percent of pixels missing).

For a small PSF the performance degradation is smoother and probably more dependent on the individual layout of the missing pixels. The reason is that the only information available to interpolate the values of the missing pixels is the PSF and the MaxEnt criterion.

If missing pixels are further apart from a known pixel (further than the radius of PSF main lobe), then that pixel will take the background value. Also, as the number of known pixels decreases noise will start to affect the estimation more.

In the literature, there are numerous results reported for deconvolution by Maximum Entropy method. They use different metrics and the experiments are done in various conditions [24-31]. Last but not least, the test images may come from domains like astronomy, spectrometry, tomography etc. or maybe synthesis images.

In order to compare our results with already existing ones, we take as reference the paper of Mainsinger et al. [23].
The authors have used similar, although not identical metrics and parameters and as in our case, simulated data. The test image is 256x256 pixels gray levels image that was distorted by convolution with a Gaussian PSF and additive white Gaussian noise. The differences consist in the fact that the authors use as metric the root mean square error (rms) instead of PSNR and characterize the PSF by Full Width of Half Maximum (FWHM) instead of standard deviation σ_{ker}. FWHM is approximately 2.335 σ_{ker}. In order to compare the results, we have converted our PSNR values into rms and the FWHM in standard deviation. Thus for 2 rms noise, our method provides a restored image rms of 0.35 at PSF standard deviation 3, compared with a rms of 0.6 at PSF of 2.35. At the same noise level, for a PSF standard deviation of 5, our method gives a rms of 0.43 while the rms of the other method is 0.49 at a PSF of 4.71. The errors are smaller by our method in the conditions of high noise.

Concerning the reconstruction from incomplete data, we have compared MaxEnt with the more recent approach of Compressive Sensing in [31]. The tests on various images have shown that CS performs generally better than MaxEnt. However, for low sparsity images and low number of samples, CS and MaxEnt reconstructions are equivalent as PSNR. As complexity, MaxEnt has a clear advantage over Compressive Sensing, which for higher image sizes becomes quickly intractable.

Similar to MaxEnt principle, other experiments were made for image registration by using the maximization of Normalized Mutual Information and Normalized Cross Correlation measures [32].

V. CONCLUSION

The paper focuses on MaxEnt application in image restoration and reconstruction from incomplete data. To illustrate the performances of this method, we have selected from the existing implementations, Cornwell algorithm known as a good compromise between simplicity and robustness to noise. In the algorithm, we modified the choice of parameter λ used for updating the solution.

We present the algorithm performances in restoring the true image from tomographic images blurred by three different PSFs and corrupted by additive noise. The algorithm performs both deburring and denoising based on the correlations introduced by the imaging system PSF. The balance between these two functions is controlled by supplying the estimated noise level as a parameter. At noise levels around 40dB PSNR the method offers good deburring and denoising performance. For higher noise levels, the algorithm performs mostly denoising than deburring. Ringing effects are present around sharp edges, particularly when there are transitions from extreme values to average ones.

Results for incomplete data are also given. We considered the case of missing pixels randomly distributed over the image. The tests show that as long as the PSF provides enough redundancy in the input image, the estimation of the missing pixels has low errors. The algorithm performs well up to 50% of missing data. For the tomographic image, the PSNR of the reconstruction is about 25dB.

VI. REFERENCES

[1] J-L. Starck, F. Murtagh, "Astronomical image and data analysis", pp.71-110, Springer Science & Business Media, 2007, doi: 10.1007/978-3-540-33025-7

[2] A. Giffin, "Maximum entropy: the universal method for inference", pp. 20-25, ProQuest, Umi Dissertation Publishing, 2011.

[3] C. Shannon, "A mathematical theory of communication." Bell System Technical Journal, no.27, pp. 379–423, 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x

[4] E.T. Jaynes, "Information theory and statistical mechanics." Physical review, no.106, p. 620, 1957, doi: 10.1103/PhysRev.106.620

[5] B.R. Frieden, "Restoring with Maximum Likelihood and Maximum Entropy", JOSA, no.62, pp.511–518, 1972, doi: 10.1364/josa.62.000511

[6] S.F. Gull, G.J. Daniell,"Image reconstruction from incomplete and noisy data. " Nature, no. 272, pp. 686–690, 1978, doi: 10.1038/272686a0

[7] J. Skilling , "Maximum Entropy and Bayesian Methods", Kluwer, pp. 45–52, 1989, doi: 10.1007/978-94-015-7860-8

[8] N.Weir, "A multi-channel method of maximum entropy image restoration". Astronomical Data Analysis Software and Systems I, vol. 25, p. 186, 1992.

[9] Tj R. Bontekoe, E. Koper, and D. J. M. Kester, "Pyramid maximum entropy images of IRAS survey data.", Astronomy and Astrophysics no.284, pp.1037-1053, 1994.

[10] E. Pantin, J.L. Starck, "Deconvolution of astronomical images using the multiscale maximum entropy method. ", Astronomy and Astrophysics Supplement Series, no. 118, pp. 575–585, 1996, doi: /10.1007/978-1-4612-1968-2_29

[11] J. Guan, L.M. Song, Z.X. Huo, "Application of a multiscale maximum entropy image restoration algorithm to HXMT observations. " Chinese physics C, pp. 0–0, 2016, doi: 10.1088/1674-1137/40/8/086203

[12] B. Zhao, G. Qin, P.Liu, "A Robust Image Tampering Detection Method Based on Maximum Entropy Criteria." Entropy 17.12 , 2015 , pp. 7948-7966, doi: 10.3390/e17127854

[13] M. Willis, B.D. Jeffs, D.G. Long, "A new look at maximum entropy image reconstruction." Signals, Systems, and Computers, 1999. Conference Record of the Thirty-Third Asilomar Conference on. IEEE, vol. 2, pp. 1272–1276, 1999, doi: 10.1109/ACSSC.1999.831911

[14] J. Skilling, R. Bryan, "Maximum entropy image reconstruction: general algorithm. " Monthly notices of the royal astronomical society, no. 211, pp.111–124, 1984, doi: 10.1093/mnras/211.1.111

[15] A. Caticha, A.Giffin, "Updating probabilities." AIP Conference Proceedings, vol. 872, no. 1. AIP, 2006, doi: 10.1063/1.2423258

[16] E.T. Jaynes, "On the rationale of maximum-entropy methods. " Proceedings of the IEEE, no. 70, pp. 939–952, 1982, doi: 10.1109/proc.1982.12425

[17] A. Jannetta, J.C. Jackson, C.J. Kotre, I.P. Birch, K.J. Robson, R. Padgett, "Mammographic image restoration using maximum entropy deconvolution. " Physics in Medicine and Biology, no. 49, p. 4997, 2004, doi: 10.1088/0031-9155/49/21/011

[18] M. K. Charter, W. T. Grandy Jr, L. H. Schick, "Maximum Entropy and Bayesian Methods.", ed. PF Fougere, pp. 325–339, Dordrecht: Kluwer, 1990, doi: /10.1007/978-94-009-0683-9

[19] B.R. Frieden, "Image enhancement and restoration." Picture Processing and Digital Filtering. Springer Berlin Heidelberg, pp.177-248, 1975, doi: 10.1007/978-3-662-41612-9_5

[20] T. Cornwell, K. Evans, "A simple maximum entropy deconvolution algorithm." Astronomy and Astrophysics, no.143, pp. 77–83, 1985.

[21] N.I. Gould, S. Leyffer, "An introduction to algorithms for nonlinear optimization." Frontiers in numerical analysis. Springer Berlin Heidelberg, pp. 109-197, 2003, doi: 10.1007/978-3-642-55692-0_4

[22] A. Mohammad-Djafari,"Entropy, information theory, information geometry and Bayesian inference in data, signal and image processing and inverse problems." Entropy no.17.6, pp. 3989-4027, 2015, doi:10.3390/e17063989

[23] K. Maisinger, M. P. Hobson, A. N. Lasenby , "Maximum-entropy image reconstruction using wavelets. " Monthly Notices of the Royal Astronomical Society, no. 347(1), pp. 339-354, 2004, doi: 10.1111/j.1365-2966.2004.07216.x

[24] T. Hedrich, et al. , "Comparison of the spatial resolution of source imaging techniques in high-density EEG and MEG." NeuroImage, no.157, pp. 531-544, 2017, doi: 10.1016/j.neuroimage.2017.06.022

[25] J.J. Martín-Sotoca, A. Saa-Requejo, J.B. Grau, A. Paz-González, A.M. Tarquis,, "Combining global and local scaling methods to detect soil pore space." Journal of Geochemical Exploration, vol. 189, pp. 72-84, 2018, doi: 10.1016/j.gexplo.2017.06.017

[26] D. Gagliardi, et al., "Estimation of the effective bone-elasticity tensor based on μCT imaging by a stochastic model. A multi-method validation." European Journal of Mechanics-A/Solids, vol. 69, pp. 147-167, 2018, doi: 10.1016/j.euromechsol.2017.10.004

[27] Y. Li, et al., "Characterization of macropore structure of Malan loess in NW China based on 3D pipe models constructed by using computed tomography technology." Journal of Asian Earth Sciences, vol. 154, pp. 271-279, 2018, doi:10.1016/j.jseaes.2017.12.028

[28] Q.L. Yu, et al., "Transverse phase space reconstruction study in Shanghai soft X-ray FEL facility." Nuclear Science and Techniques, vol. 29, no. 1, 2018, doi: 10.1007/s41365-017-0338-0

[29] J.B.Heymann, "Tomographic Reconstruction from Electron Micrographs." Cellular Imaging. Springer, Cham, pp. 209-236, 2018, doi: 10.1007/978-3-319-68997-5_8

[30] S. Shentu, et al., "Maximum entropy method for ocean acoustic tomography." Signal Processing, Communications and Computing (ICSPCC), 2017 IEEE International Conference on. IEEE, 2017, doi: 10.1109/ICSPCC.2017.8242449

[31] M.A. Petrovici, C. Damian, D. Coltuc, "Image reconstruction from incomplete measurements: Maximum Entropy versus L1 norm optimization." Signals, Circuits and Systems (ISSCS), 2017 International Symposium on. IEEE, 2017, doi: 10.1109/ISSCS.2017.8034886

[32] H. Costin, S. Bejinariu, D. Costin, "Biomedical Image Registration by means of Bacterial Foraging Paradigm", International Journal of Computers, Communications & Control (IJCCC), vol. 11, no. 3, pp. 329-345, 2016, doi: 10.15837/ijccc.2016.3.1860

[33] G. Steidl, T.Teuber, "Removing multiplicative noise by Douglas-Rachford splitting methods", Journal of Mathematical Imaging and Vision, vol. 36, no.2, pp.168-184, 2010, doi:10.1007/s10851-009-0179-5

[34] G.Aubert, J.F. Aujol, "A variational approach to removing multiplicative noise", SIAM Journal on Applied Mathematics, vol.68, no.4, pp.925–946, 2008, doi:10.1137/060671814