GEMINI
8-M Telescopes
Project

TN-I-G0054

Self-Learning Bayesian Centroid Estimation

N. Dillon and C. R. Jenkins
Royal Greenwich Observatory

August 6, 1997

1. ABSTRACT

CCD read noise is the single most important factor which determines the performance of the Shack-Hartmann wavefront sensor in photon-starved applications. We address the problem of making optimum centroid estimates in sensors employing NxN-pixel centroiding geometries, where N is generally > 2 so that read noise is even more important than in quad-cell-based sensors. Maximum-likelihood and Bayesian estimators are derived and we show that these afford excellent noise suppression whilst relaxing the constraints on alignment tolerances and static aberrations which are demanded in quad-cell applications. The estimators considered are all linear and are shown to be implementable using conventional real-time processing hardware.

Keywords: Shack-Hartmann, wavefront sensor, centroid, adaptive optics, Bayesian analysis, maximum likelihood.

2. INTRODUCTION

Practical Shack-Hartmann wavefront sensors (SHWFS) typically require detector arrays of at least 32x32 pixels and for this reason invariably use CCDs as the detective elements. The performance of such sensors when applied to astronomical adaptive optics (AO) is largely governed by read noise introduced by the CCD - typical noise levels being in the range 2-10 e^-/pixel over a relevant range of frame rates. Despite this limitation, the SHWFS has been shown to be competitive with the zero-read-noise curvature sensor provided quad-cell centroiding is used to determine wavefront slopes within each sub-aperture[1]. There are applications, however, for which the SHWFS' sub-images cannot be guaranteed to fall on quad-cell pixel boundaries. These include: (i) sensors which experience non-common-path aberrations in the reference and corrected wavefronts and (ii) open-loop sensors used for atmospheric characterisation; we refer to this mode of operation as "non-nulled". The Gemini telescope's Peripheral WFS are an example of the former case[2] and provided the original motivation for this investigation. These sensors patrol an area which lies 3-7′ off the telescope's optical axis and the Ritchey-Chrétian telescope optics therefore introduce varying amounts of astigmatism into the reference wavefronts. It should be noted that even an on-axis sensor will generally experience some degree of non-common-path aberration due to imperfect alignment of the system.

The importance of non-nulled operation is that the quantity measured by a quad cell is then sensitive to the shape of the PSF within each sub-aperture. Any variations in the PSF, e.g. due to atmospheric fluctuations, will introduce noise into the wavefront slope estimates and a particular concern is that uncertainty in the PSF will generally lead to a <u>systematic bias error</u> so that the wavefront error does not average to zero. Non-nulled operation also reduces the linear dynamic range of the SHWFS. These effects are mitigated by the use of a finer pixel scale and correspondingly larger number of pixels to sample the PSF but this approach exacerbates the effect of CCD read noise.

We are concerned here with the problem of making non-nulled centroid estimates - using non-quad-cell detector geometries - in which the propagation of CCD read noise into the centroid estimate is minimised. In section 2 we derive a classical (non-Bayesian) linear centroid estimator which attempts to make optimum use of the available data. The mathematical framework introduced here is used as the basis for derivation of a Bayesian estimator in section 3, where the adaptation of this estimator to track changing atmospheric conditions is also described. In section 4 we investigate the implementation of the Bayesian estimator including consideration of the real-time load on the wavefront processor. The results of analytical calculations and numerical simulation of the various estimators considered are presented in section 5 along with preliminary results from experimental data.

3. NON-BAYESIAN ESTIMATION

Generalisation of the standard quad-cell centroiding formula[3] leads to a simple estimator for the geometric centroid, or moment, of an image

$$\underline{\hat{c}} \propto \sum_{\text{pixels}} \underline{r_n} d'_n \tag{1}$$

where $\underline{r_n}$ are the coordinates of the centroid (normally the centre) of each pixel and d'_n the photon counts in each pixel normalised by the total count, the constant of proportionality is generally included to correct for the non-unity gain of the basic estimator. The estimate thus obtained has the virtue that it requires no knowledge of the image profile but it is clear that those data at large $|\underline{r_n}|$ make the largest contribution to the centroid whilst, for typical PSFs, these data have the lowest SNR. The noise propagation of this estimator is therefore very poor. Data-thresholding can improve the moment centroid but we are concerned here with estimators which make best use of the available data - subject to an appropriate definition of what we mean by the term “best”.

3.1 Maximum Likelihood Estimator

We begin by deriving a Maximum-Likelihood (ML) non-Bayesian estimator. The derivation of this serves to highlight an important approximation which is required for all linear model-fitting centroid estimators, and, as we shall see later, it turns out that the Bayesian estimator is closely related to the ML solution.

The first step is to adopt, a priori, a model for the PSF (distribution of irradiance) within each sub-aperture $\mathrm{M}(\underline{c},\underline{r})$ where $\underline{r}$ locates a point in the focal plane of the SHWFS and $\underline{c}$ contains the parameters of the model that we wish to solve for. The parameters of fundamental interest are the image centroids but we address the general case in which the sub-apertures are not diffraction limited so that image width is also unknown and must be treated as a free parameter. An appropriate choice of parameters is then $\underline{c} = [x, y, z]^{\mathrm{T}}$ where x, y are the image position and z some measure of image width. We could, for example, choose to let z represent the FWHM of the PSF but we shall see later that it is convenient to solve for $z = \log(\omega - \omega_{\min})$ where ω and $\omega_{\min}$ are the PSF’s FWHM and a lower-limit, e.g. due to diffraction, on the FWHM. The model predicts a photon count

$$s_n(\underline{c}) = S_{tot} \int_{\text{pixel area}} \mathrm{M}(\underline{c},\underline{r})\,\mathrm{d}\,A \tag{2}$$

within the n^th^ pixel (where S_{tot} is the total photon count), but we observe noisy data $d_n = s_n + \varepsilon_n$. In most AO applications, gaussian CCD read noise dominates over photon noise so that $\mathrm{p}(\varepsilon_n)$ can be approximated by a zero-mean gaussian with variance $\sigma_n^2 = \sigma_{CCD}^2 + \langle s_n \rangle$, where we have assumed that image motion is sufficiently small that the distribution of photon noise over the detector is roughly time-invariant (the estimator derived below is easily adapted to track medium-term variations in noise distribution if required). The likelihood of observing the ensemble of pixel values within a sub-aperture is therefore

$$\mathrm{p}(\underline{d}|\underline{c}) = \prod_{\text{pixels}} \mathrm{p}(d_n|\underline{c}) \quad \text{where} \quad \mathrm{p}(d_n|\underline{c}) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left\{\frac{-[d_n - s_n(\underline{c})]^2}{2\sigma_n^2}\right\} \tag{3a,b}$$

A Maximum Likelihood estimate for the model parameters is obtained in the conventional way by maximising the log likelihood, this corresponds to finding the least-squares solution:

$$\sum_{\text{pixels}} \frac{[d_n - s_n(\underline{c})]}{\sigma_n^2} \underline{\nabla} s_n(\underline{c}) = \underline{0} \tag{4}$$

The solution of Eq. 4 is complicated by the fact that for all PSFs of physical interest the model predicts photon counts which are non-linearly dependent upon $\underline{c}$. A non-linear (iterative) numerical solution is impractical in a real-time AO system so we need a linearised approximation to the solution. This is obtained by taking a Taylor expansion of $s_n(\underline{c})$ about a reference point which we choose to be the average value $\underline{\mu_c} = \langle \underline{c} \rangle$. In most applications we should expect $\underline{\mu_c}$ to be known to reasonable accuracy so that we can now transform to a new set of unknowns $\underline{\Delta c} = \underline{c} - \underline{\mu_c}$ where $\underline{\Delta c}$ represent small real-time variations in the model parameters about the mean values and $s_n(\underline{c}) \approx s_n(\underline{\mu_c}) + \underline{\nabla} s_n(\underline{\mu_c}) \cdot \underline{\Delta c}$. Eq. 4 can then be written in the compact matrix form

$$\mathbf{A}\underline{\Delta\hat{c}} = \underline{b} \tag{5a}$$

where $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ and $\underline{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

in which the coefficients $a_{ij} = \sum_{\text{pixels}} \frac{1}{\sigma_n^2} \left.\frac{\partial s_n}{\partial u_i}\right|_{\underline{\mu_c}} \left.\frac{\partial s_n}{\partial u_j}\right|_{\underline{\mu_c}}$, $b_i = \sum_{\text{pixels}} \frac{[d_n - s_n(\underline{\mu_c})]}{\sigma_n^2} \left.\frac{\partial s_n}{\partial u_i}\right|_{\underline{\mu_c}}$ (5b,c) with $u_1 = x$, $u_2 = y$ and $u_3 = z$. The hat denotes that the solution of these so-called "normal equations" $\underline{\Delta\hat{c}}$ is an estimate of the unknowns $\underline{\Delta c}$. Under our assumption of gaussian errors the ML solution $\underline{\hat{c}} = \underline{\mu}_c + \mathbf{A}^{-1}\underline{b}$ is a least-squares-error fit of the model PSF to the data and the errors on the fitted centroids and image-width also have gaussian distributions with variances given by the diagonal elements of the covariance matrix $\mathbf{C} = \mathbf{A}^{-1}$.

4. BAYESIAN ESTIMATION

The ML estimator described above seeks to maximise the probability of model parameters x, y, z given observations $\underline{d}$. This approach makes no assumptions about the likelihood of a given set of parameters occurring and we should expect to do better by constraining the estimate with prior knowledge of the intrinsic statistical variations displayed by x, y, z. This is the essence of Maximum Aposteriori (MAP) likelihood estimation. The basis for the technique is Bayes' theorem which states that the posterior probability of model parameters $\underline{c}$ given $\underline{d}$ (and an implicit model M) is given by

$$\mathrm{p}(\underline{c}|\underline{d}) = \frac{\mathrm{p}(\underline{d}|\underline{c}).\mathrm{p}(\underline{c})}{\mathrm{p}(\underline{d})}. \tag{6}$$

Here $\mathrm{p}(\underline{d}|\underline{c})$ is the likelihood function (Eq. 3a) and $\mathrm{p}(\underline{c})$ the prior probability of $\underline{c}$; the inclusion of this term is the important distinction between the ML and MAP approaches. $\mathrm{p}(\underline{d})$ is a normalising factor which is constant with respect to $\underline{c}$ and can therefore be ignored for our purposes. Derivation of the MAP estimator involves solving for the maximum of Eq. 6 following an approach similar to that used previously. We provide the details of this solution in section 4.2 but it is appropriate to first consider how we intend to apply classical Bayesian inference to this problem, and, specifically, how we propose to obtain the prior information which is central to the technique.

Our basic approach is to assert the independent priors $\mathrm{p}(x)$, $\mathrm{p}(y)$ and $\mathrm{p}(z)$ and use these to compute $\mathrm{p}(\underline{c}) = \mathrm{p}(x)\,\mathrm{p}(y)\,\mathrm{p}(z)$. The important point is that we cannot assume the priors to be time-invariant functions which may be known absolutely. Significant variations in atmospheric seeing are known to occur on timescales of only a few minutes[4] and it is clear that the long-term distribution of seeing parameters (which applies over periods of several days or weeks) will be appreciably broader, and therefore impose weaker constraints, than that obtained over a period of a few seconds or minutes. We therefore adopt a constant functional form for the prior but introduce additional parameters $\underline{w} = \left[w_{1,}, w_2, .., w_M\right]$ to characterise the underlying distributions at any instant so that the prior becomes $\mathrm{p}(\underline{c}|\underline{w})$. $\underline{w}$ may typically contain quantities such as the rms image motion and average image width etc. We must therefore address two distinct optimisation problems. Firstly, on the arrival of each SHWFS image frame we need an estimator which maximises the posterior likelihood (Eq. 6) with respect to unknown image parameters $\underline{c}$ and known prior $\mathrm{p}(\underline{c}|\underline{w})$ - this is the MAP estimator that we derive below. The second problem, which occurs on longer timescales over which the atmosphere is expected to display small variations (e.g. tens of seconds to a few minutes), is to update the adopted prior $\mathrm{p}(\underline{c}|\underline{w})$ with a new estimate of $\underline{w}$. In this way we periodically update the prior such that it tracks variations in the atmospheric seeing in a series of discrete but hopefully small steps. We return to the determination of $\underline{w}$ later following adoption of a specific set of priors and derivation of the MAP estimator.

4.1 Adoption of Priors

A rigorous statistical description of the image parameters (x, y, z) would undoubtedly be difficult to obtain. The problem is compounded by the likely non-stationarity of the processes and interaction of the AO servo loop with intrinsic atmospheric perturbations. Our goal, however, is to achieve a modest degree of noise suppression whilst avoiding the introduction of bias into centroid estimates - as might occur if the priors were excessively narrow or grossly incorrect in their basic form. We therefore proceed with a pragmatic choice of priors based on consideration of the physical phenomena and anticipate that this choice will prove to be adequate in most practical applications.

Atmospheric turbulence is known to lead to gaussian image motion, although motion due to telescope windshake and that induced by the AO system is admittedly less well understood. In the absence of further information we adopt gaussians for the distributions $\mathrm{p}(x)$ and $\mathrm{p}(y)$. Measurements of variations in image width at typical observing sites show that a log-normal distribution provides an accurate fit to the data[4,6,7] with a lower cut-off occurring at some small image width due to the telescope's diffraction limit or other factors which limit the best image quality. It should now be apparent why we chose to adopt $z = \log(\omega - \omega_{\min})$ where $\omega, \omega_{\min}$ are the image width and lower-limit on ω respectively, since we can now adopt a gaussian for $\mathrm{p}(z)$. Our three prior distributions are therefore all gaussian and the parameters which describe the overall prior $\mathrm{p}(\underline{c}|\underline{w})$ are $\underline{w} = \left[\mu_x, \sigma_x^2, \mu_y, \sigma_y^2, \mu_z, \sigma_z^2\right]$ which contains the mean and variance for each element of $\underline{c}$.

4.2 Maximum Aposteriori Likelihood Estimator

Armed with our adopted priors we now proceed to derive the MAP centroid estimator. Following the approach taken in section 3.1, the posterior probability given in Eq. 6 is maximised when

$$\underline{\nabla} \log[\mathrm{p}(\underline{c})] + \sum_{\text{pixels}} \frac{[d_n - s_n(\underline{c})]}{\sigma_n^2} \underline{\nabla} s_n(\underline{c}) = \underline{0}\,. \tag{7}$$

Now, if the derivatives of the log-probabilities can be expressed in the linear form $\underline{\nabla} \log[\mathrm{p}(\underline{c})] = \underline{\alpha} \cdot \underline{\Delta c} + \underline{\beta}$ then it is easy to show that the normal equations are in this case

$$\mathbf{A}' \underline{\Delta \hat{c}} = \underline{b}' \tag{8a}$$

where $$\mathbf{A}' = \mathbf{A} - \begin{bmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ 0 & 0 & \alpha_3 \end{bmatrix} = \mathbf{A} - \mathbf{I}\underline{\alpha} \text{ and } \underline{b}' = \underline{b} + \underline{\beta} \qquad (8b,c)$$

Eqs. 8 define a general MAP centroid estimator for any given $\underline{\alpha}$, $\underline{\beta}$. We choose to linearise the MAP estimator about $\underline{\mu}_c = [\mu_x, \mu_y, \mu_z]^{\mathrm{T}}$ and substitution of the priors adopted in section 4.1 yields

$$\frac{\mathrm{d}}{\mathrm{d}\,x}\ln[\mathrm{p}(x)] = -\frac{\Delta x}{\sigma_x^2}, \quad \frac{\mathrm{d}}{\mathrm{d}\,y}\ln[\mathrm{p}(y)] = -\frac{\Delta y}{\sigma_y^2}, \quad \frac{\mathrm{d}}{\mathrm{d}\,z}\ln[\mathrm{p}(z)] = -\frac{\Delta z}{\sigma_z^2} \qquad (9a,b,c)$$

These are in the linear form that we need and it can be seen that vectors $\underline{\alpha}$, $\underline{\beta}$ defined above are particularly simple, with $\underline{\alpha} = \left[-\sigma_x^{-2}, -\sigma_y^{-2}, -\sigma_z^{-2}\right]^{\mathrm{T}}$ and $\underline{\beta} = \underline{0}$. This completes the derivation of our MAP centroid estimator. It is instructive to examine the normal equations for the estimator since there is striking similarity with the ML solution. Recall Eq. 8a, $\mathbf{A}'\underline{\Delta\hat{c}} = \underline{b}'$, which must be inverted to solve for the unknowns $\underline{\Delta\hat{c}}$ where we now have

$$\mathbf{A}' = \mathbf{A} + \begin{bmatrix} \sigma_x^{-2} & 0 & 0 \\ 0 & \sigma_y^{-2} & 0 \\ 0 & 0 & \sigma_z^{-2} \end{bmatrix}, \qquad \underline{b}' = \underline{b} \qquad (10a,b)$$

and $\mathbf{A}$, $\underline{b}$ are defined in Eqs. 5 and correspond to the ML estimator. We have chosen to express our results in the above form because it can now be seen that the variances introduced in the MAP estimator show up as additive damping factors along the diagonal elements of $\mathbf{A}$. It is clear that the ML estimator is essentially a special case of the MAP estimator for which the fitted-parameters have unconstrained variances. Note also that, for a symmetric image which is roughly centred within a sub-aperture, matrix $\mathbf{A}'$ is very sparse and the covariance matrix for the MAP solution can be approximated by the diagonal matrix

$$\begin{bmatrix} (a_{11} + \sigma_x^{-2})^{-1} & 0 & 0 \\ 0 & (a_{22} + \sigma_y^{-2})^{-1} & 0 \\ 0 & 0 & (a_{33} + \sigma_z^{-2})^{-1} \end{bmatrix}$$

and recall that the diagonal elements of this matrix give the variances on the respective fitted parameters. A MAP noise reduction factor can be defined as the ratio of the RMS error on a MAP estimate to that on a corresponding ML estimate: $\left(1 + {<\varepsilon^2>}\big/{\sigma^2}\right)^{-\frac{1}{2}}$ where σ^2 and $\langle\varepsilon^2\rangle$ are the intrinsic process variance and mean-square ML-estimation error respectively. It is not surprising that the noise suppression increases as the prior is tightened ($\sigma^2 \to 0$) and as the noise increases.

4.3 Estimation & Adaptation Of The Prior

The learning procedure outlined at the beginning of section 4 enables the prior to track changing atmospheric conditions by updating $\underline{w}$ in a series of discrete steps. However the problem of determining the optimum estimates of $\underline{w}$ from the available data in a mathematically-rigorous way is undoubtedly non-trivial. The problem is that the only observables are raw CCD image data so that our estimate of $\underline{w}$ necessarily involves estimation of $\underline{c}$ as an intermediate step and this raises the question of what estimator should be used. The important point is that the estimates, $\hat{\underline{c}}$, from which we derive $\hat{\underline{w}}$ must be free from bias such that the adopted priors are not artificially constrained - it is clear that the output from the MAP estimator itself cannot be used since $\hat{\underline{w}}$ would then feed-back into itself leading to unstable or trivial solutions. A further complication is that, whatever estimator we use, $\hat{\underline{w}}$ will be subject to measurement noise.

We have not found an optimum strategy for the estimation of $\underline{w}$ which takes account of both measurement noise <u>and</u> non-stationarity in the processes of interest. The approach we have adopted thus far is to employ an ML estimator for $\hat{\underline{c}}$ in parallel with the required MAP estimator and to use the estimates from this to calculate the sample mean and variance for each of the three sequences $\hat{x}_n, \hat{y}_n, \hat{z}_n$ over a number of SHWFS frames. The number of frames over which averaging is performed, L_{update}, must in practice be a compromise between the need to achieve maximum noise suppression (L_{update} large) whilst updating the prior sufficiently quickly to track non-stationarity (L_{update} small). We believe that further analytical progress on the choice of adaptation timescale is possible but note that, in a typical AO system, an averaging period of a few seconds will typically achieve high SNR on $\hat{\underline{w}}$ whilst enabling relatively fast adaptation to changing atmospheric conditions. Our estimates of the parameters of the prior on x are therefore

$$\hat{\mu}_x = \frac{1}{L_{update}} \sum_{i=0}^{L_{update}-1} \hat{x}_{n-i} \quad \text{and} \quad \sigma_x^2 = \left[\frac{1}{L_{update}} \sum_{i=0}^{L_{update}-1} \left(\hat{x}_{n-i} - \hat{\mu}_x \right)^2 \right] - \ <\varepsilon_x^2> \tag{11}$$

where $<\varepsilon_x^2>$ is the mean square error on $\hat{x}_n$ (obtained from the covariance matrix $\mathbf{A}^{-1}$ for the ML estimator) and this term must be subtracted from the observed variance in $\hat{x}_n$ because we are interested in the <u>intrinsic</u> variance in the process x_n. Corresponding estimators are used for the parameters of the priors on y and z.

Finally, recall that the linearsing step which was needed when deriving the ML and MAP estimators required knowledge of the reference point $\underline{\mu}_c = \langle \underline{c} \rangle$. The requirement was that $\underline{\Delta c} = \underline{c} - \underline{\mu}_c$ is always small and in the case of the ML estimator we had no choice other than to adopt a constant $\underline{\mu}_c$ for the linear expansion. A refinement is possible in this case since the adaptation strategy described enables the value of $\underline{\mu}_c$ used in the MAP estimator to be updated periodically thus minimising the effect of the linearising approximation.

5. IMPLEMENTATION CONSIDERATIONS

5.1 Real-Time (Latency-Critical) Processing

There is incentive to minimise the processing load imposed by the centroid estimator in an AO system in order to minimise latency in the control loop. The ML and MAP estimators derived above are both linear and should therefore be expressible as weighted sums of the raw pixel data, $\underline{d}$. We now show that this is indeed the case.

Recall from Eqs. 8 that the centroiding processor must in general compute $\underline{\hat{c}} = \underline{\mu}_c + (\mathbf{A} - \mathbf{I}\underline{\alpha})^{-1}(\underline{b} + \underline{\beta})$ each time a set of N^2 pixel values are read from the CCD. Now, Eq. 5c can be written in the form $\underline{b} = \mathbf{B}\underline{d} - \underline{b}_0$ where $\mathbf{B}$ is a 3xN^2 matrix and $\underline{b}_0$ a 3-vector with coefficients

$$b_{ij} = \frac{1}{\sigma_{\mathrm{j}}^2}\frac{\partial s_j}{\partial u_i}\bigg|_{\underline{\mu}_c} \quad \text{and} \quad b_{0_i} = \sum_{\text{pixels}}\frac{s_n(\underline{\mu}_c)}{\sigma_{\mathrm{n}}^2}\frac{\partial s_n}{\partial u_i}\bigg|_{\underline{\mu}_c} \qquad (i = 1,2,3;\ j = 1,..,N^2) \qquad \text{(12a,b)}$$

which allows the estimator to be expressed in the required form

$$\underline{\hat{c}} = \underline{\mu}'_c + \mathbf{W}\underline{d} \qquad \text{(13a)}$$

where $\underline{\mu}'_c$ and $\mathbf{W}$ are constants given by

$$\underline{\mu}'_c = \underline{\mu}_c + (\mathbf{A} - \mathbf{I}\underline{\alpha})^{-1}(\underline{\beta} - \underline{b}_0) \quad \text{and} \quad \mathbf{W} = (\mathbf{A} - \mathbf{I}\underline{\alpha})^{-1}\mathbf{B} \qquad \text{(13b,c)}$$

Eq. 13a shows that the image centroids are given by weighted sums of the raw CCD pixel data. In a real-time application the offset terms in $\underline{\mu}'_c$ may be pre-loaded into the processor's accumulator before the first data arrive, and in this case the real-time processing load for centroid computation is nominally the same as that for the simple moment estimator given in Eq. 1. It should be acknowledged, however, that Eq. 1 can be re-arranged to take advantage of the symmetry usually present in SHWFS detector arrays and this re-arrangement reduces the number of multiply operations required for the moment centroid by a factor of N. This simplification is not possible for the ML or MAP estimators unless the fitted PSF is nulled, but we note that most practical AO systems employ DSPs in their real-time computers and the single-cycle multiply-accumulate capability of these ensures that the ML and MAP estimators impose effectively no additional processing load when compared to the simple moment estimator. The arithmetic operations required to compute $\hat{c}$ are readily implementable on either fixed- or floating-point processing hardware.

It is important to recognise that although our estimators treat the characteristic image width z as a free parameter, it is not necessary to compute z in order to benefit from its inclusion in the formulation: only the first two elements of $\underline{\hat{c}} = [\hat{x}, \hat{y}, \hat{z}]^{\mathrm{T}}$ in Eq. 13a need be computed to obtain MAP/ML centroid estimates. If a reconstructor were available which exploited the additional information contained in the image width then it may be desirable to compute z in real time, but we are not aware of any such reconstructor.

5.2 Background Processing

The estimators described here require varying amounts of background processing in addition to the real-time operations discussed above. The additional computation load associated with these operations should be taken into account when assessing the viability of the proposed centroiding algorithms.

Both the ML and MAP estimator require the total photon count, S_{tot}, to be determined and we assume this is done by simply summing the pixel values within each sub-image. It is advantageous, however, to employ a simple low-pass filter on S_{tot} (e.g. box-car average over a few image frames) in order to suppress noise. Note also that every time S_{tot} changes the MAP/ML estimator weights need to be re-computed and we therefore assume that hysteresis is used to ensure that these updates are performed relatively infrequently. This approach assumes that scintillation can be neglected.

In the case of the adaptive MAP estimator additional processing is required in the form of a separate ML estimator dedicated to estimation of the parameters of the MAP prior, along with evaluation of the statistical quantities in $\underline{w}$ (Eq. 11 etc) and computation of the optimum MAP estimator weights given in Eqs. 13 every L_{update} frames when the priors are updated. The ML estimator used here should generally compute the estimated image width in addition to image centroids. These operations clearly increases the load on the centroid processor although the ML estimates are not required with low latency and may potentially be computed following the MAP estimates if there is sufficient “idle time” before the arrival of the next SHWFS image frame.

Computation of the estimator weights in response to a significant fluctuation in S_{tot}, or an update of the MAP priors, is a relatively demanding computational task which generally involves the evaluation of transcendental functions in order to compute the matrix elements a_{ij}, b_i (Eqs. 5b,c). Furthermore, a matrix inversion - albeit only a 3x3 matrix - is required to compute the weights $\mathbf{W}$. We suggest that these operations are best suited to floating-point processing hardware rather than a fixed-point implementation. The fact that they are performed relatively infrequently clearly raises the possibility of implementing them as background operations in the centroid processor and in this case the use of a multi-tasking real-time kernel will simplify the implementation. For example, in the case of the fast-guiding sensors under development for Gemini[2], the shortest SHWFS frame period is 5 ms whilst the centroid processor is specified to generate centroid estimates within 1 ms of the end of each integration period in order to maintain optimum servo performance (latency is further degraded by up to 1 ms by other parts of the system). The centroid processor is therefore potentially idle for 80% of the time and our implementation uses this time to perform the background calculations associated with adaptation of the MAP estimator. An alternative approach, which may be attractive for a higher-order AO system, would be to employ a dedicated processor for background operations in parallel with the real-time centroid processor.

6. RESULTS

6.1 Noise Propagation

Noise propagation through the various centroid estimators has been evaluated via a combination of analytical calculation and numerical simulations. We find that idealised calculations which do not take account of the fact that the SHWFS sub-images are constantly moving, and generally have a varying FWHM, can lead to misleading conclusions however we wish to avoid the complexity of a full-blown AO system simulation. A simple model which includes these effects has therefore been adopted as the basis for comparison and this enables analytical results to be obtained in many cases. The sub-image PSF is modelled as a gaussian profile in which the FWHM has a log-normal distribution with $\langle \omega \rangle$ = 0.5″, σ^2_ω = 0.2″ and a diffraction limit at $\omega_{\min}$ = 0.15″. Image motion is modelled as gaussian perturbations in *x*, *y* about some defined $\langle x \rangle, \langle y \rangle$ with FWHM image motion given by a constant factor Γ_{xy} multiplied by the instantaneous image width (the assumption here is that at times of bad seeing overall image motion is relatively large).

Fig. 1 illustrates the sensitivity of the quad cell to non-nulled operation. The centroiding error is plotted against available signal at a fixed CCD noise level for 2x2- and 4x4-pixel moment centroid estimation (Eq. 1, with the constant of proportionality taken to be the reciprocal of the transfer function at the mean image width). The "nulled" case has $\langle x \rangle = \langle y \rangle = 0$ and $\Gamma_{xy} = 0$, whilst the "non-nulled" case has $\langle x \rangle = \langle y \rangle = 0.1''$ and $\Gamma_{xy} = 0.25$. At low signal levels detector noise dominates and the quad cell is clearly the optimum geometry irrespective of whether the PSF is nulled or non-nulled. As the signal increases, however, the quad-cell reveals its sensitivity to non-nulled operation whilst the 4x4 centroid benefits from better sampling of the PSF and is relatively immune against this effect. In this particular example the non-nulled quad-cell performs worse than the 4x4-pixel geometry at a relatively modest signal of around 250 photons/sub-aperture. Note that the limiting error which the quad-cell experiences at high SNR will propagate into the long-term average reconstructed wavefront error.

In Fig. 2 we demonstrate the immunity against CCD read noise afforded by the ML estimator when used for NxN pixel centroiding. The figure shows the RMS centroiding error for 4x4- and 6x6-pixel geometries normalised by that for a quad-cell as a function of CCD noise level. In this case the PSFs were perfectly nulled in order to give maximum credit to the quad-cell and thus obtain a conservative indication of the ML noise suppression. The signal was held constant at 100 photons. At very low noise levels we see a small benefit in all cases over the quad-cell due to the better spatial sampling in these geometries. The moment estimators degrade very rapidly with increasing noise level whilst the ML estimator is clearly much less sensitive to noise, the 4x4-

pixel geometry performing only marginally worse than a quad-cell under conditions which are known to be unduly optimistic with regard to quad-cell performance.

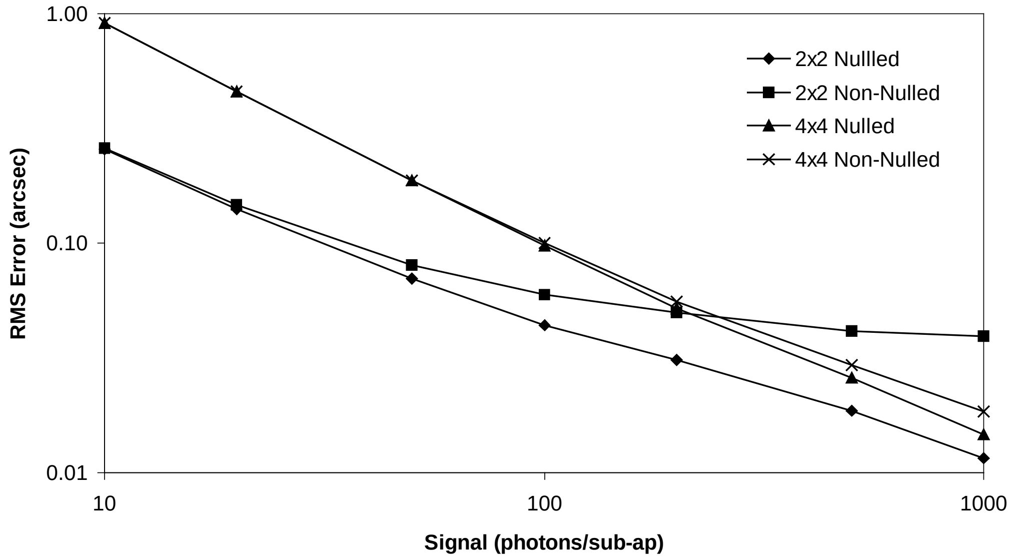

Figure 1. Noise propagation for moment-centroid estimator at a noise level of 5 photons/pixel. The sub-aperture size is 2x2″ in all cases. See the text for definition of the “nulled” and “non-nulled” cases.

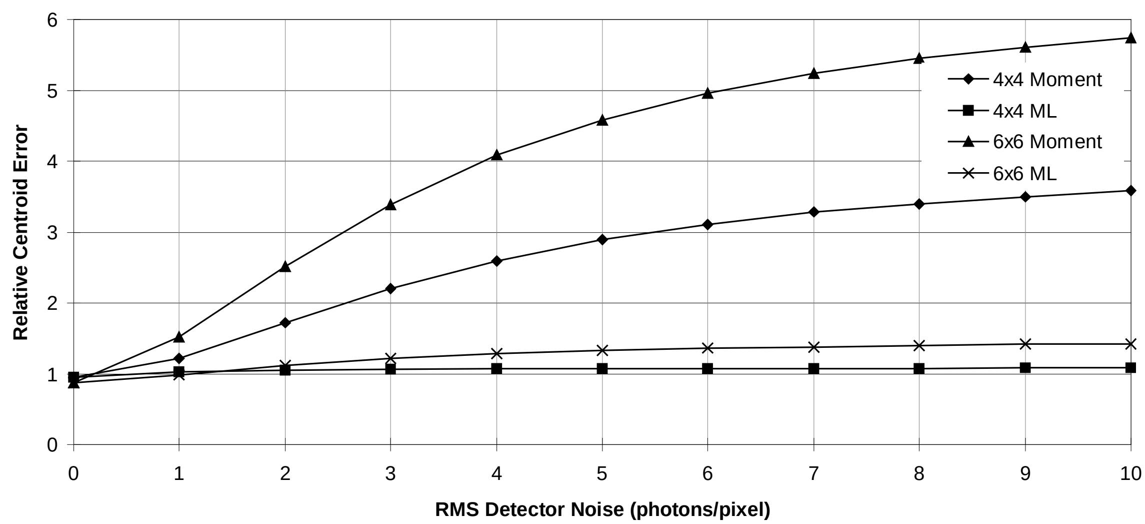

Figure 2. Noise propagation of moment-centroid and ML estimators for 4x4- and 6x6-pixel geometries.

Fig. 3 shows the results of simulations in which the ML and MAP estimators were compared. In this case the temporal variations in image position were modelled by a 1st-order auto-regressive process with Γ_{xy} = 0.1, 0.2 and 0.4, and 50% de-correlation between the centroid position occurring on successive image frames. A high-pass filter on the centroids was used to remove the low-frequency component from these variations in order to model the effect of the fast-guiding loop, the 3dB frequency for this filter was set to 0.5 (normalised by the Nyquist frequency).

Image-width variations were log-normal, as described above, but a 1st-order auto-regressive process was again used to generate the sequence of log-image-widths with 50% de-correlation occurring over 10^4 frames. The MAP priors were estimated using the approach described in section 4.3 with an update interval $L_{\text{update}} = 200$ frames. The figure shows the ratio of the RMS centroid error for MAP and ML estimators as a function of noise level and with varying amounts of image motion, Γ_{xy}. The results confirm our expectations: when the image motion is relatively small, and the noise level high, the MAP estimator provides a significant degree of noise suppression whilst at low noise levels, or when the image motion is relatively large, the MAP prior provides only a weak constraint on the data and there is little to choose from between the two estimators.

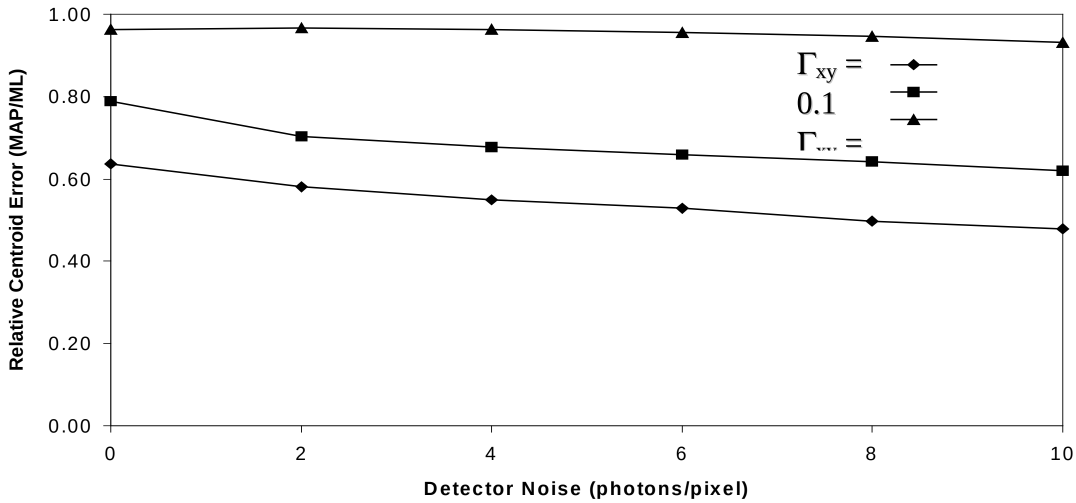

Figure 3. Ratio of MAP to ML centroiding errors for varying amounts of image motion.

6.2 Experimental Results

Open-loop speckle data in the near-IR (J, H and K′ bands) were available to us for evaluation (these data had been obtained as part of a separate study conducted by Klaus Hodapp of the University of Hawaii for the Gemini project). The data were obtained at the UH 2.2-m telescope

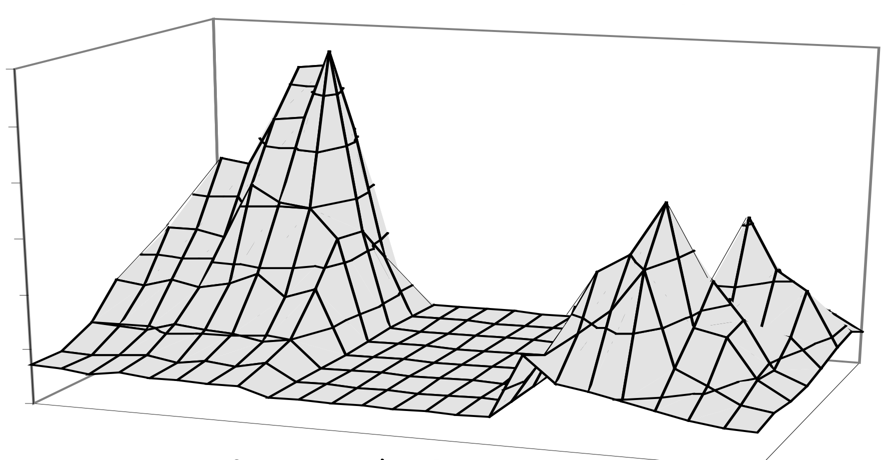

Figure 4. Two image frames at K′ with 20-s temporal separation.

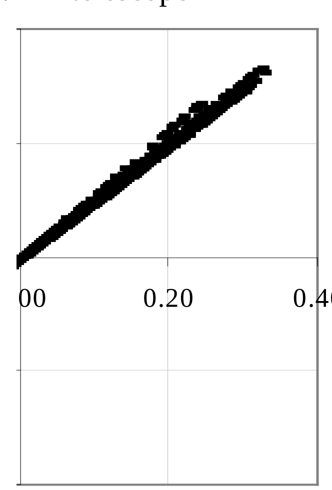

Figure 5. Comparison of ML and moment

using the full aperture with 8x8 pixels at 0.188 ″/pixel and consisted of a number of image sequences each of 45-s duration at a frame rate of 56 Hz. Seeing was poor during these observations, around 0.9″ FWHM at K′. The SNR was sufficiently high (≥ 20) that CCD read noise could be modelled by adding artificial gaussian noise to the raw data.

Fig. 4 shows a representative pair of image frames in H band with a temporal separation of 20 seconds. Speckles are evident in the PSFs and the first question we address is whether a simple model for the PSF will suffice when fitting to the data. We have therefore computed ML centroids using an assumed gaussian PSF with $\langle x \rangle, \langle y \rangle$ = 0,0 and a FWHM of 0.5″, at high SNR. The scatter plot shown in Fig. 5 shows that there is excellent agreement between the ML centroid estimates and the moment centroid and confirms the validity of the simple model-fitting technique. The small variation in the constant of proportionality which can be seen is caused by fluctuations in the image width - these fluctuations are accommodated by the ML estimator but not by the moment estimator because the normalising constant in Eq. 1 depends upon image width, whereas we necessarily chose a constant normalisation for our moment-centroid estimates.

We are cautious in our use of these data when evaluating the adaptive MAP estimator. The problem is that the power spectra for the image centroids in open-loop data contain a large amount of energy at low temporal frequencies, whereas most of this energy will be removed by the control loop in an AO system. The existence of this low-frequency energy forces us to adopt relatively weak priors on x and y in order to accommodate the observed variance in image centroids leaving little scope for MAP noise suppression. Nevertheless, as an illustration of the potential of the adaptive MAP technique we have tested a modified version of the basic algorithm using 90-seconds of K′ data binned to 4x4 pixels at a SNR of 5. The 90-s sequence was generated by splicing together two 45-s sequences which had been obtained at 320-s intervals. By treating this composite data set as a contiguous time-series we were able to assess the ability of the algorithm to track seeing variations over timescales of several minutes. A short adaptation timescale of L_{update} = 4 frames was adopted in order to allow the low-frequency centroid excursions to be tracked - this approach effectively causes the image motion seen by the MAP priors to be high-pass filtered and thus allows relatively tight priors (small variance on x, y) to be used over each 4-frame interval. The modification referred to above was to use the first 18-seconds of data to obtain an accurate estimate of the intrinsic variance on x and y over a sequence of 4 frames by averaging the sample variance over many 4-frame periods. These variance were then held <u>constant</u> following the initial training period whilst the local average of x and y were updated at the end of each 4-frame adaptation period in the conventional way described in section 4.3. Fig. 6 shows the noiseless moment centroid along with adaptive MAP estimates during and after the training period. The benefit afforded by the MAP estimator is obvious, however we re-iterate the fact that this open-loop tracking problem is somewhat artificial and clearly has significant potential for noise reduction by exploiting temporal correlations in the data, using, for example, conventional Wiener or Kalman filters in conjunction with a non-adaptive ML estimator.

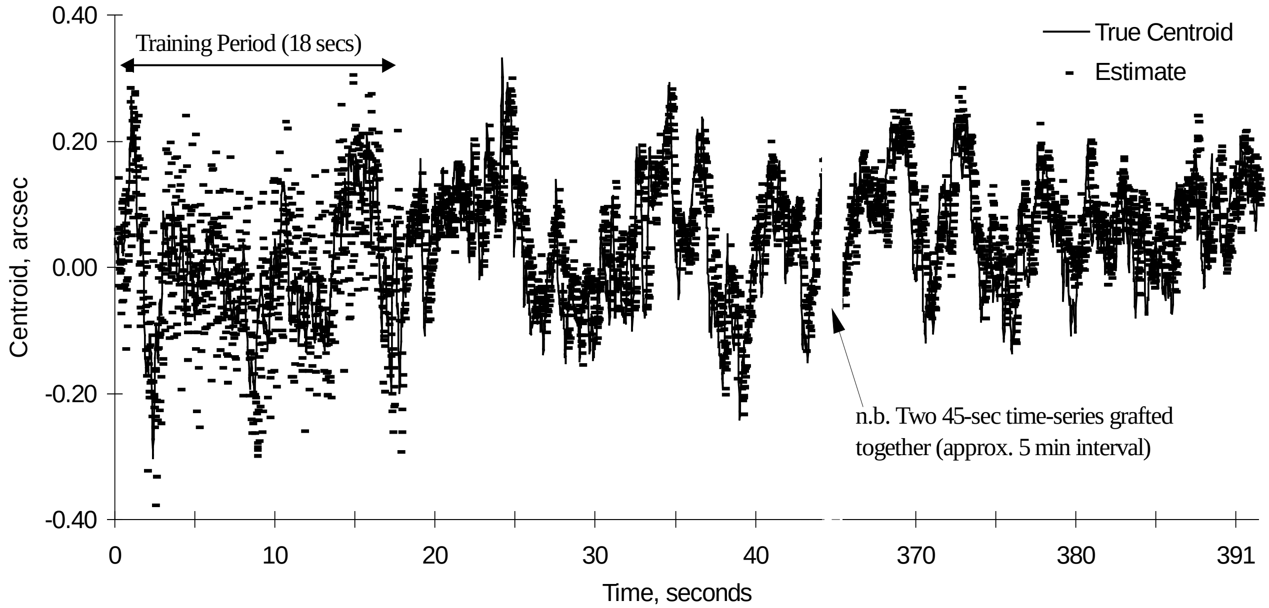

Figure 6. Demonstration of adaptive MAP algorithm, modified to track low-frequency image motion in the open-loop data.

7. CONCLUSIONS

A general form of estimator which can implement Bayesian (MAP) or Maximum Likelihood (ML) centroid estimation has been described. The estimator requires a user-provided model for the PSF within each sub-aperture and the only restriction on the choice of PSF is that it must be differentiable over its entire range. In addition, the MAP estimator requires a user-provided model for the statistical distribution of centroids parameters (image position and image width). A self-learning algorithm enables this prior distribution to adapt in order to track temporal variations due to atmospheric fluctuations.

The MAP and ML estimators provide a high degree of immunity against CCD read noise and enable the use of many-pixel (non quad-cell) detector geometries within each sub-aperture of the SHWFS, whilst achieving a SNR which is competitive with an idealised quad-cell. For example, under typical operating conditions (100-photon signal, 5 photons/pixel read noise), a 6x6-pixel ML centroid has a noise level around 30% worse than that of an ideal quad-cell whilst a 4x4 ML estimate degrades the sensitivity by less than 10% with respect to a quad-cell. In view of the fact that the idealised quad-cell is a goal which may not be achieved in many applications we believe that many-pixel centroiding in conjunction with optimum estimators such as those described here is likely to become more widely adopted in future AO systems.

The benefits afforded by the MAP estimator are largest when the intrinsic centroid variations that we seek to estimate are smaller than the ML estimation error. For a given amount of image motion, the benefits increase as the available signal (photon count) decreases whilst at high signal levels the MAP estimator performance converges on that of an ML solution. This behaviour is a characteristic of the Bayesian approach and has been highlighted by Solomon et al[8] in the context of a Bayesian wavefront reconstructor. We do not claim that the ability of the MAP estimator to suppress noise in very-low-SNR applications will necessarily extend the performance of AO systems to correspondingly lower signal levels, but believe that the graceful degradation of performance which the MAP estimator provides as the available SNR decreases is an important and valuable feature of this approach.

The computation of MAP/ML centroid estimates requires one multiply-accumulate operation per pixel per centroid. In the case of a DSP-based centroid processor, the ML or non-adaptive MAP estimators impose the same processing load as a simple moment centroid and may be implemented as a “drop-in” replacement for the latter with relatively little effort. An adaptive MAP estimator requires additional computational operations in order to track slow changes in atmospheric seeing. These latter operations may be implemented on the centroid processor if it has available processing capacity, but are otherwise suited to parallel implementation on a separate floating-point processor (CISC, RISC or DSP) which operates in a supervisory role, periodically downloading optimum estimator weights to the real-time centroid processor.

8. ACKNOWLEDGEMENTS

We thank Klaus Hodapp, of the University of Hawaii, for making the IR data described in section 5 available to us. The work described here was undertaken as part of the UK’s work on the Acquisition & Guidance system for the Gemini telescopes. The Gemini 8-m telescopes project is managed by the Association of Universities for research in astronomy, for the National Science Foundation, under an international partnership agreement.

9. REFERENCES

1. Rigaut, F. et al., “Comparison of curvature-based and Shack-Hartmann-based adaptive optics for the Gemini telescope”, *Appl. Opt.*, **36**, pp 2856-2868. 1988.
2. Oschmann J. et al., “Gemini 8 Meter Telescopes Active and Adaptive Optics Update”, *OSA/ESO Topical Meeting on Adaptive Optics Garching*, ESO Conf. Proc. No. 54. pp 83-88. 1996.
3. Tyler G. A. & Fried D.L., “Image-position error associated with a quadrant detector”, *J. Opt. Soc. Am.*, 72, pp 804-808. 1982.
4. Racine, R, “Temporal Fluctuations Of Atmospheric Seeing”, *Proc. Ast. Soc. Pac.*, 108, pp 372-374. 1997.
5. Roddier F., “The Effects of Atmospheric Turbulence in Optical Astronomy”, in *Prog. In Optics*, X1X, pp 280-376. 1981.
6. Racine, R. et al., “Mirror, Dome, and Natural Seeing at CFHT”, *Proc. Ast. Soc. Pac*, 103, 1020-1032. 1991.
7. Wilson, R., Private communication (preliminary results from site-testing on the WHT at La Palma Observatory).
8. Solomon C.J. et al., “Bayesian Estimation of Atmospherically Distorted Wavefronts Using Shack-Hartmann Sensors”, *Opt. Rev.*, 2, No. 3, pp 217-220. 1995.