Maximum-likelihood estimation of an astronomical image from a sequence at low photon levels

Mireille Guillaume, Pierre Melon, and Philippe Réfrégier
Laboratoire Signal et Image, Ecole Nationale Supérieure de Physique de Marseille, Domaine Universitaire de Saint-Jérôme, 13 397 Marseille Cedex 20, France

Antoine Llebaria
Laboratoire d’Astronomie Spatiale, Centre National de la Recherche Scientifique, BP 8, traverse du Siphon, 13 376 Marseille Cedex 12, France

Received January 5, 1998; revised manuscript received May 15, 1998; accepted July 8, 1998

We consider the problem of estimating one nonblurred and cleaned image from a sequence of \( P \) randomly translated images corrupted with Poisson noise. We develop a new algorithm based on maximum-likelihood (ML) estimation for two unknown parameters: the reconstructed image itself and the set of translations of the low-light-level images. We demonstrate that the ML reconstructed image is proportional to the sum of the low-light-level images after correcting for the unknown movement and that its entropy is minimal. The images of the sequence are matched together by means of an iterative minimum-entropy algorithm, where a systematic search under displacements for the images is performed. We develop a fast version of this algorithm, and we present results for simulated images and experimental data. The probability of good matching of a low-level image sequence is estimated numerically when the light level of the images in the sequence decreases, corresponding to small numbers of photons detected (down to 20) in each image of the sequence. We compare these results with those obtained when the low-light-level images are matched to a known reference, i.e., the linear correlation method, and with those from the optimal one, when the noise has a Poisson distribution. This approach is applied to astronomical images that are acquired by photocounting from a balloon-borne ultraviolet imaging telescope. © 1998 Optical Society of America [S0740-3232(98)00611-5]

OCIS code: 100.0100.

1. INTRODUCTION

The reconstruction of images from noisy and blurred data is a classical problem in image processing. We study the case of correcting for both Poisson noise and blur, which are conflicting effects in the measurement of images when the scene and the detector are in relative movement. Indeed, when the time acquisition for one image increases, the Poisson noise decreases as the blur increases. During the last few years, many papers have dealt with scene matching and photon-limited images, particularly in the field of astronomical imaging and medical imaging.

Northcott et al.\(^6\) and Ayers et al.\(^5\) have studied methods based on the triple correlation or the bispectrum technique to restore images degraded by atmospheric turbulence, resulting in variations of the optical transfer function. These techniques rely on taking a series of measurements with sufficiently short exposure times to freeze the effects of atmospheric turbulence. The modulus of the Fourier transform of the reconstructed image is estimated with the speckle interferometry technique, and triple correlation is used to obtain the phase of the Fourier transform of the reconstructed image. Although the low-light-level images are corrupted by Poisson noise, the triple correlation and bispectrum techniques are not optimal in regard to the statistical description of these images.

We consider here the case where the optical transfer function does not change during the measurements and where only translations have to be corrected. We research the optimal estimation of the reconstructed image in the maximum-likelihood (ML) sense.

The ML estimation of the translation between two images is determined by the location of the maximum of their linear intercorrelation when the observed images are corrupted by Gaussian additive noise. It has been shown that taking into account the precise nature of the statistics can greatly improve performance. In some cases the reference image is known, and the problem reduces to that of matching two images. In Ref. 6 photon-limited images are cross correlated with a reference scene, but this method is not optimal when the noise has a Poisson distribution. Recently, we have developed an optimal algorithm to match one low-light-level image with a given reference.

The case for which the reference is unknown has been considered by Slocumb and Snyder,\(^5\) who have considered the estimate of the translation of one image with quantum-limited position sensing. This work is closely related to the one that we address in this paper, but they implement an expectation-maximization algorithm for ML estimation of the displacements. Schulz and Snyder\(^4\) also consider the problem of image reconstruc-
tion from quantum-limited data, but they do not calculate the unknown displacements, whereas we do, and they also implement an expectation-maximization algorithm for maximizing the likelihood function. Many papers in the field of restoration of astronomical or medical images with Poisson noise use this iterative algorithm to obtain the maximum of the likelihood function when the reference image is not accessible. For example, the Richardson–Lucy algorithm has been extensively used in the last few years to reconstruct blurred images corrupted with Poisson noise, especially in the context of Hubble Space Telescope experiments.

We present here a new method based on the focal corrector anastigmat astronomical experiment. In the experiment the images are acquired from a balloon-borne photocounting system. The final image is built from the list of addresses of photoevents, which are delivered in temporal information is available in the film frames, and it is possible to determine a sequence of low-level images ranging from 20 ms to time unit, negligible because the gondola is suspended by a cable of 100-m length, and we analyze the movement that is due to translations in the plane perpendicular to the cable. The log likelihood is then

\[ l(r(i), J) = \sum_{p=1}^{P} \left\{ -\frac{\Delta T}{T} r(i) + s_p(i + j_p) \right\} \times \ln \left\{ \frac{\Delta T}{T} r(i) \right\} - \ln[s_p(i + j_p)] \right\}. \] (4)

The ML estimate \( r_{\text{ML}} \) is obtained by

\[ \frac{\partial l(r(i), J)}{\partial r(i)} = 0, \] (5)

and a simple calculation leads to

\[ r_{\text{ML}}(i, J) = \frac{T}{P \Delta T} \sum_{p=1}^{P} s_p(i + j_p). \] (6)

Since no a priori knowledge about the unknown translations \( J \) is available, we choose to consider the ML estimation of this parameter. It has been proven that if the reference image is known, the optimal estimation of \( j_p \) is obtained by maximizing

\[ \delta l(j_p) = \sum_{i=1}^{N} s_p(i + j_p) \ln[r(i)]. \] (7)

When the reference \( r \) is not known, one must consider the whole sequence of \( P \) noisy and shifted images. We determine the ML estimation \( J_{\text{ML}} \) of \( J \):

\[ l(r, J) = \sum_{p} \sum_{i} \ln[P(s(i)/r, j_p)], \] (8)

\[ J_{\text{ML}} = \arg \max_{J} \{l(r, J)\}. \] (9)

In that situation, because the value of \( r \) is unknown, one classical approach is to implement an expectation-
minimization algorithm to find the value of $\bar{J}$ that maximizes the likelihood function $l(r, J)$ (see Refs. 15, 16, and 23). However, in the considered experiment the temporal information has been preserved during the acquisition, and the unknown image $r$ can be replaced by its estimate $r_{\text{ML}}$ in Eq. (9), giving

$$J_{\text{ML}} = \arg \max_J \left( \sum_i \sum_p s_p(i + j_p) \ln[r_{\text{ML}}(i, J)] - r_{\text{ML}}(i, J) - \ln[s_p(i + j_p)!] \right).$$

The second and third terms on the right-hand side of the equation are independent of $\bar{J}$ because of the assumption of periodicity. This assumption is made to simplify the mathematical development and, as we consider here correction of small translations (up to 25-pixel width for $256 \times 256$ images), it fits well with the observed data. Indeed, on account of the star-tracker acquisition system, there is no luminous object located at the boundaries of the image. However, for large translations, this hypothesis is not fulfilled. In that case these corrective terms should be considered for the determination of $J_{\text{ML}}$.

Under the hypothesis of periodicity, the vector $\bar{J}_{\text{ML}}$ is found to be

$$\bar{J}_{\text{ML}} = \arg \min_J \left( -\sum_i r_{\text{ML}}(i, J) \ln[r_{\text{ML}}(i, J)] \right).$$

The expression to be minimized is

$$E[r_{\text{ML}}(\bar{J})] = -\sum_i r_{\text{ML}}(i, \bar{J}) \ln[r_{\text{ML}}(i, \bar{J})]$$

and represents the entropy of the image $r_{\text{ML}}(i, \bar{J})$.

We have thus demonstrated that the ML estimation of the reference image $r(i)$ is the image obtained by considering the translation values between the low-photon images that minimize the entropy. One can remark that the minimum-entropy image is also the most ordered image, since entropy is a measure of disorder. However, up to now, no explicit solution has been found, and Eq. (11) will be resolved iteratively.

We can note that there is no problem of regularization for the ML estimation. Indeed, in Eq. (11), when the estimated reference $r_{\text{ML}}(i, J)$ has value zero at pixel $i$ (no photon detected), the value of $r_{\text{ML}}(i, J) \ln[r_{\text{ML}}(i, J)]$ is chosen equal to zero, and therefore there is no possible divergence for the ML estimation. Furthermore, the entropy criterion is stable, since limited variations of the observed images $s_p(i)$ lead to limited variation of the entropy.

3. FAST ALGORITHM FOR THE RECONSTRUCTION OF $r(i)$

We consider the problem of finding the vector of translations $\bar{J} = (j_1, \ldots, j_p)$ such that the entropy of the sum image $\Sigma_p s(i + j_p)$ is minimum. We implement an iterative algorithm of systematic research over the allowed displacements for each image.

At step $k$ the estimated value for vector translation is $\bar{J}^k = (j_{1}^{k}, j_{2}^{k}, \ldots, j_{p}^{k})$. The estimated value of the reference is $r(i, J^k)$, and the entropy is $E[r(J^k)]$. If each image $p$ of the sequence, we calculate $E[r(J_{1}^{k}, j_{2}^{k}, \ldots, j_{p}^{k} - \Delta j_1, j_{2}^{k+1}, \ldots, j_{p}^{k})]$ for all values of $\Delta j$ within a window of $m \times m$-pixel width. The next value, $j_{p+1}^{k} = j_{p}^{k} - \Delta j$, is the one for which $E[r(J_{1}^{k}, j_{2}^{k}, \ldots, j_{p}^{k} - \Delta j_1, j_{2}^{k+1}, \ldots, j_{p}^{k})]$ is minimum. The iterations stop when $E[r(J^k)] = E[r(J^{k+1})]$ after one presentation of the whole set of low-photon-level images.

However, this iterative algorithm is very time consuming. Indeed, for $2000$ images of size $256 \times 256$ pixels, with a $5 \times 5$-pixel-width search window, and for ten iterations on the whole sequence, the total computation time is $40$ h with a Sun Sparc 10 workstation. Because of the high number of images in one sequence, large memory capacity is also necessary.

We take advantage of the low photon level (only few pixels have nonzero value in each low-level image), and we develop a fast algorithm inspired by the Nieto-Lebaria algorithm,21 in which the variation of $E[r(J^k)]$ is determined on tables of photoevent addresses rather than on images. For example, in the experimental conditions described below, the computation time can be reduced from 40 h for the direct calculation to 5 min with the fast algorithm. We consider the variation of entropy that is due to the contribution of one image $s_p$. Let us introduce $A_p^k(i) = \Sigma_{j \in \mathcal{J}_p} s_p(i + j_p)$, which is the sum of the $P - 1$ other shifted images of the sequence at step $k$ of the iterative process.

We have implemented the following algorithm for a sequence of $P$ images (see Appendix A):

- Initial value of the estimated reference image is $r^0(i) = \Sigma_{p \neq 1} s_p(i)$, and initial entropy is $-\Sigma_{p \neq 1} s_p(i) \ln r^0(i)]$.
- Choose the size $m \times m$ of the search window.
- For each step $k$:
  1. Calculate $A_p^k(i) = r^k(i) - s_p(i + j_p)$.
  2. For each $\Delta j$ included in the exploration area, calculate the variation: $\text{Var}(k, P) = \sum_{i \in \mathcal{S}_p} [A_p^k(i) + s_p(i + j_p(\Delta j))] \times \ln [A_p^k(i) + s_p(i + j_p - \Delta j)] - A_p^k(i) \ln [A_p^k(i)]$, where $\mathcal{S}_p$ is the set of pixels having a nonzero value in the image $s_p(i + j_p - \Delta j)$.
  3. Choose the value $\Delta j_{\text{opt}}$ that maximizes $\text{Var}(k, P)$, and set $j_{p+1}^{k} = j_{p}^{k} - \Delta j_{\text{opt}}$.

Note the following:

1. Since $\text{Card}(\mathcal{J}_p) \ll N$, we obtain a fast algorithm.
2. In practice, for all the performed simulations, a convergence for $E(\bar{J})$ is always attained with fewer than ten iterations (i.e., ten presentations of all the images).
3. The size $m \times m$ of the search window can be adapted to the amplitude of the translations to avoid some local minima of $E(\bar{J})$. 

Guillaume et al.
4. RESULTS FOR SIMULATED IMAGES
To analyze the optimal processor of Eq. (11), we have generated sequences of translated images corrupted with Poisson noise. The reference image for this generation is a simulated sky with a uniform distribution of stars having the same size (equal to the point spread function of the optical instrument) and intensities distributed according to a realistic law. The reference image and one noisy image are shown in Fig. 1. The displacement of each image is randomly generated according to the probability law \( P(j_p) \).

The parameters for this simulation are as follows:
1. The temporal coefficient for the generation of Poisson noise, \( \lambda = \Delta T/T \), which is proportional to the average number of photons in one low-light-level image. In practice, \( T \) is kept constant, and \( \Delta T \) varies.
2. The density probability law \( P(j_p) \) for the displacement \( j_p \) of one image. Actually, when searching for the ML estimation of the displacements, we make the implicit assumption that the density probability of the translation \( j_p \) is uniform. Indeed, if we know the density probability of the displacements, we can improve the proposed algorithm by using this information in a maximum a posteriori approach. However, it is interesting to test the efficiency of the ML estimation for nonuniform displacements.
3. The number \( P \) of images constituting the sequence. We have generated noisy images with various values of these parameters, and we have applied the processor of Eq. (11).

During all presented simulations, and according to the data obtained with the focal corrector anstigmat experiment, we study only a small translation correction (up to 10-pixel width for \( 256 \times 256 \) images).
A. Evolution of Entropy

We study in this subsection the convergence of entropy. \( E_{\text{calc}}^k \) and \( E_{\text{true}} \) are, respectively, the calculated entropy at step \( k \) of the algorithm and the entropy of the reference \( r \). \( E_r^k = E_{\text{calc}}^k / E_{\text{true}} \) is the reduced entropy of the sequence (\( E_{\text{calc}}^k \) and \( E_{\text{true}} \) are negative; thus \( E_r^k \) is positive and must be maximized to minimize \( E_{\text{calc}}^k \)).

We show in Fig. 2 the evolution of \( E_r^k \) relative to the number of iterations \( k \) for many values of the noise parameter \( \lambda \) and for \( P = 2000 \) images. For all values of \( \lambda \), a convergence is attained with a few iterations on the whole sequence. The true value of the reference image entropy is reached for values of \( \lambda \) larger than \( 5 \times 10^{-4} \) (corresponding to images with more than 140 photons in the mean). For smaller values of \( \lambda \), there are not enough photons in the images to match the sequence.

In Fig. 3 we show the evolution of \( E_r^k \) for translations \( j_p \) generated according to uniform and Gaussian density probabilities \( P(j_p) \) for two standard deviation values and for no displacement. We conclude that for small translations the convergence value of the entropy does not depend on the precise form of \( P(J) \) but that the delay of convergence increases with standard deviation of the translations.

In Fig. 4 the values of convergence for the entropy have been reported, relative to the number of images \( P \) in the sequence, for various values of \( \lambda \). We can see that if \( \lambda \) is greater than \( 5 \times 10^{-4} \) (corresponding to 140 photons in one image of the sequence) and if \( P \) is greater than 2000, the value of the entropy \( E_{\text{true}} \) is reached or is well approximated.

B. Evaluation of the Reconstructed Image

In this subsection we assess the quality of the reconstruction by using two empirical criteria: the probability of correct matching of a low-light-level image and the standard deviation error in the reconstructed image. Of course, these measurements are done on simulated sequence images so as to compare the estimated values of \( \bar{J} \) and \( r \) with the true ones.
In Fig. 5 the probabilities of perfect matching of the images are estimated, relative to the number of images \( P \) of the sequence, for various values of \( \lambda \). These probabilities have been estimated with five generations of \( P \) shifted and noisy sequences by calculating the proportion of exact (zero-pixel-error) determinations of \( j_p \). We note that the number of images \( P \) in the sequence must be greater than a certain value (which is here between 2000 and 5000) to obtain significant results. On the contrary, if \( P \) increases over this value, there is not much gain. The probability of perfect matching (zero pixel error) depends strongly on the number of photons present in each image. 280 photons per image are necessary to obtain 80% perfectly matched images.

In Fig. 6 the probabilities of perfect matching are calculated relative to \( \lambda \) for many values of \( P \). We have compared these probabilities with those obtained with the optimal algorithm of Eq. (7) and with the linear correlation algorithm when the reference is known. For sequences of more than 1000 images, the minimum-entropy algorithm is more efficient than linear correlation with known reference.

To estimate a measurement error in the reconstructed image itself, we have calculated the error standard deviation \( \sigma \) that is due to a false estimation of \( \bar{j}_p \):

\[
\sigma = \left( \frac{1}{P} \sum_{p=1}^{P} (\Delta j_p)^2 \right)^{1/2},
\]

where \( \Delta j_p = j_p - j_p^{ML} \), \( j_p \) being the real value of translation. The result is shown in Fig. 7. When the sequence contains more than 5000 images, the error standard deviation depends only on the temporal parameter \( \lambda \) for each image \( s_p \). For more than 140 photons per image, we obtain less than 1-pixel standard deviation error.

The reconstructed image is shown in Fig. 8(a) without correction and in Fig. 8(b) after position correction with 224 photons per image in the mean and a 5000-image sequence.

5. RESULTS FOR EXPERIMENTAL DATA

We show in Figs. 9(a) and 9(b) the experimental data Messier 3 from the focal corrector anastigmat experiment before and after the proposed algorithm is applied, respectively. A list of 609,879 photoevents allows the reconstruction of a 1024 × 1024 image. In Fig. 10 the section of the intensity of the most luminous star is shown for no correction and for correction with low-level images made with 250 and 25 photons, corresponding to 2429 and 24,297 images, respectively. There is no improvement in making sequences with fewer than 125 photons per image. However, for this 1024 × 1024 image, the result does not deteriorate when the number of photons decreases. In particular, no ringing effects can be seen, as is the case for linear correlation with known reference.

6. CONCLUSION

In this paper we have presented the maximum-likelihood estimation of an astronomical image when temporal information is available in the collected data. The reconstructed image is obtained by optimal matching of a sequence of high-temporal-resolution images. For small translations of the sensor, this matching is obtained when the entropy of the reconstructed image is minimum. A fast algorithm is developed to estimate the minimum-entropy reconstructed image. We have shown that for all the performed simulations and for experimental data, the minimum entropy is attained in a few minutes with a classical workstation, and that when the number of photons and the number of images are sufficient (more than 150 photons per low-photon-level image and more than 2000 images in the simulations), and for small translations, the reconstructed image has entropy close to the entropy of the reference.

The error standard deviation is smaller than 1 pixel for the reconstructed image, and the result obtained with the minimum-entropy method and unknown reference is more efficient than that from the classical linear correlation algorithm with known reference.
The presented processor has been developed under the assumption of small translations for the sequence. This assumption has two consequences: The first is on the choice of the presented search algorithm for the minimum entropy. The case of large-translation corrections could be implemented with a stochastic algorithm to avoid local minima. The other consequence is that we can assume that the images are periodic, which is necessary to establish the minimum-entropy algorithm. For large-translation corrections, if this hypothesis were no longer valid, additive terms would have to be included in the algorithm in the optimization process.

These results have been obtained without any a priori knowledge about the translations. They open up interesting perspectives for developing an optimal algorithm when some knowledge of the translations is available, in which case the maximum a posteriori estimate should be used. This topic is currently the subject of investigation in our laboratory.

APPENDIX A

Let

\[ E[r(J^k)] = -\sum_{i=1}^{N} [A_p^k(i) + s_p(i + j^k_p)] \]

\[ \times \ln[A_p^k(i) + s_p(i + j^k_p)], \]

\[ E[r(\overline{J}^k)] = -\sum_{i=1}^{N} [A_p^k(i) + s_p(i + j^k_p - \Delta j)] \]

\[ \times \ln[A_p^k(i) + s_p(i + j^k_p - \Delta j)], \]

which can also be written as

\[ E[r(\overline{J}^k)] = -\sum_{i \in D_p} A_p^k(i) \ln[A_p^k(i)] \]

or

\[ E[r(\overline{J}^k)] = -\sum_{i \in D_p} [A_p^k(i) + s_p(i + j^k_p - \Delta j)] \]

\[ \times \ln[A_p^k(i) + s_p(i + j^k_p - \Delta j)] \]

\[ - \sum_{i=1}^{N} A_p^k(i) \ln[A_p^k(i)] + \sum_{i \in D_p} A_p^k(i) \ln[A_p^k(i)], \]

and thus \( E[r(\overline{J}^k)] = E(A_p^k) - \text{Var}(k, p), \) where

\[ \text{Var}(k, p) = \sum_{i \in D_p} [A_p^k(i) + s_p(i + j^k_p - \Delta j)] \]

\[ \times \ln[A_p^k(i) + s_p(i + j^k_p - \Delta j)] \]

\[ - A_p^k(i) \ln[A_p^k(i)]. \]
Since
\[
\Delta_{j^{\text{opt}}} = \arg \min_{\Delta_j} \{E[r(\hat{J}_{j^{\text{opt}}})]\} = \arg \max_{\Delta_j} \{\text{Var}(k, p)\},
\]
it is equivalent to maximize \(\text{Var}(k, p)\) or to minimize \(E[r(\hat{J}_{j^{\text{opt}}})]\) over \(\Delta_j\).

The advantage of calculating \(\text{Var}(k, p)\) is that the number of pixels in \(D_0\) is in the range of 100 instead of 65,000 for \(256 \times 256\) images.

REFERENCES