Addition of Images with Varying Seeing

Nick Kaiser
Pan-STARRS Project Team
Institute for Astronomy
July 10, 2004
## Revision History

<table>
<thead>
<tr>
<th>Revision Number</th>
<th>Release Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>2001.07.01</td>
<td>First draft</td>
</tr>
</tbody>
</table>
Abstract

We consider the problem of combining astronomical images with varying seeing and or noise. We assume that the images are sky-noise or read-noise dominated and that the counts are sufficiently high that the noise fluctuations are effectively Gaussian. Under these conditions there is a unique optimal way to combine images with varying seeing and or noise properties: convolve each image with the reflection of its point spread function and then accumulate with weight inversely proportional to the sky variance. This is useful for upcoming large imaging surveys such as the LSST since it means that there is no need to store the whole stream of images, rather one can simply store an accumulated image, and this single accumulated image is optimal for all applications. This optimal accumulated image can be post-processed in various ways, several of which we discuss.
List of Figures

1  Aliasing in Good Seeing ........................................... 13
2  Aliasing in Exceptional Seeing ................................. 14
3  Image Accumulation Pipeline ................................. 17
4  Matched Image Subtraction .................................... 18
1 Introduction

The motivation for this study was consideration of ‘Large Synoptic Survey Telescopes’ (LSSTs) (ref DRA) such as the proposed DMT (ref ??) and the POI/Pan-STARRS project (ref ??). Such instruments will repeatedly image the sky to faint magnitudes and are expected to produce two basic image data products: an accumulated image of the static sky, and a stream of difference images from which transient, variable or moving objects may be detected. Being ground-based, these facilities will take data under a wide range of seeing conditions. This raises the question of whether there is a unique optimal way to accumulate the static sky image, or do different scientific goals require different analysis techniques. This is a question of some practical importance, since in the latter case an optimal treatment would require storing the several petabytes of data that such systems will generate in total whereas a single accumulated image would only require a much smaller amount of memory store. While focused primarily on the static sky image, this question also has relevance for the difference image since the better one can generate the static sky image the better one can construct difference images. Better difference images in turn will allow one to better reject transients from the static sky image. While motivated by LSST applications, the problem of combining multiple CCD images arises in many contexts, so the results obtained here may be of broader applicability.

Here we shall show that in the limit that one has large numbers of photons per pixel, and in the limit that the objects in the field are faint compared to the sky background, the likelihood of the true sky (or ‘object scene’) is a function only of a single accumulated image. This image is formed by taking the source image data — considered as a set of delta functions placed at the centers of the pixels and with weight proportional to the measured count of electrons — and convolving with the point spread function (PSF) and then accumulating with weight inversely proportional to the sky + read-noise variance. Since, by Bayes’ theorem, the likelihood tells us how to update our state of knowledge about the object scene in the light of the data, this optimal accumulated image — along with auxiliary information regarding the pixel locations on the sky, PSFs and sky variance — contains all of the information in the input data stream. The optimal cumulative image is, in a sense, a completely lossless compression of the entire input data stream. As the cumulative image is a continuous function of position, it is somewhat dubious to claim that this is a compression, but in practice the cumulative image will be approximated by some pixellisation on a grid somewhat finer than the original images, so for large numbers of input images the algorithm does indeed provide effectively lossless compression.

The cumulative image only contains all of the information in the limit that the counts are large — so the fluctuations in the counts obey Gaussian statistics — and that the noise is dominated by the shot-noise of the sky background and/or read noise. For the LSST projects the former is a good approximation for any plausible integration time and the latter is a good approximation nearly everywhere on the sky. One square arcsecond of ‘dark’ sky has magnitude $m \sim 20$ in the $I$-band, for instance, so with FWHM $\sim 0''6$ seeing, the central surface brightness for a star is approximately equal to the sky background for a 20th magnitude star. For fainter point sources, and for essentially all work involving galaxies, we may neglect the contribution of the signal counts to the noise.

There is a considerable body of literature addressing this specific problem and other closely related problems, but, somewhat worryingly, no unified view has emerged. For example, Fischer & Kochanski (1994) have considered the problem of combining images with varying seeing, but they find a variety of different optimal weighting schemes for different applications. This is because they did not consider the possibility of convolving the images with their PSFs, which is central to the algorithm here. An iterative approach to combining images with different PSFs was proposed by Lucy & Hook (1992), but the relation of this to the present algorithm...
remains unclear. Considerable effort has been applied to combining images from the heavily under-sampled WFPC images from HST, and again a diverse range of solutions have been proposed, including the heuristic ‘drizzling’ technique (Hook & Fruchter 2000; Casertano et al. 2000) and the more rigorously motivated Fourier image reconstruction technique Lauer 1999a, Lauer 1999b. The problem of determining the PSF for heavily under-sampled images has been considered in detail by Anderson & King (2000).

In what follows we first consider the simple case of well sampled images §3, for which the optimality of the proposed algorithm is most clearly understood. We then turn to the more interesting problem of discretely, and generally non-uniformly, sampled data. We attack this from the point of view of probability theory; we write down the likelihood functional for the true ‘object scene’ \( f(\mathbf{r}) \). We show that this depends on the data only through the optimal cumulative image described above. We also discuss a number of complications that we ignore in deriving the optical image addition scheme. These include the problem of non-Gaussian noise in the form of cosmic rays and other instrumental artefact; the problem of determining the PSF; how a slight modification of the method results in a strong suppression of PSF anisotropy, which is useful for weak lensing observations.

We stress that our analysis only dictates what is the quantity that needs to be accumulated. There are many ways to post-process this accumulated image which can, and should, be fine tuned to particular applications. In two accompanying papers we will consider the detection and counting of point sources and also the problem of recovering from the optimal accumulated image a ‘de-aliased’ image.

## 2 The Imaging Model

We are interested here in determining the surface brightness of the sky as a function of celestial coordinates \( f(\alpha, \delta) \). Rather than work in angular coordinates it is convenient to work with a projection of the sky onto a tangent plane, and to denote the projected surface brightness by \( f(\mathbf{r}) = f(\alpha, \delta) \). We will assume that we are dealing with exposures much longer than the speckle coherence time, so that the PSF will have a smooth form with a well defined central maximum. This means that we should be able to construct a well-defined smooth mapping from sky coordinates \( \mathbf{r} \) to detector coordinates \( \mathbf{x} \).

### 2.1 The Point Spread Function

Diffraction theory, along with a prescription for the phase fluctuations imposed by the atmosphere, provides us with the classical Poynting flux in the focal plane

\[
F(\mathbf{x}) = V(\mathbf{x}) \times [f(\mathbf{r}(\mathbf{x})) \otimes g_x(\mathbf{x})]
\]  

(1)

where \( V(\mathbf{x}) \) is a smooth function describing the vignetting of the telescope, \( g_x(\mathbf{x}) \) is the point spread function, and the convolution operator \( \otimes \) is defined such that \( (a \otimes b)_r \equiv \int d^2r' a(\mathbf{r}')b(\mathbf{r} - \mathbf{r}') \). We will assume that the PSF is normalised: \( \int d^2x \ g_x(\mathbf{x}) = 1 \). For long exposures, the PSF \( g_x(\mathbf{x}) \) is the convolution of the atmospheric PSF and the pupil PSF. In general, these components of the PSF will generally have a core with fairly rapidly falling wings. For instance, any pupil with sharp edges — as is usually the case — will have extended wings falling as \( g \sim 1/r^3 \). For pure Kolmogorov turbulence in a large telescope the PSF has wings with \( g \propto 1/r^{11/3} \),
but in real systems the PSF often tends towards a shallower slope \( g \sim 1/r^2 \) at large radius. This is usually attributed to fine scale mirror roughness.

An ideal energy measuring detector would record the Poynting flux pattern \( F(\mathbf{x}) \). In a CCD, electrons are created at a rate proportional to the local photon flux. The occupation number for the photons is very small — if it were not, the CCD would melt — so the electrons generated have Poisson statistics. Once the electrons are created they are attracted towards the gates where they become confined to a egg-crate-like set of potential wells. The transport of the electrons from their point of creation to these potential wells introduces a further smearing of the image, in addition to the diffraction theory PSF. There is some probability \( P(x_p|\mathbf{x}) \) that an electron created at position \( \mathbf{x} \) will be counted in the \( p \)th pixel. This is the pixel PSF \( g_{p|x}(\mathbf{x}) \). To a crude approximation, this is just a box function (i.e. unity if both components of \( \mathbf{x} \) are less than one half the pixel size and zero otherwise). More realistically, this will also incorporate the effects of charge diffusion, which allows photons landing outside (inside) the box to (fail to) be registered in that pixel, and there may also be intra-pixel variations in sensitivity in front-side illuminated devices, and in the red for back-side illuminated devices. With this model, the expectation value of the count of electrons in the \( p \)th pixel is

\[
\mu_p = V(\mathbf{x}) \times (f \otimes g)_{x_p} \tag{2}
\]

where \( g(\mathbf{x}) \) is now the convolution of the atmospheric PSF, the pupil PSF and the pixel PSF.

In the following analysis we shall make much use of the Fourier transform of the PSF, i.e. the optical transfer function (OTF) \( g(k) \). This tells us how the harmonic components of the object scene are represented in the image formed on the focal plane. The total OTF is, by the convolution theorem, the product of the transfer functions for the atmosphere, pupil and pixel. These components have the following general properties:

- The atmospheric OTF is an exponential; \( g(k) = \exp(-\alpha k^{5/3}) \) with \( \alpha \) a constant. High frequency features in the image are exponentially suppressed by atmospheric seeing.
- The pupil OTF is the auto-correlation of the pupil function \( g(k) = \int d^2y \ A(y) \ \tilde{A}(y + k\lambda/2\pi) \). For a circular pupil of diameter \( D \) this vanishes for angular wave-numbers \( |k| > 2\pi D/\lambda \), so high frequencies are entirely removed.
- If we model the pixel PSF as a simple box then the pixel OTF is a two dimensional ‘sinc’ function factor \( g_{p|x}(k) \sim \text{sinc}(k_xd)\text{sinc}(k_yd) \) with \( \text{sinc}(y) = \sin(y/2)/(y/2) \) and \( d \) the pixel size. The sinc function removes spatial frequencies which lie along the ‘rib-lines’ \( k_x, k_y = n\pi/d \) for \( n \neq 0 \). Charge diffusion can be modelled as a Gaussian smearing, and this would introduce a Gaussian factor in the pixel OTF.

The removal, or exponential suppression, of certain spatial frequencies severely constrains ones ability to reconstruct the true object scene.

When combining images it is more convenient to work in sky coordinates \( \mathbf{r} \) rather than detector coordinates \( \mathbf{x} \). We can readily define an equivalent PSF in sky coordinates \( g(\mathbf{r}) \) such that the expected count in the \( p \)th pixel is

\[
\mu_p = V(\mathbf{r}, \mathbf{x}) \times (f \otimes g)_{r_p} \tag{3}
\]

\(^1\text{The transform convention here is } f(k) = \int d^2r \ f(\mathbf{r})e^{-ikr} \text{ and we will mostly use the same symbol for the transform, relying on the argument to indicate what space we are working in. We will use } \mathbf{k}, \mathbf{k}' \text{ etc for wave-vectors. Occasionally, we use the notation } f(k) = \hat{f} \text{ if no arguments are given.}\)
where $V(\mathbf{r}, \mathbf{x})$ is a smooth function of sky and detector coordinates, and the total PSF $g(\mathbf{r})$ is again assumed to be normalised to $\int d^2 r \, g(\mathbf{r}) = 1$. The total PSF can be defined operationally: If we take a series of exposures such that the center of a star is placed at a set of positions $\{\mathbf{r}\}$ then the mean count in the $p$th pixel is $\mu_p \propto g(\mathbf{r}_p - \mathbf{r})$. As we shall see, all that we need in the analysis is the total PSF; we do not need to know how the PSF breaks down into the various components. The calibration function $V(\mathbf{r}, \mathbf{x})$ can also be defined operationally, say by making measurements of the total flux of standard stars and interpolating.

### 2.2 Pixel Count Statistics

Neglecting read-noise, digitization-noise, etc., the probability that the count for a pixel is $n$ given the expected count $\mu$ is given by the Poisson distribution

$$P(n|\mu) = \frac{\mu^n e^{-\mu}}{n!}.$$  \hspace{1cm} (4)

This is also the likelihood of $\mu$ given count $n$: $L(\mu) = P(n|\mu)$. The likelihood of the true object scene $f(\mathbf{r})$ is then a product over pixels

$$L(f(\mathbf{r})) = P(\text{data}|f(\mathbf{r})) = \prod_p \frac{\mu_p^n e^{-\mu_p}}{n_p!}$$  \hspace{1cm} (5)

with $\mu_p$ given by equation (3). For large $\mu$, the Poisson distribution (4) becomes sharply peaked. The natural logarithm of $P(n|\mu)$, or the log-likelihood of $\mu$, is

$$\mathcal{L}(\mu) = \log P(n|\mu) = n \log \mu - \mu + \text{constant}$$  \hspace{1cm} (6)

where by ‘constant’ we mean independent of the model parameter $\mu$. Differentiating with respect to $\mu$ we find that the likelihood is peaked for $\mu_0 = n$ and performing a Taylor expansion around the maximum we have

$$\log P(n|\mu) \simeq n \log n - \frac{1}{2} n \left(\frac{\mu - \mu_0}{\mu_0}\right)^2 + \ldots = \text{constant} - \frac{1}{2} \frac{(n - \mu)^2}{n} + \ldots$$  \hspace{1cm} (7)

This means that the likelihood of $\mu$ becomes, for large counts, a Gaussian: $L(\mu) \propto \exp(-(n - \mu)^2/2n)$. The likelihood of $f(\mathbf{r})$ is therefore the product of a set of Gaussians, one per pixel:

$$P(\text{data}|f(\mathbf{r})) = \prod_p \frac{1}{\sqrt{2\pi} \sigma_p} \exp \left( -\frac{1}{2} \frac{(f_p - (f \otimes g)_{\mathbf{r}_p})^2}{\sigma_p^2} \right)$$  \hspace{1cm} (8)

with $f_p = n_p - \bar{n}$, with $\bar{n}$ the mean count from the sky background and where $\sigma_p^2 = n_p$. Additional sources of noise such as read noise, digitization etc, can be incorporated by adding to $\sigma_p^2$ the extra variance. Finally, for sources which are faint compared to the atmospheric emission the variance $\sigma_p^2$ becomes independent of the signal, and we have $\sigma_p^2 = \bar{n}$, augmented by any extra sources of noise variance.
### 3 Accumulation of Well Sampled Images

#### 3.1 Optimal Addition of Images

Let's consider first the limiting, though seldom realised, case that the pixel size is much smaller than the actual width of the PSF. Such ‘well sampled’ images can be treated mathematically as continuous functions of position, and can trivially be interpolated to a common coordinate system (once we have located stars and found the transformation from chip coordinates to sky coordinates). We can model a collection of images, labeled by image number $i$, as

$$f_i(r) = (g_i \otimes f)_r + n_i(r)$$  \hspace{1cm} (9)$$

where $f(r)$ is the true object scene, $g_i(r)$ is the PSF for the $i$th image, and $n_i(r)$ is the noise. We will assume here that the photon counts are sky noise dominated; though for bright objects this will not be the case. With this assumption, the noise $n(r)$ is a statistically homogeneous random field. We will assume that the mean sky has been subtracted. We will also assume that the PSF for each image $g_i(r)$ has been determined from stars in the image, allowing for variation across the image if necessary (see §5.2).

If we Fourier transform (9) becomes

$$f_i(k) = g_i(k) f(k) + n_i(k).$$  \hspace{1cm} (10)$$

Now since the sky noise is statistically homogeneous, distinct Fourier components $n_i(k)$ are uncorrelated. Moreover, Poisson noise has a flat spectrum so $\langle |n_i(k)|^2 \rangle = \sigma_i^2$ is independent of frequency. Thus each measured Fourier component is the true transform times $g_i(k)$ — which is the optical transfer function (OTF) — plus an independent random fluctuation with constant rms value $\sigma_i$. The optimal estimate of $f(k)$ is then trivially derived: each image gives an unbiased estimate

$$\hat{f}(k)_i = f_i(k)/|g_i(k)|^2 = \sigma_i^2/|g_i(k)|^2,$$

which has variance $\langle |\delta \hat{f}(k)|^2 \rangle = \langle |n(k)|^2 \rangle/|g_i(k)|^2 = \sigma^2/|g_i(k)|^2$, and the optimally weighted linear combination of these is

$$\hat{f}(k) = \sum_i w_i f_i(k) / \sum_i w_i$$

with weight which is just the inverse of the variance $w_i = |g_i(k)|^2/\sigma_i^2$. The optimal estimator is then simply

$$\hat{f}(k) = \frac{\sum g_i^*(k) f_i(k)/\sigma_i^2}{\sum |g_i(k)|^2/\sigma_i^2}.$$  \hspace{1cm} (11)$$

Now the key feature here is that the data $f_i(k)$ appear in a sum over images with multiplicative weights $g_i^*(k)/\sigma^2$, so back in real space, the optimal estimator must involve an average of images $f_i(r)$ convolved with their individual PSFs $g_i(r)$ and weighted by $1/\sigma_i^2$. More precisely, convolved with the transform of $g_i^*(k)$ which is the reflection of the PSF $g_r(\mathbf{r})$, which we denote by $g^!(\mathbf{r})$. Thus the data will appear in the real-space estimator only through the appearance of the function

$$\varphi(r) = \sum_i (g_i^! \otimes f_i)_r/\sigma_i^2.$$  \hspace{1cm} (12)$$

This equation tells us that we can accumulate an optimal image by convolving each image with (the reflection of) its PSF and adding to the accumulated sum with weight $\propto 1/\sigma_i^2$.

At first sight this might seem crazy. We go to great lengths and expense to build telescopes on high mountains in order to get good seeing, and then once we have obtained images we are told to immediately degrade them...
to make the PSF worse. However, unlike atmospheric and instrumental seeing, there is no information loss involved in this convolution as it is applied after the noise has been realized — this convolution is exactly reversible. On reflection, the algorithm should not seem too unreasonable. Imagine one has a set of images with the same noise but widely varying seeing. Clearly all of the images contain the low frequency components of the signal in equal measure, so we should average all the images with equal weight to optimally recover the low-\(k\) information. The high frequency information, however, will be almost entirely absent from very bad seeing images, and giving these any weight in the summation will just add noise. In general we clearly want some way to continuously control the weighting of the various images as a function of spatial frequency, and this is supplied by equation (12).

### 3.2 Alternative Representations

If we try to turn (11) into a practical real-space estimator of the object scene we immediately run into problems. Clearly, we cannot just transform this; the numerator contains a flat noise component \(n_{i}(k)\) multiplied by the OTF \(g^*(k)\) which falls off exponentially fast at high frequency. The denominator, however, contains a factor \(|g_{i}(k)|^2\) which falls off even faster. This should not come as a surprise since, as we have already noted, the optical system either removes or exponentially suppresses high spatial frequency content, so recovering these would require an exponentially difficult or impossible deconvolution. If we were to try this, we would obtain a signal with a constant OTF, and therefore with a delta-function PSF, but with noise component diverging at high frequencies with variance \(\sum_{i} |g_{i}(k)|^2/\sigma_i^2\). What if instead we try to recover an image which, like the source images, has a flat noise spectrum. This simply involves multiplying by $\sqrt{\sum_{i} |g_{i}(k)|^2/\sigma_i^2}$ to give

$$\hat{f}_i(k) = \frac{\sum_{i} g_{i}^*(k) f_{i}(k)/\sigma_i^2}{\sqrt{\sum_{i} |g_{i}(k)|^2/\sigma_i^2} \sqrt{\sum_{i} 1/\sigma_i^2}} \tag{13}$$

where the suffix ‘fs’ denotes flat (noise) spectrum. It is easy to see that the expectation value is \(\langle \hat{f}_i(k) \rangle = g_{fs}(k) f(k)\), where the effective OTF is

$$g_{fs}(k) = \frac{\sqrt{\sum_{i} |g_{i}(k)|^2/\sigma_i^2}}{\sqrt{\sum_{i} 1/\sigma_i^2}} \tag{14}$$

and is simply the weighted rms of the individual OTFs. Note that the effective OTF is real, so the corresponding effective PSF is symmetric. An example of a flat noise spectrum image generated in this way is shown in figure 3.

The inverse transform of \(\hat{f}_i(k)\) is perfectly regular — one can simply evaluate (13) at all frequencies where the effective OTF in the denominator exceeds some tiny value and set the transform to zero otherwise, and the resulting numerical transform, aside from smoothing of the noise at microscopic scales, looks just like a single image as would have been obtained from an instrument with OTF \(g(k) = g_{fs}(k)\). We stress that the flat spectrum average is simply a convenient representation of the optimal average of the data; (13) contains precisely the same information as (11). It is particularly useful if one is using simple photometry packages which assume
that the sky noise is uncorrelated in order to estimate errors. We could have chosen any other desired effective
OTF, though with the proviso that trying to recover exponentially suppressed spatial frequencies will result in
exponential amplification of the noise.

### 3.3 Relation to Wiener Filtering

The kind of problem we are concerned with here is often tackled using Wiener filtering. For example, imagine
we are provided with a single image \( c(\mathbf{r}) \) which is the convolution of some desired true image \( f(\mathbf{r}) \) with a kernel
\( g \) plus noise:

\[
c(\mathbf{r}) = (g \ast f)_\mathbf{r} + n(\mathbf{r}).
\]

In the absence of noise one might be tempted to simply divide \( c(\mathbf{k}) \) by \( g(\mathbf{k}) \) to obtain an estimate of the true
image scene transform

\[
\hat{f}(\mathbf{k}) = \frac{c(\mathbf{k})}{g(\mathbf{k})}.
\]

The Wiener filter estimator is a modification

\[
\hat{f}(\mathbf{k}) = W(\mathbf{k}) \frac{c(\mathbf{k})}{g(\mathbf{k})}
\]

where the filter \( W(\mathbf{k}) \) is chosen in order to minimize the mean square deviation between \( \hat{f}(\mathbf{r}) \) and \( f(\mathbf{r}) \):

\[
\int d^2r \ (\hat{f}(\mathbf{r}) - f(\mathbf{r}))^2 = \int d^2k \ (\hat{f}(\mathbf{k}) - f(\mathbf{k}))^2.
\]

The optimal filter, in this sense, is readily shown to be

\[
W(\mathbf{k}) = \frac{P_S(\mathbf{k})}{P_S(\mathbf{k}) + P_N(\mathbf{k})}
\]

where \( P_S(\mathbf{k}) \) and \( P_N(\mathbf{k}) \) are the signal and noise components of the power spectrum of \( c(\mathbf{r}) \).

This is very different from what we are doing. The Wiener filter tells us how to combine information from
different spatial frequencies such that the synthesised image \( \hat{f}(\mathbf{r}) \) is as close as possible, in a least squares
sense, to the true object scene \( f(\mathbf{r}) \). The Wiener filter does not change the information content, since it is a
reversible filtering. The estimator (11) or (13) in contrast, tells us how to combine the information from a set of
images at a single spatial frequency. This is highly irreversible; if a weighting other than the optimal one is used
then there will be loss of information. Another distinction is that the Wiener filter requires some estimate of the
signal power in the input image, whereas for the optimal accumulator we need only estimate the noise variance
\( \sigma_n^2 \). Wiener filtering is a potentially useful ‘post-processing’ of the optimal accumulated image, which may
help the viewer distinguish the signal from the noise, but should not be confused with the optimal algorithm for
combining images.

### 3.4 Real-Space Formulation

A relevant feature of (13) for our purposes is that operationally, one can simply accumulate all the relevant
information by forming the sum (12) which is simply evaluated and well behaved since, as discussed, the PSF
is generally fairly compact. We also need to be able to compute the terms in the denominator of (13), which we can do if we keep track of the auxiliary quantities $\sum_i (g_i \otimes g_i)_x / \sigma_i^2$ — which is essentially the transform of the effective OTF (14) — and also $\sum_i 1/ \sigma_i^2$. Armed with these sums it then easy, at any time, to apply a well conditioned Fourier space division to generate a real space estimator with any desired transfer function, be it flat spectrum, Wiener filter or whatever. In particular, one can generate an image which has effective PSF identical to $g_i \otimes f_i$, the convolution of the most recently acquired image with its PSF. This matched image can therefore be subtracted from the last (convolved) image to reveal transient objects.

4 Discretely Sampled Images

4.1 The Image Probability Functional

We will now generalize the analysis to treat discretely sampled images where the pixel scale is not very fine as compared to the PSF width (as is usually the case). In the imaging model described above, a CCD camera supplies one with a set of discrete image samples, which we will denote by $f_p$. For a static sky or object scene the pixel values are then the values of the convolution of the object scene with the PSF, plus an uncorrelated random noise component:

$$f_p = (f \otimes g)_{x_p} + n_p$$  \hspace{1cm} (20)

For a single image, these samples will typically lie on a uniform grid in focal plane coordinates (and therefore on a nearly uniformly sampled grid in sky coordinates). Multiple images provide one with multiple overlapping grids of samples, each of which will have its own atmospheric PSF etc.

The likelihood is the probability of the data given the hypothesis (i.e the true object scene $f(x)$). Since the noise values are uncorrelated, $L(f(x))$ is simply the product of a set of Gaussian distributions (8) The significance of the likelihood is that, according to Bayes’ theorem, our state knowledge about the object scene after making the observations — the posterior probability $P_{\text{post}}(f(x))$ — is equal to the prior probability — whatever that happens to be — times the likelihood:

$$P_{\text{post}}(f(x)) \propto L(f(x))P_{\text{prior}}(f(x)).$$  \hspace{1cm} (21)

Now one’s choice of prior is, like one’s choice of soccer team or religion, a subjective, though perhaps contentious, matter. It may, for instance, be conditioned by other observations, though devout ‘Bayesians’ would hold that there are particular preferred ‘priors’ even in the absence of any information. Regardless of one’s prior prejudice or knowledge, the data in question enter only in the likelihood, so we need not concern ourselves with choice of prior.

The likelihood here is a functional; i.e. a scalar function of the continuous object scene $f(x)$. To simplify the maths, let us imagine discretizing the object scene on some fictitious, but very very fine, uniform grid $f_x = f(x)$ where the points $x$ on the grid are labelled in an arbitrary manner by the 1-dimensional index $x$ (we will reserve the suffix $p$ to denote actual pixel values). We can then think of the image as a large $N$-component vector $\mathbf{f} = \{f_0, f_1, \ldots, f_{N-1}\}$. The convolution can then be written as a discrete sum, which in turn is a dot-product $$(f \otimes g)_{x_p} = \sum_x f_x g_{px} = \mathbf{f} \cdot \mathbf{g}_p$$ where $\mathbf{g}_p$ is the image-vector with components with
\[ g_{px} \equiv g(r_p - r_x). \] In index notation, the log-likelihood (8) is

\[
\mathcal{L}(f_x) = -\frac{1}{2} \sum_p (f_p - \sum_x f_g g_{px})^2 / \sigma_p^2
\]

\[
= \sum_x f_x \sum_p f_p g_{px} / \sigma_p^2 - \frac{1}{2} \sum_x \sum_x' f_x f_x' \sum g_{px} g_{px'} / \sigma_p^2 + \text{constant}
\]  

(22)

or, in vector notation,

\[
\mathcal{L}(\mathbf{f}) = \mathbf{f} \cdot \varphi - \frac{1}{2} \mathbf{f} \cdot \mathbf{A} \cdot \mathbf{f} + \text{constant}
\]

(23)

where we have defined the image-vector

\[
\varphi = \sum_p f_p g_p / \sigma_p^2
\]

(24)

and the image-matrix

\[
\mathbf{A} \equiv \sum_p g_p g_p / \sigma_p^2.
\]

(25)

(26)

Re-arranging terms — essentially completing the square — we can write the log-likelihood as

\[
\mathcal{L}(\mathbf{f}) = -\frac{1}{2} (\mathbf{f} - \mathbf{\bar{f}}) \cdot \mathbf{A} \cdot (\mathbf{f} - \mathbf{\bar{f}}) + \text{constant}
\]

(27)

with

\[
\mathbf{\bar{f}} \equiv \mathbf{A}^{-1} \cdot \varphi.
\]

(28)

Exponentiating \( \mathcal{L}(\mathbf{f}) \) we obtain the likelihood

\[
L(\mathbf{f}) = \sqrt{\frac{|\mathbf{A}|}{(2\pi)^N}} \exp \left(-\frac{1}{2} (\mathbf{f} - \mathbf{\bar{f}}) \cdot \mathbf{A} \cdot (\mathbf{f} - \mathbf{\bar{f}}) \right)
\]

(29)

where we have filled in the normalization pre-factor by inspection. The likelihood is thus a shifted Gaussian, where the mean (i.e. most likely, or, for flat prior, most probable) image-vector is given by \( \mathbf{\bar{f}} \).

It might seems that (28) and (25) provide a fairly straightforward, though perhaps computationally expensive, way to compute the most likely object scene. However, as already discussed, it is futile to hope to be able to recover the true object scene because of the nature of the OTFs we are dealing with. To see more clearly why this fails it may be helpful to make connection with the analysis for continuously sampled images. In the limit of a large number of images the pixel samples become very dense, and the the sum over pixels in (26) for example becomes effectively a 2-dimensional integral:

\[
\mathbf{A}(\mathbf{r}, \mathbf{r}') = \sum_p \mathbf{g}_p \mathbf{g}_p / \sigma_p^2 \rightarrow \int d^2 r_p \ g(\mathbf{r}_p - \mathbf{r}) g(\mathbf{r}_p - \mathbf{r}') / \sigma^2 \propto (g^\dagger \otimes g)_{\mathbf{r}'-\mathbf{r}}.
\]

(30)

Similarly, if we take the dot product of \( \mathbf{A} \) with some image-vector \( \mathbf{v} \), this is a further convolution:

\[
\mathbf{A} \cdot \mathbf{v} = \sum_x A_{xy} v_y \rightarrow \int d^2 x \ A(\mathbf{x}, \mathbf{y}) v(\mathbf{y}) = g^\dagger \otimes g \otimes v.
\]

(31)
For densely sampled images then, $A$ is an operator which effects a double convolution with the PSF. The inverse of $A$ appearing in (28) is therefore the (double) deconvolution operator.

This tells us that multiplication by $A^{-1}$ in (28) cannot be well defined since it is tantamount to performing a physically impossible deconvolution. The maximum likelihood image-vector defined in (28) is therefore similarly ill defined. The relevant point here, however, is that the data $f_p$ enter the likelihood only in the image-vector sum $\varphi = \sum_p f_p g_p / \sigma_p^2$ or equivalently in the image

$$\varphi(r) = \sum_p f_p g_p (r_p - r) / \sigma_p^2.$$  

(32)

This is rather general as it allows each pixel to have its own PSF and noise value. It is very often a rather good approximation to assume that all of the pixels in a given image (or within some extended region of a given image) have effectively the same PSF and noise. It is then useful to define the function

$$f_i(r) \equiv \sum \sum f_p \delta(r - r_p),$$

(33)

where the sum is only over those pixels in the $i$th image. This is just a set of impulses at the locations of the pixels with weight equal to the pixel value. The optimal cumulative image can then be written as

$$\varphi(r) = \sum_i (g_i^1 \otimes f_i) r / \sigma_i^2.$$  

(34)

This is exactly the same as in (12) save that the continuously sampled data are replaced by the impulsive function $f_i(r)$. The likelihood contains all of the information contained in all of the original images which is relevant to the question of what is the true object scene. Thus the single cumulative image (34) also preserves all of the useful information.

### 4.2 Statistics of the Accumulated Image

The accumulated image $\varphi(r)$ contains both a signal and noise component. Substituting (20) for $f_p$ in (32) gives

$$\varphi(r) = \varphi_S(r) + \varphi_N(r) = \sum_p (g \otimes f) r_p \frac{g(r_p - r)}{\sigma_p^2} + \sum_p n_p \frac{g(r_p - r)}{\sigma_p^2}.$$  

(35)

#### 4.2.1 Noise Properties

The covariance of the noise fluctuations $\varphi_N$ is

$$\langle \varphi_N(r) \varphi_N(r') \rangle = \sum_p \sum_{p'} \langle n_p n_{p'} \rangle \frac{g(r_p - r) g(r_{p'} - r')}{\sigma_p^2 \sigma_{p'}^2}$$

(36)

$$= \sum_p \frac{g(r_p - r) g(r_p - r')}{\sigma_p^2} = \sum_p \frac{g_p g_p}{\sigma_p^2} = A(r, r').$$  

(37)

In the limit that we have a large number of images this becomes just a function of $r - r'$ and the noise becomes a statistically homogeneous process, with two-point function $A(0, r)$ and it is then straightforward to apply an appropriate Fourier space filtering to $\varphi(r)$ to render the noise spectrum flat just as in the continuous case.
4.2.2 Signal Properties

What about the properties of the signal component \( \varphi_S(\mathbf{r}) \)? This is

\[
\varphi_S(\mathbf{r}) = \sum_p (g \otimes f)_{p} \frac{g(\mathbf{r}_p - \mathbf{r})}{\sigma_p^2} = \sum_p \int d^2 r' \delta(\mathbf{r} - \mathbf{r}_p)(f \otimes g)_{r'} g(\mathbf{r}' - \mathbf{r})/\sigma_p^2. \tag{38}
\]

If we specialize to the case that the PSF and noise are the same for all pixels in a given image and define the function

\[
c_i(\mathbf{r}) \equiv \sum_{p \subset i} \delta(\mathbf{r} - \mathbf{r}_p) \tag{39}
\]

where the notation \( p \subset i \) means we are summing only over the subset of pixels in the \( i \)th image, this becomes

\[
\varphi_S(\mathbf{r}) = \sum_i \int d^2 r' g_i(\mathbf{r}' - \mathbf{r})c_i(\mathbf{r}') (f \otimes g_i)_{r'}/\sigma_i^2 = \sum_i g_i^\dagger \otimes [c_i \times (g_i \otimes f)]/\sigma_i^2. \tag{40}
\]

The signal content in the optimal summed image \( \varphi(\mathbf{r}) \) is therefore the convolution of the true object scene with the PSF \( (g_i \otimes f) \), sampled at the locations of the pixels \( (c_i \times (g_i \otimes f)) \) and then finally re-convolved with the PSF.

4.2.3 Aliasing

The transform of \( \varphi_S(\mathbf{r}) \) is

\[
\varphi_S(\mathbf{k}) = \sum_i \tilde{g}_i^* \otimes \tilde{c}_i \otimes (\tilde{g}_i \times \tilde{f})]/\sigma_i^2. \tag{41}
\]

In the case of well-sampled images we had \( \varphi = \tilde{g}^* \tilde{g} \tilde{f} \). The extra convolution with the transform of the pixel pattern \( \tilde{c}_i \) above results in aliasing.

Consider first the situation where the distortion is negligible and the pixel grids are aligned with the coordinate axes. The transform \( \tilde{c} = c_i(\mathbf{k}) \) of the comb function \( c_i(\mathbf{r}) \) is then also a comb function:

\[
c_i(\mathbf{k}) = e^{i\mathbf{k} \cdot \Delta \mathbf{x}_i} \sum_n \delta(\mathbf{k} - \Delta \mathbf{k} \mathbf{n}) \tag{42}
\]

where \( \Delta \mathbf{x}_i \) is the offset of the origin of the comb, \( \Delta \mathbf{k} \equiv 2\pi/\Delta x \), with \( \Delta x \) the pitch of the comb, and where \( \mathbf{n} \) is a vector with integer valued components. If the seeing disk is much larger than the pixel size then \( g(\mathbf{k}) \) is very compact — width \( \sim 1/\text{FWHM} \ll \Delta \mathbf{k} \) — and the prefactor \( g^*(\mathbf{k}) \) in (41) then limits the contribution to \( \varphi_S(\mathbf{k}) \) to the \( \mathbf{n} = 0 \) component, and we have \( \varphi_S(\mathbf{k}) \simeq \tilde{g}^* \tilde{g} \tilde{f} \), or equivalently \( \varphi_S(\mathbf{r}) \simeq g^\dagger \otimes g \otimes f \). If the PSF is not well sampled, however, the \( \mathbf{n} \neq 0 \) terms contribute and result in aliasing.

For a single image, the aliased and true signal are roughly equal at the ‘folding frequency’ \( k_{\text{fold}} = \pi/\Delta x \). For Kolmogorov seeing the OTF is \( g(\mathbf{k}) = e^{-k^2/2} \) with structure function \( S = 6.88(\lambda k/2\pi r_0)^{5/3} \), where \( r_0 \) is the Fried length. The seeing width is \( \text{FWHM} \simeq 1.0 \times \lambda/r_0 \) so the OTF is \( g(\mathbf{k}) = \exp(-0.16(k \text{FWHM})^{5/3}) \) or, at the folding frequency, and expressed in decibels, the atmospheric attenuation is

\[
10 \log_{10}(g(k_{\text{fold}})) = -4.68 \left( \frac{\text{FWHM}}{\Delta x} \right)^{5/3} \text{dB}. \tag{43}
\]
Thus, if the pixel size is half the seeing width, for example, these spatial frequencies are suppressed by about 15dB in amplitude (30dB in power). An often used rule of thumb is that a pixel size of $1/3$ of the seeing FWHM is required to adequately sample the seeing. The above formula shows that with this sampling, the modes at the folding frequency that are strongly affected by aliasing are in fact attenuated by about 30dB in amplitude, so this is quite a conservative sampling rate.

For multiple images, the effect of aliasing becomes weaker. This is because the contributions to $\varphi(k)$ contain phase factors $e^{i2\pi n \Delta x}$. The zeroth order ($n = 0$) contributions simply add with phase factor of unity, while the dominant aliased components $n_x, n_y = \pm 1$ yield a net amplitude which is smaller by a factor $\sum e^{i2\pi n \Delta x}/\Delta x$ which has expectation value $1/\sqrt{\text{images}}$. Thus, simply by accumulating lots of images one will eventually overcome the aliasing effect. However, the convergence is quite slow. This is illustrated in figures 1 2. The first of these has seeing FWHM twice the pixel size, corresponding to very good seeing (FWHM = 0\".6) on Mauna Kea and 0\".3 pixel size. The difference between the optimal image (d) and the doubly convolved original (e) is very small, and as expected, appears at around the folding frequency. The lower panels show the result of applying a deconvolving filter to give a flat noise spectrum representation. The effect of aliasing is clearly extremely small for pixel size one half of the FWHM, but very strong for the second case where the pixel size and seeing FWHM are equal. This would correspond to FWHM = 0\".3, which is unheard of even on Mauna Kea.

We assumed above that the field distortion was negligible and that the pixel grids were aligned with the coordinate axes. In general this will not be the case. However, in any small patch of the image we can take the pixel coordinates to be a linear transformation applied to a regular Cartesian grid and it is not difficult to show that the main results are quite general.

### 4.3 Flat Source Normalization

.......

### 4.4 Photometric Accuracy

.......

### 4.5 Likelihood for Image Differences

This analysis is readily extended to the problem of detecting transients. Say we have one collection of samples $f_{p1}$ at one epoch and another set of images $f_{p2}$ at another. Information about changes in the true object scenes $f_1, f_2$ at these two epochs are encoded in the joint conditional probability distribution $P(f_1, f_2|f_{p1}, f_{p2})$. But since the noise fluctuations in the sets of images are independent this factorizes, so $P(f, f'|f_p, f'_p) = P(f|f_p)P(f'|f'_p)$, each of which is fully determined by the appropriate cumulated image.

If we set $f'(x) = f(x) + d(x)$ and integrate (or ‘marginalize’) over the distribution of static image $f(x)$, we find
Figure 1: Aliasing in the optimal summed image for ‘good seeing’ conditions. Panel a shows the commonly used test image of Lena. Panels b, c show two of 16 samples of this which were made by convolving with Kolmogorov seeing; applying a random shift with interpolation; re-binning $5 \times 5$. The FWHM of the seeing was twice the final pixel size. Panel d shows the optimal sum of the 12 images, and panel e shows the original a convolved twice with the PSF (seeing and pixel response). Panel f shows the difference highly amplified. Panel g shows the result of dividing the image by $\max(|g_{\text{fs}}(k)|, 0.01)$ in Fourier space to give a ‘flat-spectrum’ representation of the optimal image (though in this example the image is noise-free). Panel h shows the same filter applied to e. The low level residuals seen in panels f, i are the result of aliasing.
Figure 2: Aliasing in the optimal summed image for ‘exceptional seeing’ conditions. As in figure 1 except that here the FWHM was equal to the pixel size. The effect of aliasing is now much stronger.
that the resulting probability distribution for the difference image vector \( \mathbf{d} \) is
\[
P(\mathbf{d}) \propto \exp \left( -\frac{1}{2} (\mathbf{d} - (\mathbf{f}_2 - \mathbf{f}_1)) \cdot \frac{A_1 A_2}{A_1 + A_2} \cdot (\mathbf{d} - (\mathbf{f}_2 - \mathbf{f}_1)) \right).
\] (44)

This says that the most likely difference image is just \( \overline{\mathbf{d}}(\mathbf{r}) = \overline{\mathbf{f}_2}(\mathbf{r}) - \overline{\mathbf{f}_1}(\mathbf{r}) \). Note that if we have a much larger number of images in collection 1 than in collection 2 — as would be the case if 1 is the accumulated image and collection 2 are the most recently collected set of images — then \( A_1 \gg A_2 \), so \( A_1 A_2/(A_1 + A_2) \to A_2 \), and the width of the distribution function is determined solely by the uncertainty in the most recent image.

## 5 Other Considerations

### 5.1 Cosmic Ray Rejection

What we are proposing here is that the accumulated image is made simply by averaging the convolved images. This is not usually done, since this averaging is not robust against cosmic rays, which can give very high counts. More commonly, some kind of median averaging or averaging after rejection of outliers is performed. This presents something of a problem for the algorithm presented here where the convolution of the source images with their PSFs will smear out cosmic ray hits, and this makes them harder to reject. Provided the images are reasonably well sampled it is usually possible to distinguish cosmic ray hits from real sky features in the source images. However, there may be applications where other considerations prefer quite poor sampling. A better solution, and that adopted by the Pan-STARRS/POI project, is to make simultaneous multiple images with several telescopes. This is clearly a great advantage for a project that emphasises detection of transient, variable or moving objects. This provides very good rejection of cosmic rays; a feature that appears in the sum of 4 images as say a 5-sigma event will, if it is really a cosmic ray or other artefact, appear at the 10-sigma level in just one of the source images, and so will be relatively easy to remove. Also, once one has a reasonably well determined cumulative image we will in fact be able to perform avsigclip filtering; this will be necessary to remove e.g. low-altitude satellite trails which will not get removed by the primary CR filter.

### 5.2 PSF Modeling and Anisotropy Nulling

The PSFs in real telescopes vary with position on the focal plane. This is easy to cope with provided that the variation is smooth on scales at least as large as the mean separation between moderately bright stars. In CFHT12K data we typically have 50-100 samples of the PSF over each 7' \times 14' chip, and we find that a simple linear polynomial (for each pixel of the PSF model image - clarify ??) is adequate. Note that from the point of view of obtaining an optimal summed image we do not really require an extremely precise estimate of the PSF. The PSF is used to weight the contributions from the various images. The loss of information if we make an error in the weighting scheme is, quite generally, only quadratic in the error.

By the same measure, one can, if one wishes, rotate the PSF by 90° before convolving. Provided the PSF is not too anisotropic this will not result in much information loss. The reason that one might want to do this is that is will null out quadrupole anisotropy in the PSF introduced by e.g. telescope oscillations. This does not
work for all anisotropies — coma aberration, for instance, leaves a residual — but it is still a major advantage for weak lensing observations. This does require an accurate model for the PSF.

5.3 Bright Objects

We have assumed throughout that the objects we are trying to photometer are faint as compared to the glow from the earth’s atmosphere. The reason for making this restriction is that the variance in the pixel count $\sigma_p^2$ then becomes independent of the signal. The analysis can be generalized to avoid this, but then the matrix $A$ defined in (26) becomes dependent on the object scene and it is no longer true that the data enter in the likelihood only in the image $\varphi(\mathbf{r})$. Whether this is important depends to some extent on the application and also on how well one can really model PFSs etc; bright objects permit, in principle, very accurate measurements, but only if systematic errors can be kept very small. Perhaps the best solution for LSST projects is simply to keep all of the pixels in and around bright objects. This is a very small fraction of the sky so the extra cost of storage is negligible.

6 Pipeline Simulation

To test the likely performance of this optimal image combination and subtraction scheme we have implemented a pipeline, or rather a pair of pipelines, to generate test data with realistic pixel size and noise levels etc and to perform the basic pipeline reduction tasks.

We took as our model for the intrinsic sky one of the HDF images. We convolved these with Kolmogorov seeing with a realistic log-normal distribution of seeing widths. We then mapped this onto the ‘focal plane’ with random shifts and small rotations (to recognize the fact that we cannot point telescopes precisely). We added fake cosmic rays of random amplitudes (typically close to noise level as these are the most difficult to treat). We then sampled these with $0''.32$ pixel size and added Gaussian random noise to simulate the sky noise level expected for 60s integrations in the $R$-band. In the real images, we will have a large number of stars for registration and PSF measurement. Here we only simulate about 1 square arc-minute of data, so we added a grid of artificial point sources before applying seeing, with magnitudes in the range $m_R = 20 - 21$. The source image and an example simulated data image are shown in the upper panels of figure 3.

These images were generated in sets of four with identical seeing, but with different shifts, rotations, noise and cosmic rays to simulate the POI’s simultaneous imaging. These were processed as follows; stars were detected with ‘findpeaks’ and used to determine the mapping from focal plane to sky coordinates. These stars were also used to determine the PSF. The images were re-convolved and then mapped back to sky coordinates, and a median of each set of four taken to remove cosmic rays. The resulting stream of images were then averaged. The maximum signal to noise image is shown in the lower left corner of figure 3, along with a ‘flat-noise spectrum’ version which has had high frequencies boosted to render the sky noise uncorrelated. Figure 4 shows the high quality of image subtraction that is possible with matched filtering.

The results of these simulations are similar to the quality of image warping and PSF modeling that we have been able to achieve routinely in reducing e.g. CFH12K data. The main difference here is that we are simulating short exposures with relatively small telescopes, so there is more noise in the star positions and shapes. This does not
Figure 3: Upper left panel shows one WFPC2 image from the HDF. Upper right shows a simulated CCD image generated as described in the text. A few faint cosmic rays are visible as elongated streaks. The lower-left panel shows the optimally weighted combination of about 100 such exposures, and the bottom right panel shows the same image filtered to give a flat noise spectrum — this is useful for e.g. simple photometry packages that assume the sky noise is uncorrelated.
Figure 4: Illustration of subtraction of a PSF-matched image. The left panel shows one of the fake exposures from the simulation. The image has been convolved with its PSF (as this is the optimal image for detecting point sources). We chose one of the very best seeing images as these are the hardest to match. The center panel shows an image generated from the accumulated image which has been filtered in Fourier space to match the PSF. The right hand panel shows the subtraction, which is extremely good.

seem to have impaired the results. The image subtraction for very bright saturated objects in the real images is, not surprisingly, rather poor, but such objects are not used either for registration or PSF modeling.
7 Conclusions and Discussion

We have shown that there is a clearly defined optimal way to combine images of differing seeing and noise level (34). This is the quantity that the pipeline in LSST projects should generate.

Review of results ....

Acknowledgements....

References


