Mon. Not. R. Astron. Soc. **296**, 339–346 (1998)

An optimal extraction algorithm for imaging photometry

Tim Naylor
Department of Physics, Keele University, Staffordshire, ST5 5BG

Accepted 1997 November 12. Received 1997 October 21; in original form 1997 September 8

ABSTRACT

This paper is primarily an investigation of whether the 'optimal extraction' techniques used in CCD spectroscopy can be applied to imaging photometry. It is found that using such techniques provides a gain of around 10 per cent in signal-to-noise ratio over normal aperture photometry. Formally, it is shown to be equivalent to profile fitting, but offers advantages of robust error estimation, freedom from bias introduced by mis-estimating the point spread function, and convenience. In addition some other techniques are presented, which can be applied to profile fitting, aperture photometry and the 'optimal' photometry. Code implementing these algorithms is available at http://www.astro.keele.ac.uk/~timn/.

Key words: methods: data analysis – techniques: image processing – techniques: photometric.

1 INTRODUCTION

Most astronomical photometry is now carried out using imaging detectors, which normally provide spatially resolved profiles of several stars at once. This allows one to measure the brightness of a star using just those parts of the profile which provide the best signal-to-noise ratio, and to correct for the resulting loss of flux by examining the profile of a brighter star or stars. Time-series differential aperture photometry is foremost amongst these techniques. In its simplest form the target star and a 'local standard' star are measured in each frame, using an aperture which maximizes the signal-to-noise ratio for the target. Dividing the flux of the target star by the local standard then corrects for the lost flux, and a light curve of the target with respect to the local standard can be constructed. More complex schemes using several local standard stars exist (e.g. Gilliland & Brown 1988 for cases where colour terms are important, and Honeycutt 1992 for cases where the local standards available differ from image to image), but for simplicity I shall refer throughout the paper to dividing by a single star. Aperture photometry will not work, however, if there is another star whose flux contaminates the aperture of one of the stars. Solving this problem has led to profile fitting, in which the sky-subtracted flux is modelled as a series of point sources, each spread across the detector according to the point spread function (PSF). This technique allows overlapping profiles to be disentangled. Profile fitting has two distinct forms. In the first an analytical PSF is fitted to the data, and the total flux is extracted from the normalization (e.g. Penny & Dickens 1986). This technique has the advantage of speed (the functions have analytical derivatives), but it is unclear what the effect of the mismatch between the observed and model PSF is. This problem can be partially solved by observed profile fitting (e.g. Stetson 1987), but the problem here is that the centre of a star can fall at any point within a pixel. This means that the PSF has to be able to be sampled on to different grid positions, which leads to several stars being required to define it.

In Shahbaz, Naylor & Charles (1994) we speculated on the possibility of there being an 'optimum' way of extracting the information from imaging photometry. An analogous problem exists in long slit spectroscopy, where the obvious way to extract a spectrum is to add those columns of the detector which contain the light from the target, and subtract the background light. It has been known for over a decade, however, that the way to ensure the maximum signal-to-noise ratio is to weight the columns according to the fraction of the signal they contain (Hewett et al. 1985; Horne 1986; Robertson 1986). Since the improvement in signal-to-noise ratio should scale as the power of the number of dimensions, more significant improvements can be made for photometry than for spectroscopy. In this paper, therefore, I address the question of whether such a scheme can be applied to imaging photometry, and if it offers significant advantages over aperture photometry and profile fitting. At the same time I discuss various improvements that can be applied to all methods of imaging photometry.

My own interest is primarily time-series photometry, where the brightness of the local standard is already known, or relative measurements of the target are sufficient for the scientific programme. In this rather limited case, the techniques presented here are a complete solution to the photometry problem. Alternatively, if the scientific programme involves transferring the photometry to an apparent magnitude system, then the differential magnitudes derived using the algorithms presented here must be corrected for aperture losses by curve-of-growth techniques, and for bandpass differences between the system used and the standard system. Such corrections are discussed in the standard texts and elsewhere (e.g. Stetson 1990; Miles 1998), and I will not discuss them further here.

The rest of the paper is laid out as follows. The aim of Section 2 is to present a derivation of the weights required for the optimal scheme. I will develop the notation by considering basic aperture photometry, which will then allow a comparison between the two methods. Section 3 is a consideration of practical issues, including determining where the star is in the frame, what the sky level is, and finding a robust way of estimating the individual pixel variances. Finally I will present a comparison between the optimally weighted extraction and profile fitting.

2 THE THEORY

Consider a two-dimensional detector, where the positions are are quantized (usually into pixels), and where the values at these positions are uncorrelated in the statistical sense. The noise in pixel (i,j) is given by the square root of its variance, $V_{i,j}^{\frac{1}{2}}$, which usually has contributions from three different sources. Most obviously there will be photon noise from sources and sky. Secondly, there will be noise from the detector and associated electronics, such as readout noise and noise associated with the dark current. Finally, there may be noise resulting from data reduction problems, such as imperfect flat-fielding. The flux level in each pixel I will refer to as counts, although for calculating variances, the conversion from counts to detected photons (the gain) needs to be known.

2.1 The notation

Consider a telescope which delivers an image to the focal plane the flux of which at any point is defined by the PSF $P(x-x_0, y-y_0)$, where x and y are positions in the focal plane, and (x_0, y_0) are the coordinates of the centre of the profile. I assume that P is independent of x_0 and y_0, i.e. does not vary with the position of its centre in the focal plane; this is a requirement for small aperture photometry. The integral of P over all area is defined to be one, so that the profile for a star with a total of T counts is $TP(x-x_0, y-y_0)$. The flux that is detected from the star in a pixel whose coordinates are (i,j) is

$$f = \int_{x=i-\frac{1}{2}}^{x=i+\frac{1}{2}} \int_{y=j-\frac{1}{2}}^{y=j+\frac{1}{2}} TP(x-x_0, y-y_0)\mathrm{d}x\mathrm{d}y$$
$$= T\int_{\mathrm{pix}(i,j)} P(x-x_0, y-y_0)\mathrm{d}A = TP_{i,j}^{\mathrm{I}}; \qquad (1)$$

where the limit pix(i,j) is taken to mean the integral over the area of pixel (i,j). I define this integral to be $P_{i,j}^{\mathrm{I}}$, where the superscript is a reminder that this is the integral of the PSF over a given pixel. An important alternative interpretation of $P_{i,j}^{\mathrm{I}}$ is that it is the fraction of light of the star which falls in pixel (i,j). For simple aperture photometry these integrals are summed with some condition such as

$$(i-x_0)^2 + (j-y_0)^2 < R^2, \qquad (2)$$

where R is the software aperture radius. Thus the measured number of counts is

$$F = T\sum_{i,j} P_{i,j}^{\mathrm{I}}. \qquad (3)$$

Here the sum is over all pixels (i,j) which satisfy the criterion in equation (2). If the pixellation is small compared to the PSF, i.e. the spatial profile is well sampled, then the sum tends to an integral, and the equation simplifies to

$$F = T\int_R P(x-x_0, y-y_0)\mathrm{d}A, \qquad (4)$$

where the integral is taken over the area of the aperture, in a definition similar to that in equation (2).

2.2 The aperture size

Equation (3) allows us to derive the well known result that, for

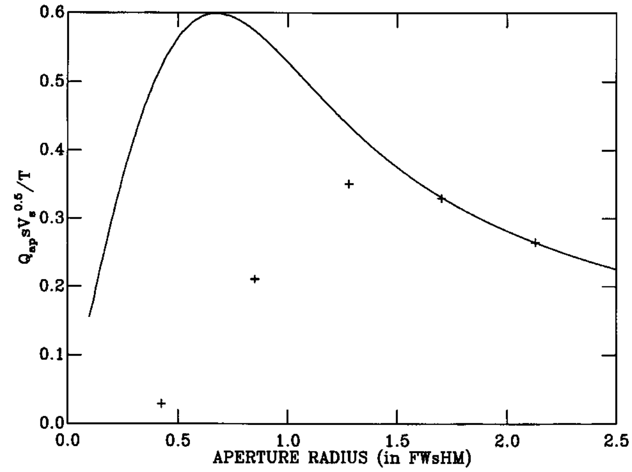

Figure 1. The solid line shows the theoretical signal-to-noise ratio for a well sampled Gaussian PSF as a function of aperture radius. The calculation is for a faint star, such that the noise is dominated by counts from the sky. The crosses are the signal-to-noise ratio obtained for data from the Keele 0.6-m telescope with an ST-6 CCD camera. The FWHM seeing was approximately 2 pixel, and a simple aperture was used.

comparing stars within the same frame, the aperture need not encompass the entire flux from the star. For a given aperture size the integral will be the same for all stars in the frame, and acts as a multiplier to the total flux. Thus the ratio of fluxes between two stars is independent of the aperture size used.

In principle the noise in the aperture (and thus the ideal aperture size) is a simple matter of calculation. There are two extreme cases for the optimum size aperture. If the counts from the star are greater than V everywhere on the detector, i.e. the star is very bright, the best signal-to-noise ratio will be obtained by using an aperture which encompasses the entire detector. At the other extreme lies the sky-limited case, in which the counts from the sky within the aperture are much greater than the counts from the target, and thus the star is very faint. In this case a rather small aperture is required. Stars in the intermediate brightness regime require aperture sizes somewhat larger than the sky-limited case. What makes the sky-limited case mathematically appealing is that the noise is independent of the brightness of the target. The noise is given by the noise per pixel, multiplied by the square root of the number of pixels, and thus the signal-to-noise ratio is

$$Q_{\mathrm{ap}} = \frac{T}{sV_{\mathrm{s}}^{\frac{1}{2}}}\left(\frac{F/T}{R_{\mathrm{ap}}\sqrt{\pi}}\right), \qquad (5)$$

where V_{s} is the variance for a single pixel containing only sky counts, s is the full width half-maximum (FWHM) seeing, R_{ap} is the aperture radius in units of the seeing, and F is the measured flux within the aperture. The term in square brackets depends upon the shape of the PSF, which for simplicity I will assume to be Gaussian, allowing numerical integration to give the curve shown in Fig. 1. This shows that the optimum signal-to-noise ratio is obtained for an aperture the radius of which is two-thirds of that of the FWHM seeing, when the numerical term in square brackets is 0.60.

In practice, since P is unknown, for time-series photometry it is usual to find the ideal aperture size by trial and error using a constant star of similar magnitude to the target. The scatter of the relative magnitude of this star between the frames of the run yields the signal-to-noise ratio for each aperture, allowing the ideal aperture to be chosen (e.g. Howell 1989). This results in a curve similar in shape to the theoretical results shown in Fig. 1. The shape of the curve is normally explained as follows. Small apertures encompass too few photons from the star to yield a good signal-to-noise ratio, whilst large apertures contain the noise from many sky pixels. However, there are two frequently ignored contributions to this curve. First, inaccuracies in estimating the sky brightness can contribute to the noise for large apertures. The second contribution is a result of the ideal aperture normally being quite small, such that the 'fine pixellation' assumption between equations (1) and (3) is not true. This means that one must return to equation (1) where, unfortunately, the sum of the integrals depends on which pixels are summed, and hence exactly where the centre of the star falls with respect to the pixel grid. This variability in the sum of integrals I shall refer to as aperture noise (Shahbaz et al. 1994), since it can also be viewed as a noise inherent in re-sampling the pixels into the aperture. The effect of aperture noise is shown by the crosses in Fig. 1. They are the measured signal-to-noise ratio for a simple aperture, with real data where the seeing was approximately 2 pixel FWHM. At large apertures the measurements match the theory very well, but the signal-to-noise ratio falls short of that for fine pixellation if the aperture is smaller than 1.5 pixel.

Aperture noise can be ameliorated, although not cured, by calculating what fraction of a pixel close to the aperture edge lies within the aperture, and adding only that fraction to the total. Practically this is achieved by introducing weighting factors for each pixel ($\omega_{i,j}$) such that

$$F = \sum_{i,j} \omega(i - x_0, j - y_0)(D_{i,j} - S_{i,j}) = \sum_{i,j} \omega_{i,j}(D_{i,j} - S_{i,j}), \qquad (6)$$

where $D_{i,j}$ and $S_{i,j}$ are the data value and estimated sky value for each pixel. Since the flux from the star is changing across the pixel, such a 'soft-edged' aperture is still only an approximation to the flux within an aperture before the data were sampled on to the pixel grid. Thus, the weighting factors used can be very simple (Da Costa 1992), such as

$$\omega(i - x_0, j - y_0) = \begin{cases} 1, & \text{if } (i-x_0)^2 + (j-y_0)^2 < (R-0.5)^2; \\ 0.5 - R + \sqrt{(i-x_0)^2 + (j-y_0)^2}, & \text{if } (R-0.5)^2 < (i-x_0)^2 + (j-y_0)^2 < (R+0.5)^2; \\ 0, & \text{if } (i-x_0)^2 + (j-y_0)^2 > (R+0.5)^2. \end{cases} \qquad (7)$$

2.3 Optimum weighting

Equation (6) can be generalized to an optimum weighting scheme, following Pratt (1978) (see also Hewett et al. 1985, Horne 1986 and Robertson 1986). If the signal in a pixel ($D_{i,j} - S_{i,j}$) is divided by the expected fraction of the total flux in that pixel ($P^{\mathrm{I}}_{i,j}$), then it represents a measure of the total flux. Thus each pixel can be used as independent estimator of the total flux in the profile. To combine independent estimates of the same value in the most advantageous way, one weights them by the inverse of the square of their errors, which in this case is just $(P^{\mathrm{I}}_{i,j})^2/V_{i,j}$. To keep the normalization correct one must divide by the sum of weights, which gives

$$F = \frac{\sum_{i,j} P^{\mathrm{E}}_{i,j}(D_{i,j} - S_{i,j})/V_{i,j}}{\sum_{i,j} (P^{\mathrm{E}}_{i,j})^2/V_{i,j}}, \qquad (8)$$

where the estimated PSF P^{E} has been used instead of the true PSF P^{I}. This equation can be simplified to

$$F = \sum_{i,j} W_{i,j}(D_{i,j} - S_{i,j}), \qquad (9)$$

where the weight mask, $W_{i,j}$, is defined as

$$W_{k,l} = \frac{P^{\mathrm{E}}_{k,l}/V_{k,l}}{\sum_{i,j} (P^{\mathrm{E}}_{i,j})^2/V_{i,j}}. \qquad (10)$$

The variance of the measured flux is given by

$$\mathrm{var}[F] = \sum_{i,j} W^2_{i,j} V_{i,j}. \qquad (11)$$

Since the optimum weight mask is not, in general, equal to a function similar to that given in equation (7), this proof shows that aperture photometry can never extract all the available signal-to-noise ratio, even with a carefully chosen soft-edged aperture.

An intuitive interpretation of equations (9) and (10) is that optimal extraction is the ultimate soft aperture. The technique is similar to summing the flux into a series of annular apertures, and then adding these apertures with optimal weights. This preserves the link with conventional aperture work, and makes it obvious that if the weights are the same for two stars in the image, the ratio of fluxes will always be correct, by an argument analogous to that set out in Section 2.2. I shall return to this point in a more formal mathematical sense in Section 3.3. Yet another possible interpretation is that the mask is the optimal filter for the data, similar to the filters used for source detection.

The signal-to-noise ratio for a star extracted with optimal weights is given by equations (9), (10) and (11) as

$$Q_{\mathrm{opt}} = F\sqrt{\sum_{i,j} (P^{\mathrm{I}}_{i,j})^2/V_{i,j}}. \qquad (12)$$

If I assume the sky-limited case, and fine pixellation, the sum becomes an integral of the PSF squared, which is analytic for a Gaussian. I also replace F with T, since it should be a good estimate of T, to obtain

$$Q_{\mathrm{opt}} = \frac{T}{sV_{\mathrm{s}}^{\frac{1}{2}}}\sqrt{\frac{2\ln 2}{\pi}}. \qquad (13)$$

This equation can be compared with (5). In this case the numerical term is 0.66, compared directly with the 0.60 obtained for aperture photometry. This shows that the optimum extraction offers an increase in signal-to-noise ratio of about 10 per cent for well sampled data.

3 PRACTICAL CONSIDERATIONS

The obvious way to proceed at this point is to use equations (9) and (10) to make an optimal extraction. The quantities needed are the sky level, S, and variance for each pixel, V, and an estimate of the PSF, P^{E}. Calculating the first two of these quantities is straightforward (see Sections 3.1 and 3.2), and P^{E} can be found by fitting a bright, isolated star. In practice an analytical function cannot give a perfect match between P^{E} and P^{I}, and the effects of this mismatch are discussed in Section 3.3. As I will show, this leads to the modification of the variances to be used in equation (10) to those given in equation (20). With this issue clarified one can then proceed to fit a star to determine the PSF P^{E} (see Section 3.4), which is then used to estimate the counts from the star. Bringing the whole procedure together is outlined in Section 3.6.

3.1 Sky determination

Sky determination is crucial for stars for which the majority of the counts in the aperture come from the sky, not the target. This problem is particularly acute for infrared photometry, where the sky is about 12th magnitude per square arcsecond at K. This means that for a star of say K=17, the sky must be determined to a fraction of a per cent to avoid errors in the sky determination dominating the error budget. It is straightforward to estimate the sky area that must be used for aperture photometry, by assuming that the sky is determined from the mean of the sky box. If the sky is to contribute no more than 5 per cent of the error budget, then the total sky counts within the aperture must be estimated ten times better than the total flux in the aperture is measured. The estimated sky level has the greatest effect when the star is faint, such that all the noise in the aperture will originate from the sky counts within it. If there are N_{A} pixels in the aperture, and N_{S} pixels are used to determine the sky, I can calculate the noise from the sky determination (S_{e}) and apply the condition thus,

$$S_{\mathrm{e}} = N_{\mathrm{A}}\sqrt{V/N_{\mathrm{S}}} < \sqrt{VN_{\mathrm{A}}}/10; \tag{14}$$

which leads to the conclusion that the ratio of sky to star pixels should be about 100. All useful sky determination methods are somewhat noisier than taking the mean, so that a ratio of 200 is normally close to the required accuracy. To repeat this calculation for the weighted extraction, I assume the variances in equations (10) and (11) are all equal, which gives

$$S_{\mathrm{e}} = \sqrt{V/N_{\mathrm{S}}}\frac{\sum_{i,j} P^{\mathrm{I}}_{i,j}}{\sum_{i,j} (P^{\mathrm{I}}_{i,j})^2} = \frac{\sqrt{V/N_{\mathrm{S}}}}{\sum_{i,j} (P^{\mathrm{I}}_{i,j})^2} < \frac{\sqrt{V}}{10\sqrt{\sum_{i,j} (P^{\mathrm{I}}_{i,j})^2}}, \tag{15}$$

and hence

$$N_{\mathrm{S}} > \frac{100}{\sum_{i,j} (P^{\mathrm{I}}_{i,j})^2}\,. \tag{16}$$

Once again I assume a Gaussian to integrate $(P^{\mathrm{I}})^2$, which gives the result that the number of pixels in the sky should be a hundred times $s^2\pi/2\ln 2$ (where s is FWHM seeing), or perhaps twice this, depending on the sky determination method.

As has been discussed in many other papers, the sky should be measured from a region placed symmetrically about the target, so that any gradients in the sky are corrected for to first order. The combination of this and the above size constraint makes it almost inevitable that there will be stars within the sky area. Worse still, even if all the detected stars could be excised, faint, undetected stars will still bias the sky determination. Most of the methods suggested to date use the fact that the modal pixel value in the sky box should be a good estimator of the sky even if stars are present. Practically, the mode is very noisy, as shown by the histogram of sky values in Fig. 2. This has led to the scheme suggested by Da Costa (1992) which uses iterative clipping and the approximate formula that the mode is three times the median less twice the mean. This takes advantage of the fact that both the median and the mean require a much smaller number of pixels than the mode for an accurate determination. An alternative approach is to fit the histogram in Fig. 2, as suggested by Bijaoui (1980) and Irwin (1985). In this case I use a skewed Gaussian of the form

$$N = \begin{cases} N_0 \exp\left[\!\left[-\ln 2\left\{\frac{\ln[1+\gamma(D-S)]}{\beta}\right\}^2\right]\!\right], \\ \qquad \text{where } \gamma = \sqrt{\ln 2}\,\frac{(1-\mathrm{e}^{-\beta})}{V_{\mathrm{s}}} \quad \text{if} \quad D-S>0; \\ N_0 \exp\left[\!\left[-\left\{\frac{\ln 2[(D-S)]}{V_{\mathrm{s}}}\right\}^2\right]\!\right], \\ \qquad\qquad\qquad\qquad \text{if} \quad D-S<0. \end{cases} \tag{17}$$

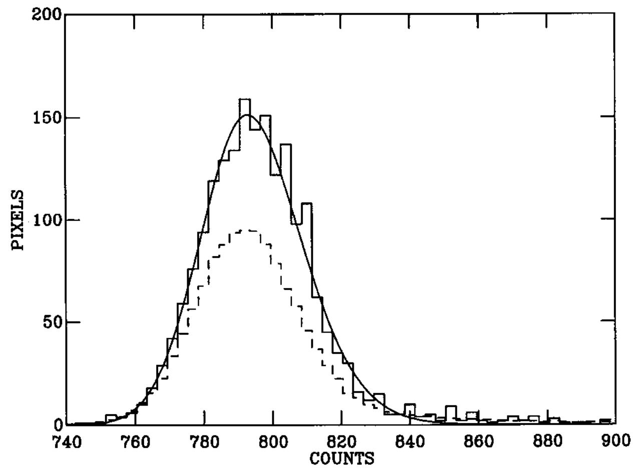

Figure 2. The solid histogram is the number of pixels with a given number of counts measured from a region of a CCD image of size determined by equation (16). The solid curve is a fit to the same data, used to determine the sky level. For comparison, the dotted histogram is for a sky region centred on the same point as the solid histogram, but of a much larger area.

Here N is the number of pixels with a given number of counts D. The constants being fitted are the sky level S, V_{s} and β. β is the skew, which must always be positive, and gives the extended tail to the right. The fitting process has the advantage that it will give errors for the sky determination, and perhaps more importantly, an accurate estimate of the variance of the sky, which is the constant V_{s}.

Although the above technique offers its best gains in the infrared, there is one aspect of IR data reduction which can make this problematical. For point source photometry, a standard technique is to move the source around the array (by moving the telescope) between exposures. This results in a group of frames, each with the stars in a different position. A flat-field is then created by first normalizing each frame in the group to the same mean counts, and then making a new frame where each pixel is the median of that pixel's value in all the frames. If this is then used to flat-field one of the frames in the group, roughly 1/nth of the pixels in the flat-fielded frame will have the same value (where n is the number of frames in the group). This is because those pixels have effectively been divided by their own value multiplied by the normalization of the flat-field. Aside from spoiling the statistical properties of the sky noise, this will introduce a sharp spike in the sky histogram, making the fitting impossible. Fortunately the solution is simple (if tedious to code). Each frame must be flat-fielded with the median of all the frames in the group excluding itself. Since a median requires an odd number of frames, this means that the group, often referred to as a 'jitter pattern' should consist of an even number of images.

3.2 Noise determination

Ab initio calculations of the error in a photometric measurement require a knowledge of all the sources of noise in the detector, and as a result often underestimate the true error. As described in Shahbaz et al. (1994), it is possible to use the variance of the sky to improve the error estimate. In that paper it was assumed that there was an extra term in the expression describing the noise, which was linear in counts, i.e. of the form expected if there were a noise associated with flat-fielding. However, there are computational advantages to assuming it is additive, i.e. of the form expected for an extra readout noise. In practice, for faint stars, it makes no difference which form is used, since the excess counts in the star aperture are too small to affect a term linear in counts. The advantage of an additive term is that if the variance of the sky ($V_{i,j}$) is measured by the histogram technique described above, the total noise in any pixel is given by

$$V_{i,j} = V_s + \frac{(D_{i,j} - S_{i,j})}{\sqrt{g}}, \tag{18}$$

where g is the number of detected photons per count, i.e. the gain. Obviously this error estimate is independent of errors in the measurement of dark current or readout noise. It is also more robust against small errors in knowledge of the gain than methods which require the number of photons in the sky. The only time it breaks down is for bright stars where the sky makes little or no contribution to the noise budget, and the gain has been incorrectly measured.

3.3 The weights

In general, there will be a mismatch between the estimated PSF P^{E}, and the actual PSF P^{I}, which means that equations (9) and (10) will not provide a good estimate of the flux. In this section I shall show how the effects of this mismatch can be eliminated. This immunity from the effects of mis-estimating the PSF is one of the key advantages the optimal extraction has when compared with profile fitting. If there is a profile mismatch, the estimated flux is given by substituting the actual PSF for $D_{i,j} - S_{i,j}$ in equations (9) and (10). Thus,

$$F = T\frac{\sum_{i,j} P^{\mathrm{E}}_{i,j} P^{\mathrm{I}}_{i,j}/V_{i,j}}{\sum_{i,j} (P^{\mathrm{E}}_{i,j})^2/V_{i,j}}. \tag{19}$$

If the two sums are identical for all stars, then there is a direct analogy with the argument for aperture photometry used in Section 2.2, in that although the absolute flux may be wrong, the flux ratio of any two stars is correct. Unfortunately this is not the case; the variances appear in the sum, and as these depend on the counts in each pixel, they will be different for bright and faint stars. For high-precision work the problem is even more acute, since two stars with the same number of total counts will, due to noise, have slightly different counts in each pixel, and hence slightly different variances. The only solution to the problem is to use the same weights for the extraction of all stars. Clearly this weight will not be optimal for all the targets, but the way to ensure the most accurate flux ratios is to use the correct weights for the faintest star of interest. If the flux for this star is established as approximately F' from fitting (see Section 3.6), then

$$V_{i,j} = V_s + \frac{F' P^{\mathrm{E}}_{i,j}}{\sqrt{g}}. \tag{20}$$

If the very faintest stars of interest are sky-limited, as is often the case, then the variances for equation (10) are simply V_s, which gives it a particularly simple form.

The other practical consideration for the weight mask is where to terminate it, since mathematically the weights run out to infinity. The mask should always be cut off before other stars make a significant contribution to the flux, or the edge of the image is reached. If there is no such limitation, I normally terminate the mask at a radius of two FWHM, beyond which adding further pixels rarely makes any difference to the signal-to-noise ratio. Using the simple condition that the centre of the pixel must be within a given radius of the star centre will result in the equivalent of aperture noise for the optimal extraction. Since the pixels in question have very low weights, this is rarely a problem. Where very small radius masks are required, however, multiplying the weights, $W_{i,j}$, by the weights $\omega_{i,j}$ given by the soft aperture condition in equation (7), can give an improvement in the signal-to-noise ratio.

3.4 Determining the PSF

The profile of a bright, isolated star contains most of the information required to determine the PSF. The fundamental problem in extracting this information is how to resample the observed PSF on to the pixel grid of the star of interest. The obvious solution to this would appear to be to fit the bright, isolated star with an analytical function, specifically a two-dimensional Gaussian, and use this as P. This is, in fact, the correct procedure, but in adopting it one should be aware of the approximations that have been made. Most obviously, an analytical function will not be a perfect match to the observed PSF, but as I have already shown, this will not result in biased photometry. There is a more subtle effect, though, which is that the pixel values of the PSF star represent the PSF integrated over the area of the pixel P^{I}, rather than P. Once again, though, this simply results in a slightly imperfect estimate of the PSF, which does not present a problem.

So, a practical solution to estimating P is to fit a star with a two-dimensional Gaussian, for which the Marquardt algorithm works well. An initial estimate of the seeing is required, but since a bright star is being fitted the procedure is robust against even major errors in this estimate. Estimates of the position are also required, but these can be supplied from a star-searching routine. The only parameters of interest from the fit are those that determine the shape of P; after fitting, the normalization should be set to ensure that the integral of P is one. Next P must be integrated to P^{E} for the pixel grids of the stars of interest. Normally, this can be achieved to sufficient accuracy by evaluating P at 25 grid points spread uniformly across the pixel.

Having made a series of approximations to P, a crucial question is what fraction of the available signal-to-noise ratio has been extracted. This is equivalent to asking if the estimated PSF is close enough to the real PSF for our purposes, or if a more sophisticated model is required. Consider the χ^2 of the data with respect to the true PSF. This would simply be the number of pixels (N_{P}) over which the test was performed. Hence,

$$N_{\mathrm{P}} = \sum_{i,j} \frac{[FP^{\mathrm{I}}_{i,j} - (D_{i,j} - S_{i,j})]^2}{V_{i,j}}. \tag{21}$$

By expanding the squared term in this equation, and using equations (8) and (12) (substituting P^{E} for P^{I} and Q_{avail} for Q_{opt}), this leads to a general expression for the total signal-to-noise ratio available,

$$Q_{\mathrm{avail}} = \sqrt{\sum_{i,j} \frac{(D_{i,j} - S_{i,j})^2}{V_{i,j}} - N_{\mathrm{P}}}. \tag{22}$$

Evaluating this equation for a star for which the weight mask has been optimized allows a direct comparison between the signal-to-noise ratio available and that extracted, from which a decision can be made as to whether P^{E} is a good enough match to P^{I}. As stated above, normal practice is to weight for the very faintest stars, at which point the calculation becomes quite noisy, primarily because $D_{i,j} - S_{i,j}$ becomes very small. The simplest way to overcome this is to measure the available signal-to-noise ratio for a somewhat brighter star, which will yield a conservative estimate of how well the estimated PSF matches the data. Practically, a two-dimensional Gaussian seems to be an adequate profile, often extracting over 99 per cent of the available signal-to-noise ratio, except in cases of pathological tracking errors or poor telescope alignment.

The referee of this paper pointed out that it is crucial that the PSF star used is not a double star (or indeed an extended object). Thus it is best to restrict the choice of PSF star to those in a user-supplied list, from which a program chooses the brightest non-saturated star available on the detector. As an added precaution, several stars from a good seeing image can be fitted to check for consistency between their measured FWHMs.

3.5 Determining the star positions

Once the parameters of the shape of the PSF has been determined (for the two-dimensional Gaussian this means the FWHMs), one can fit the target stars in the same way as the PSF star, but with these parameters fixed. This provides very reliable measurements of the star positions, much better than those obtained from simple centroiding.

3.6 Using the algorithms

For single images the reduction procedure is straightforward.

(1) The brightest single non-saturated star in the image (the PSF star) is fitted to measure the parameters that define the shape of the point spread function, as described in Section 3.4. Integrating this over individual pixels defines P^{E} in equation (10).

(2) All stars in the image are then fitted with a PSF, the shape parameters of which are fixed, to determine their position (see Section 3.5).

(3) The normalization from the fit to the faintest star of interest is used to define F' in equation (20).

(4) Equations (20) and (10) are then used to define the weight mask for each star, which is then applied using equation (9) to obtain the flux estimate.

For time-series work, the situation is complicated by the fact that the sum of all the images can provide a very accurate position of the target star with respect to the local standard. Using these positions will provide more accurate flux estimates than using the position of the star measured in the single frame being analysed, for the following reason. Consider what happens if the centres of the weight masks of both the target and local standard are shifted by an equal amount from their respective star peaks. By the same argument as that used in Section 2.2, the fractional light losses are identical, leaving the relative flux unchanged. This allows one to remove errors due to the centring of the masks, even though the absolute positions of the stars may not be well determined. This is particularly important for eclipsing binaries in which the lowest fluxes are close to zero, where if the position of the star is determined separately in each frame, a fitting algorithm searching for the best position will systematically overestimate the flux (see Fig. 3).

The first step for time-series photometry is to find positions for all the stars in all the frames, by fitting a PSF defined by the brightest star in each frame. Practically, this can be achieved by using a star-search algorithm to give approximate positions and fluxes, which are then refined by the fitting process. These refined positions are used to define the offsets between each image, using a pattern-matching algorithm. For each star whose flux is to be measured, an approximate position is found in the first frame, and the offsets used to calculate an expected position in every subsequent image. If the difference between the expected and measured X and Y positions of each star is averaged over all frames, it can be used to correct the initial estimated position. This corrected position, in combination

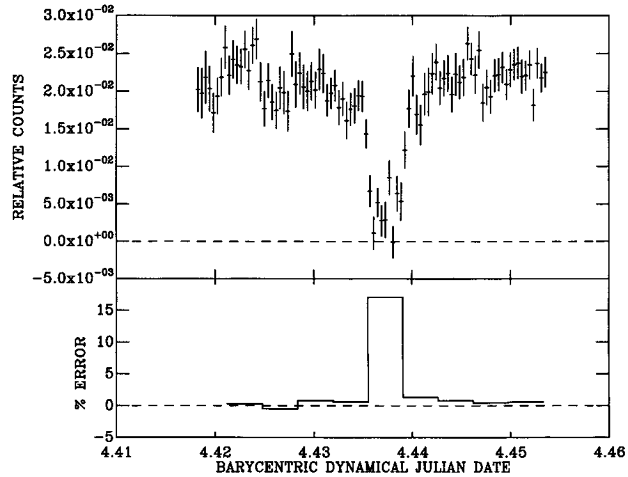

Figure 3. Upper panel: an eclipse of HT Cas observed with the Keele 0.6-m telescope and extracted with the optimal extraction algorithm, using the same relative positions of the stars for all frames. Lower panel: the difference between the flux extracted with the positions of the stars determined in each frame, and with fixed relative positions. The differences have been averaged into groups of nine frames for clarity. Note how the extraction with individually determined positions systematically over-estimates the flux, with large errors during eclipse.

with the offsets between the frames, provides an accurate position for each star in each frame. The brightest star in each frame is now fitted again to determine refined PSF parameters, but with its position fixed. These parameters, along with the accurate positions of the target stars, are all that is needed to calculate the weights for the optimal extraction.

Finally, it is worth stating explicitly that the PSF star gains no special significance. It can be used as a local standard star (indeed, since both local standards and PSF stars need to be bright, there are normally compelling reasons for doing this), or as a comparison star. Indeed, there is no reason why the PSF star cannot be the target star.

4 CROWDED FIELDS

Although it is normally stated that aperture photometry (and by implication optimal photometry) cannot be used where the profiles overlap, it is possible to use the PSF star to estimate the contamination from crowding stars. After extracting the flux from both the target and its contaminator in the normal way (resulting in the measured fluxes F_1^{m} and F_2^{m}) the weight mask is placed in the same position with respect to the PSF star, as the target has with respect to the contaminating star. This then gives an estimate of the contaminating flux as a fraction of the total flux in the contaminating star (K_1). Then the true flux in star 1 is given by

$$F_1^{\mathrm{t}} = F_1^{\mathrm{m}} - F_2^{\mathrm{t}} K_1. \tag{23}$$

An analogous equation exists for star 2, which allows a solution to be found for the true fluxes.

The method is not quite as simple in practice, since the position of the peak of a star is affected by the underlying wings of the companion. If both stars are fitted simultaneously, the derived position of the stars will only be affected by the wings of the companions in so far as the wings deviate from the analytical form of the PSF. This correction can be estimated by introducing an artificial star next to the PSF star, with the correct flux ratio. If these are fitted in an identical fashion to the two stars to be de-blended, the actual position of the star centres can be compared with that measured from the fitting, and hence the shift calculated. If the artificial star is a straight copy of the PSF star this presents an interpolation problem, since the two centres can only be separated by an integer number of pixels in X and Y. A practical solution to this problem is to measure the shifts with a copy of the PSF star being placed at the four integer positions which bracket the actual offset, and interpolate between them.

This means that for a single frame determining the flux is the following iterative procedure.

(1) Determine the analytical PSF by fitting the PSF star.

(2) Fit both stars simultaneously with this PSF, to make the initial estimate of the star positions.

(3) Extract the flux for both stars using the weight mask defined by steps 1 and 2.

(4) Estimate the contamination for each star by placing the weight mask at the appropriate position with respect to the PSF star.

(5) Use the solutions to equation 23 to estimate the corrected flux for each star.

(6) Place artificial stars next to the PSF star to calculate the correction to the measured position, and use these positions in step 3 to correct the flux.

For time-series photometry the actual separations of the stars need only be determined once, and hence the procedure is not iterative for subsequent frames, and so is much simpler.

Practically I have found that it is the interpolation of the position corrections which limits the accuracy of the technique. Where the gradient of the PSF is changing most rapidly, this interpolation becomes poor. The most obvious instance of this is where a star lies close to another star, and the detector poorly samples the PSF. The inaccuracies are strongly dependent on the PSF, and the relative brightness of the two stars, and clearly would have to be simulated for any particular case. To give some idea of the order of magnitude of these effects, I simulated a star (using an observed PSF) with a companion 3 times fainter, and with a PSF of 2 pixels FWHM. When the stars were separated by 2 pixels in both X and Y the flux of the faint star was recovered to better than 0.3 per cent. If the separation was decreased to 1 pixel, however, when the interpolation becomes dubious, the flux error was approximately 10 per cent for the companion.

Despite the complexity of calculating the stellar separations, the method is practicable for time-series work, where the stars are separated by 2 or more pixels. No assumptions have been made about the shape of the PSF, other than that it is constant over the field of view, which gives this technique an advantage over analytic profile fitting. The only re-sampling that has been done is on to a coarser grid, which gives it an advantage over observed profile fitting.

5 A COMPARISON WITH THE OTHER METHODS

A profile-fitting program will aim to get the minimum value for χ^2, i.e. to minimize

$$\chi^2 = \sum_{i,j} \frac{(TP^{\mathrm{E}}_{i,j} - (D_{i,j} - S_{i,j}))^2}{V_{i,j}} . \quad (24)$$

Differentiating with respect to T to find where the minimum χ^2 occurs yields

$$F = \frac{\sum_{i,j} P^{\mathrm{E}}_{i,j}(D_{i,j} - S_{i,j})/V_{i,j}}{\sum_{i,j} (P^{\mathrm{E}}_{i,j})^2/V_{i,j}} . \quad (25)$$

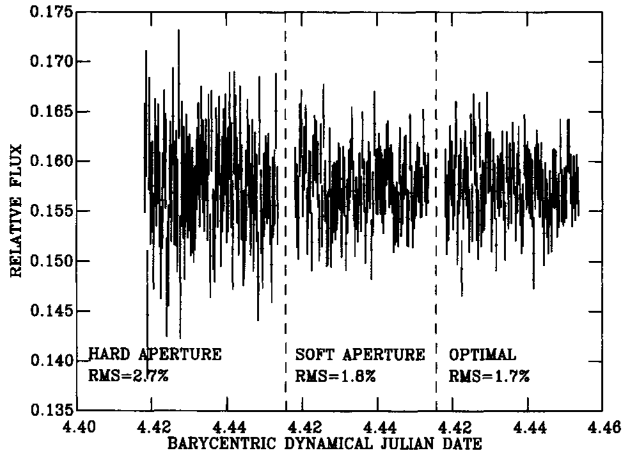

Figure 4. Three different extractions of the same data, using a simple aperture, a soft aperture and the optimal extraction. This illustrates how for poorly sampled data (the seeing is approximately 2 pixels), a soft aperture approximates to an optimal extraction. For both of the aperture reductions, the aperture radius which gives the best signal-to-noise ratio has been chosen (4.0 and 2.0 pixels for the hard and soft apertures respectively).

This is identical to equation (9). In Section 3.3 I showed that this form of the optimal extraction has a practical problem. It contains the variances of each pixel, and unless the estimated PSF is an exact match to the actual PSF, this will introduce a bias which changes systematically with magnitude. All profile fitting must contain this bias, although it is clearly a much more serious problem for fitting analytical functions than for observed PSFs constructed from other stars in the image.

The optimal extraction has a further, and perhaps more important advantage over profile fitting, in that equation (11) supplies a simple and robust error estimate. In contrast, any fitting technique will only provide good error estimates if the model is a good fit. It is unlikely that an analytic PSF will ever be a good fit to real data in the χ^2 sense, and so errors from such a technique are suspect. For observed profile fitting the problem is less severe, but even here it is normal to first adjust the pixel errors to aid convergence, and then to adjust them further to force a χ^2 of one (Stetson 1987).

The most basic advantage optimal extraction has over aperture photometry is that the mask is automatically optimized for each frame. This contrasts with trying many apertures to find the optimal one, in a procedure which can only be done globally for a group of frames. The optimal extraction also has an advantage of approximately 10 per cent in signal-to-noise ratio for well sampled data (Section 2.3). As the sampling becomes worse, the optimal extraction should give a larger improvement in signal-to-noise ratio, since aperture photometry suffers from re-sampling or aperture noise (Section 2.2). As Fig. 4 shows this is confirmed by experiments with a simple aperture. What is at first surprising, though, is that for poorly sampled data a soft aperture does almost as well as the optimal weighting. The reason for this is that a soft edged aperture with a scalelength similar to the width of the PSF (i.e. one or two pixels) is quite close to an optimal mask in effect.

A final set of advantages that the optimal extraction has over both profile fitting and aperture photometry relate to ease of use. Foremost amongst these is the simplicity of creating an adequate PSF. Automated fitting of a single bright star with an analytic function is all that is required, a process that can be undertaken by the program without human intervention. Since only one star is required, accurate photometry is possible when there is only a small number of stars in the field of view. Both the automation and the ability to work with a small number of stars are crucial in time-series work, where there are very many frames, all 'windowed' to the smallest possible area of sky to achieve the fastest CCD read times, and thus the highest time resolutions. Furthermore, the number of parameters one must adjust to perfect the extraction is small, specifically just the cut-off radius, giving one confidence that the best possible photometry has been performed.

The optimal extraction is a more complicated calculation than aperture photometry, and so should have a penalty in terms of CPU time. In fact this turns out to be rather small, presumably because of the I/O overheads associated with any image analysis. On a DEC AlphaStation 200 4/100 it takes 1.1 s per frame to fit four stars and determine the offsets relative to another image. It then takes a further 1.2 s to carry out both aperture and optimal photometry on the four stars. For aperture photometry alone the offsets would still have to be calculated, but the time to carry out the photometry is slightly faster at 0.7 s per frame using the code of Shahbaz et al. (1994).

6 CONCLUSIONS

I have presented a new method of extracting the flux from images of stars that offers a significant gain in signal-to-noise ratio over aperture photometry. It also has advantages over profile fitting in terms of ease of use, robust error estimation, and freedom from magnitude-dependent bias introduced by mis-estimating the stellar PSF. Code implementing the algorithms presented here is available from the author's website at http://www.astro.keele.ac.uk/~timn/.

ACKNOWLEDGMENTS

I would like to thank all those who have taken the time to discuss the photometry problem with me over the years, especially Tariq Shahbaz, Bill Welsh and Mike Irwin. The data for this paper were obtained at Keele Observatory, which is maintained with help of the amateur astronomers of Keele AstroSoc. The data reduction was carried out on the Keele Starlink Node, with the ARK software. The author is supported by a PPARC Advanced Fellowship.

REFERENCES

Bijaoui A., 1980, A&A, 84, 81

Da Costa G. S., 1992, in Howell S. B., ed., ASP Conf. Ser. 23, Astronomical CCD Observing and Reduction Techniques. Astron. Soc. Pac., San Francisco, p. 90

Gilliland R. L., Brown T. M., 1988, PASP, 100, 754

Hewett P. C., Irwin M. J., Bunclark P. B., Bridgeland M. T., Kibblewhite E. J., He X. T., Smith M. G., 1985, MNRAS, 213, 971

Honeycutt K. R. 1992, PASP, 104, 435

Horne K., 1986, PASP, 98, 609

Howell S. B., 1989, PASP, 101, 616

Irwin M. J., 1985, MNRAS, 214, 575

Miles R., 1998, J. Br. Astron. Assoc., in press

Penny A. J., Dickens R. J., 1986, MNRAS, 220, 845

Pratt W. K., 1978, Digital Image Processing. John Wiley and Sons, New York

Robertson J. G., 1986, PASP, 98, 1220

Shahbaz T., Naylor T., Charles P. A., 1994, MNRAS, 268, 756

Stetson P. B., 1987, PASP, 99, 191

Stetson P. B., 1990, PASP, 102, 932

This paper has been typeset from a TEX/LATEX file prepared by the author.