The Astrophysical Journal, 830:27 (23pp), 2016 October 10 doi:10.3847/0004-637X/830/1/27

PROPER IMAGE SUBTRACTION—OPTIMAL TRANSIENT DETECTION, PHOTOMETRY, AND HYPOTHESIS TESTING

Barak Zackay, Eran O. Ofek, and Avishay Gal-Yam
Benoziyo Center for Astrophysics, Weizmann Institute of Science, 76100 Rehovot, Israel; bzackay@gmail.com, eran.ofek@weizmann.ac.il
Received 2016 January 11; revised 2016 May 23; accepted 2016 June 15; published 2016 October 4

ABSTRACT

Transient detection and flux measurement via image subtraction stand at the base of time domain astronomy. Due to the varying seeing conditions, the image subtraction process is non-trivial, and existing solutions suffer from a variety of problems. Starting from basic statistical principles, we develop the optimal statistic for transient detection, flux measurement, and any image-difference hypothesis testing. We derive a closed-form statistic that: (1) is mathematically proven to be the optimal transient detection statistic in the limit of background-dominated noise, (2) is numerically stable, (3) for accurately registered, adequately sampled images, does not leave subtraction or deconvolution artifacts, (4) allows automatic transient detection to the theoretical sensitivity limit by providing credible detection significance, (5) has uncorrelated white noise, (6) is a sufficient statistic for any further statistical test on the difference image, and, in particular, allows us to distinguish particle hits and other image artifacts from real transients, (7) is symmetric to the exchange of the new and reference images, (8) is at least an order of magnitude faster to compute than some popular methods, and (9) is straightforward to implement. Furthermore, we present extensions of this method that make it resilient to registration errors, color-refraction errors, and any noise source that can be modeled. In addition, we show that the optimal way to prepare a reference image is the proper image coaddition presented in Zackay & Ofek. We demonstrate this method on simulated data and real observations from the PTF data release 2. We provide an implementation of this algorithm in MATLAB and Python.

Key words: gravitational lensing: micro – methods: data analysis – methods: statistical – surveys – techniques: image processing – techniques: photometric

1. INTRODUCTION

Detection of previously unknown transient sources is at the base of many fields of astronomy. Examples include the searches for supernovae, microlensing events and light echos. To remove a constant complex background, it is useful to perform digital image subtraction, a problem that has proven to be hard to tackle, with several suggested solutions (e.g., Phillips & Davis 1995; Alard & Lupton 1998; Bramich 2008; Gal-Yam et al. 2008; Yuan & Akerlof 2008). Probably the most popular algorithms are by Alard & Lupton (1998) and Bramich (2008).

Current methods have several problems and limitations. An important difficulty in image subtraction is that the point-spread function (PSF) of images taken from the ground is varying.[1] In some cases, the subtraction is based on a numerically unstable process (deconvolution) that may generate subtraction artifacts. Combined with ill-defined error propagation[2], it is difficult to decide if a transient candidate is real or rather due to a subtraction artifact. Finally, there is no proof that any of the methods we are currently using are optimal. As we will show in this paper, none of these algorithms are optimal. One hint for this is that some of these methods are not symmetric to exchange of the reference image and the new image, while the problem is symmetric. Another hint is that none of the methods define the matched filter that one should use in order to detect transients in the difference image.

In the Alard & Lupton (1998) and Bramich (2008) class of solutions, a complex inversion problem needs to be solved. This inversion problem can be regarded as a regularization effort (e.g., Becker et al. 2012) on the partial deconvolution done by Phillips & Davis (1995). Apart from being computationally slow, this inversion problem is in itself an effective deconvolution, and the numerical instability of the deconvolution process cannot be swept under the rug. These algorithms explore the trade-off between ringing artifacts in the subtraction image, that are due to the effective division in the Fourier plane, and residuals from the constant-in-time sky that are due to a failure of equalizing the PSFs of the reference and the new images. For example, if the PSF of the new image is sharper than the PSF of the reference in some axes, then these methods find no good solutions leading to multiple image artifacts.

These artifacts, along with residuals caused by registration errors, appear as false positive signals that hinder the automatic detection of transients. The current state of the art solution to this problem is to train a machine-learning algorithm (e.g., Bloom et al. 2012; Goldstein et al. 2015; Wright et al. 2015) to filter most of the artifacts and reduce the number of false positives to the minimum. However, this solution is partial and human scanners are required to sift through all remaining candidate detections and decide which is real and which is not (e.g., Gal-Yam et al. 2011; Smith et al. 2011).

This elaborate process can undermine the successful operation of transient searches in many ways. First, employing many human scanners can be cumbersome and expensive. Current surveys are spending considerable manpower on

[1] Sometimes this is relevant also for space-based observation.

[2] There are several reasons why the current methods propagate the errors incorrectly. One reason is that convolution generates correlated noise, which is typically ignored. Second is that usually the errors in the reference image are not projected correctly.

candidate sifting (e.g., Palomar Transient Factory; PTF). Without further dramatic improvement, this use of human scanners is unscalable, and is unfeasible for future surveys like ZTF (Bellm et al. 2015) and LSST (Ivezic et al. 2008). Second, having humans in the loop introduces a time delay in the transient detection. This can compromise science cases in which it is of utmost importance to make rapid follow-up observations of new transients (e.g., Cenko et al. 2013, 2015; Gal-Yam et al. 2014). Moreover, our experience is that at least some machine-learning algorithms throw away real obvious transients. Furthermore, the human scanning step makes it difficult to estimate the completeness of transient surveys as human scanners are difficult to properly simulate. Another problem is that even human scanners can be unsure if a transient is real or an artifact, and many surveys adopt the methodology of accepting only candidates that are persistent in two or more consecutive observations[3] (e.g., Gal-Yam et al. 2011; Baltay et al. 2013). This methodology trades the survey speed with the increased credibility of the candidates, and causes an additional time delay in transient detection. Last, human scanning makes it difficult to detect transients at the faintest limit because it is hard for humans to objectively quantify the false alarm probability.

In this paper, we present a closed-form solution for image subtraction in general, and transient detection in particular. Starting with the most basic statistical principles, we solve the problem of transient detection under the assumption that both the reference and the new images have white Gaussian noise (e.g., the background-noise- or read-noise-dominated limit). We then characterize the statistical behavior of our closed-form transient detection statistic under the influence of source noise and astrometric errors. Based on this analysis, we then construct a correction term to the transient detection statistic that prevents false positive detections in the vicinity of bright objects and due to registration errors. Our solution is always numerically stable, is trivial to implement and analyze, and is significantly faster computationally than the popular algorithms (e.g., Alard & Lupton 1998; Bramich 2008). We extend the transient detection statistic to the situation of multiple references, and show that the optimal reference image for image subtraction is the proper coaddition image given in Zackay & Ofek (2015b). Finally, we show that the transient detection statistic is the maximal signal-to-noise ratio (S/N) estimator for transient flux measurement in the background-dominated noise limit.

We further develop the optimal transient detection statistic into a difference image statistic that has white noise. Then, we show that any statistical measurement or decision on the data can be performed optimally and intuitively on this difference image, which we call the proper image subtraction statistic. This image has many good qualities, for instance, in the case of no difference between the reference and the new image, it has expectancy zero everywhere and uncorrelated additive Gaussian noise. It has an effective PSF that, by match filtering[4], reproduces the optimal transient detection statistic. Using this image, it is possible to detect and filter out particle hits in both the reference image and the new image, separating these artifacts from real transients. Another potential use of this image is the optimal detection of photometric variability and the astrometric motion of stars, that works in arbitrarily dense environments.

We demonstrate the efficacy of our algorithm on simulated and real images that are part of the PTF (Law et al. 2009), data release 2.

The outline of the paper is as follows. In Section 2, we review the state of the art image subtraction methods, while in Section 3 we derive our optimal transient detection and image subtraction algorithm. In Section 4, we discuss the properties of the derived image subtraction statistic. A step by step summary of the image subtraction process is presented in Section 5. In Section 6, we present tests on simulated and real data, while in Section 7 we describe our code, which is available online. In Section 8, we discuss the implementation details, and we conclude in Section 9.

2. BRIEF OVERVIEW AND ANALYSIS OF EXISTING METHODS FOR IMAGE SUBTRACTION

Previously suggested solutions for image subtraction can be divided into two variants. The first, and more popular variant, can be referred to as regularized partial deconvolution. Solutions we include in this family are Phillips & Davis (1995), Alard & Lupton (1998), and Bramich (2008). Gal-Yam et al. (2008) suggested a second variant, which we call cross filtering, while Yuan & Akerlof (2008) advocated for a mix of the two methods.

Denoting the new image by N and its PSF by P_N, the reference image by R and its PSF by P_R, the first approach attempts to find a convolution kernel k such that

$$N - k \otimes R \cong 0. \tag{1}$$

Here, $\otimes$ represents convolution.

The first solution for finding the kernel k was given by Phillips & Davis (1995). They suggested to perform a deconvolution solution in Fourier space:

$$\hat{k} = \frac{\widehat{P_n}}{\widehat{P_r}} \cong \frac{\widehat{N}}{\widehat{R}}, \tag{2}$$

were $\widehat{\ }$ represents Fourier transform. However, this solution is numerically unstable because the deconvolution operation can (and many times does) involve division by small numbers. This problem is apparent from Equation (2), where the denominator might approach zero as fast or faster than the numerator. Given that any measurement process contains noise, this division operation amplifies the noise in Fourier space, which in turn generates correlated noise in real space. The extreme cases of this correlated noise are the characteristic ringing and sinusoidal artifacts that deconvolved images suffer from.

Alard & Lupton (1998) suggested a practical way to mitigate the numerical instability problem. Restricting k to a small stamp and representing it as a set of basis functions, and noting that Equation (1) is linear, they suggested to solve for k using linear least squares. Alard & Lupton (1998) suggested the use of a set of basis functions, which are linear combinations of Gaussians multiplied by low degree polynomials. Later on, Bramich (2008) suggested to solve for the values of a pixelized kernel. All of the above methods can be viewed as regularizations of the deconvolution method of Phillips & Davis (1995)—i.e., restricting the solutions for the kernel k to finite size and to some set of logical solutions. Even though the numerical stability of these algorithms is much better than that

[3] This step is also required for unknown minor planet identification.

[4] Also called cross-correlation of the images with its PSF. See, e.g., Zackay & Ofek (2015a) for a derivation of the matched filter solution.

of Equation (2), they still have several problems. First, the division by zero problem is still there and it can become especially pronounced when the new image has a narrower PSF (including a PSF that is narrower in any single axis) compared to the reference image. It is interesting to note that these methods are not symmetric to the exchange of the new and reference image, while the problem is symmetric to this exchange. Second, although these methods are intuitive, they lack statistical justification, there is no rigorous proof they cause no information loss, and it is unclear what further image processing should be applied. For example, do we need to apply another matched filter to the subtracted image in order to detect transients? If so, which filter should we use? Third, using these methods the resulting pixel noise is correlated and there is no simple analytic prescription on how to set a detection threshold for transient search.[5] Therefore, it is hard to decide if a detected source is real or an artifact, or to quantify the probability of it being a false positive. In addition, in the effort of suppressing the deconvolution artifacts, these solutions sacrifice the cancellation of the constant-in-time image. This will cause large and pronounced subtraction artifacts, that will prevent identification of transients that are substantially fainter than their hosting environment. Finally, using inversion methods for image subtraction (i.e., linear least squares) makes the subtraction process slow, compared with, e.g., the Fourier space solution of Phillips & Davis (1995).

The cross-filtering solution suggested by Gal-Yam et al. (2008) is to convolve the new image with the PSF of the reference image and to convolve the reference image with the PSF of the new image:

$$S_{\mathrm{GY08}} = P_r \otimes N - P_n \otimes R. \tag{3}$$

This solution is always numerically stable and leaves no subtraction artifacts. The problem with this solution is, again, the lack of statistical justification, and that the matched filter for source detection is not specified.

Yuan & Akerlof (2008) suggested to apply kernels for both R and N, both chosen from a family of PSFs determined by few parameters, and to drive the solution toward spatially small kernels by adding the effective PSF area to the loss function.

It is worthwhile to note that the problem of subtracting two images, and minimizing the resulting difference image in the least square sense has an infinite number of solutions (see also Yuan & Akerlof 2008). For example, the linear equation

$$K_{\mathrm{r}} \otimes R - K_{\mathrm{n}} \otimes N \cong 0, \tag{4}$$

where K_{r} and K_{n} are arbitrary kernels, has an infinite number of solutions. This is because for any K_{r}, we can find K_{n} that satisfies Equation (4) in the least squares sense. It is clear from this simple analysis that all subtraction methods mentioned are focused on making the PSF of the two images identical, with very little attention to the maximization of the S/N of a transient source that appears in one of the images. In a sense, these methods do not solve the transient detection problem, but a different problem, which is how to make two images as similar as possible using convolution. In this paper, we rigorously derive a method that cancels the constant-in-time image *and* maximizes the S/N of a transient source at the same time. We note that there are several ways to derive this method. Here we will derive it from first principles via modeling the transient detection with simple hypothesis testing and using the lemma of Neyman & Pearson (1933). Additional derivations are given in the appendices.

[5] One method to estimate the noise level is using Bootstrap simulations (e.g., Ofek et al. 2014).

3. STATISTICAL DERIVATION

Given the numerous problems with existing image subtraction methods, we would like to place the transient detection problem on firm statistical grounds. In Section 3.1, we outline the derivation and formulae of our image subtraction statistics. Given that the full derivation is tedious, we defer it to Appendix A. In Section 3.2, we show that the best way to build a reference image, for the purpose of image subtraction, is to use the image coaddition algorithm of Zackay & Ofek (2015b). Our derivation in Section 3.1 assumes that the images are background-noise dominated (i.e., the objects we care about have source noise that is lower than the background noise). This causes an underestimation of the noise near bright sources. In Section 3.3, we present a simple correction to the image subtraction formulae that takes care of the source noise and other errors, like registration noise. In Section 3.4 we present an accurate treatment of astrometric shifts, noise, and color-refraction errors. In Section 3.5, we outline a possible method to equalize the flux zero points of the new and reference images. In Section 3.6, we provide an algorithm for optimal PSF photometry in the subtraction image, while in Section 3.7 we describe how this method can be used for cosmic-ray, bad pixel, and reflection-ghost identification.

3.1. Transient Source Detection Using Image Subtraction

Here we derive, from first principles, an optimal method for transient source detection, under the assumptions that the images are background-noise dominated, and the noise is Gaussian and independent.[6]

R and N are the background-subtracted reference image and the background-subtracted new image, respectively. Denoted by T is the background-subtracted true constant sky image. Denoted by P_r and P_n are the PSFs of the reference image and the new image, respectively. P_r and P_n are normalized to have unit sum. We assume that P_n, P_r, and the flux-based zero points[7] of the new image (F_n) and reference image (F_r) are known. We present a method for finding F_n and F_r in Section 3.5, and the PSF measurements are discussed in Section 8.2. The expression for the reference image is

$$R = F_r T \otimes P_r + \epsilon_r, \tag{5}$$

where ϵ_r is the additive noise component of the image R.

Given the null hypothesis, $\mathcal{H}_0$, that states there are no new sources in the new image we can write

$$N_{|\mathcal{H}_0} = F_n T \otimes P_n + \epsilon_n. \tag{6}$$

Given the alternative hypothesis, $\mathcal{H}_1(q, \alpha)$, that states there is a new point source at position q with flux α in the new image, we

[6] In practice, the pixels may be slightly correlated due to charge repulsion and charge diffusion in a CCD.

[7] Following Zackay & Ofek (2015a, 2015b), this factor represents the product of atmospheric transparency, telescope, and detector transmission and integration time.

can write

$$N_{|\mathcal{H}_1(q,\alpha)} = F_n T \otimes P_n + \alpha F_n \delta(q) \otimes P_n + \epsilon_n, \quad (7)$$

where $\delta(q)$ denotes a two-dimensional image with one at position q, and zero otherwise. We assume that the dominant source of noise is the background noise, ϵ_r and ϵ_n both satisfy that all pairs of pixels are uncorrelated—i.e., that for all pairs of pixels x_1, x_2 for which $x_1 \neq x_2$:

$$\mathrm{Cov}(\epsilon_r[x_1], \epsilon_r[x_2]) = 0, \mathrm{Cov}(\epsilon_n[x_1], \epsilon_n[x_2]) = 0, \quad (8)$$

and that all pixels have spatially uniform variance[8]:

$$V(\epsilon_r[x]) = \sigma_r^2, V(\epsilon_n[x]) = \sigma_n^2. \quad (9)$$

Because both hypotheses are simple[9], we can use the Neyman–Pearson lemma (Neyman & Pearson 1933), which states that the most powerful[10] statistic for deciding between two simple hypotheses is the likelihood ratio test:

$$\mathcal{L}(q, \alpha) = \frac{\mathcal{P}(N, R|\mathcal{H}_0)}{\mathcal{P}(N, R|\mathcal{H}_1(q, \alpha))}, \quad (10)$$

where $\mathcal{P}$ denotes probability. A critical point is that we do not have any prior information or assumptions on T. Therefore, we cannot calculate the probabilities $\mathcal{P}(N, R|\mathcal{H}_0)$ and $\mathcal{P}(N, R|\mathcal{H}_1(q, \alpha))$ directly. However, we can calculate their ratio by developing the expression using the law of conditional probabilities

$$\mathcal{L}(q, \alpha) = \frac{\mathcal{P}(N|R, \mathcal{H}_0)\mathcal{P}(R|\mathcal{H}_0)}{\mathcal{P}(N|R, \mathcal{H}_1(q, \alpha))\mathcal{P}(R|\mathcal{H}_1(q, \alpha))}. \quad (11)$$

Next, we can use the fact that $\mathcal{H}_0$ and $\mathcal{H}_1$ predict the same likelihood to the reference and cancel out the last multiplicative terms in the numerator and denominator.

After some algebra, which is detailed in Appendix A, we can find the optimal statistic for source detection

$$\widehat{S} \equiv \widehat{\frac{1}{\alpha}\log\mathcal{L}} = \frac{F_n F_r^2 \overline{\widehat{P_n}}|\widehat{P_r}|^2\widehat{N} - F_r F_n^2 \overline{\widehat{P_r}}|\widehat{P_n}|^2\widehat{R}}{\sigma_r^2 F_n^2|\widehat{P_n}|^2 + \sigma_n^2 F_r^2|\widehat{P_r}|^2}, \quad (12)$$

where the over-line symbol denotes the complex conjugate operation. We note that by putting the over-line sign above the hat sign, we mean that the complex conjugate operation follows the Fourier transform operation. This statistic (or score image) is simply the log-likelihood ratio test between the two hypotheses. This score is calculated simultanously for all values of α, while each pixel in the score image refers to a different q position. It is important to note that Equation (12) is a matched filter image and no further filtering is required. In order to find transients all we need to do is identify local maxima (or minima) in S. The significance of a local maximum, in units of sigmas, is given by its value divided by the standard deviation of the image S.

Since Equation (12) is a matched filter image, its pixels are correlated, and any hypothesis testing or measurement, other than transient detection and photometry (see Section 3.6), requires a knowledge of the covariance between the pixels. An example for such hypothesis testing is cosmic-ray identification via image subtraction, or searching for variable nebulosity (e.g., light echos). In order to have an image subtraction method that is optimal for all purposes and easy to use, we need to identify an image whose pixel noise is uncorrelated, and that cross-correlating this image with its own PSF returns Equation (12). In Appendix A, we identify such an image as

$$\widehat{D} = \frac{F_r\widehat{P_r}\widehat{N} - F_n\widehat{P_n}\widehat{R}}{\sqrt{\sigma_n^2 F_r^2|\widehat{P_r}|^2 + \sigma_r^2 F_n^2|\widehat{P_n}|^2}}. \quad (13)$$

The PSF of this image, normalized to have unit sum, is given by

$$\widehat{P_D} = \frac{F_r F_n \widehat{P_r}\widehat{P_n}}{F_D\sqrt{\sigma_n^2 F_r^2|\widehat{P_r}|^2 + \sigma_r^2 F_n^2|\widehat{P_n}|^2}}, \quad (14)$$

where F_D is the flux-based zero point of the subtraction image, which is given by

$$F_D = \frac{F_r F_n}{\sqrt{\sigma_n^2 F_r^2 + \sigma_r^2 F_n^2}}. \quad (15)$$

Indeed, using this difference image D and its PSF, we can verify that the cross-correlation of D with P_D returns

$$S = F_D D \otimes \overleftarrow{P_D}, \quad (16)$$

where the backward arrow sign denotes coordinate reversal (i.e., $\overleftarrow{P}(x, y) = P(-x, -y)$). Alternatively, in Fourier space

$$\widehat{S} = F_D\widehat{D}\overline{\widehat{P_D}}. \quad (17)$$

It is important to note that, in the background-dominated noise limit, D is a proper image, and hence we call it the proper subtraction image. As in Zackay & Ofek (2015b), we define a proper image to be an image whose noise is independent and identically[11] distributed (i.i.d). This means that D can be used for any hypothesis testing or measurement, without the need for the covariance between the pixels. Furthermore, in Appendix E, we present a proof that D and P_D are in fact sufficient statistics[12] for any hypothesis testing or measurement.

Equation (13) and its PSF (Equation (14)) are adequate for the detection of objects whose original shape was convolved with the telescope and atmosphere PSF. However, particle hit events do not share this PSF. In Appendix E, we derive the PSF in the difference image D, of a δ-function in N or R. The PSF in the difference image D of a δ-function in N is

$$\widehat{P_{D_N}} = \frac{F_r\widehat{P_r}}{F_{D_n}\sqrt{\sigma_n^2 F_r^2|\widehat{P_r}|^2 + \sigma_r^2 F_n^2|\widehat{P_n}|^2}}, \quad (18)$$

[8] As the convolution is a local operation, this assumption can be relaxed (see the discussion in Zackay & Ofek 2015a).

[9] A simple hypothesis has no unknown parameters. We are applying the hypothesis testing to each value of α and q separately. See additional discussion in Appendix A.

[10] The power of a binary hypothesis test is the probability that the test correctly rejects the null hypothesis when the alternative hypothesis is true.

[11] In practice, the noise levels need to be identical only locally (on scales that are twice the PSF size) because the convolution is a local operation. In the vicinity of bright stars, D is not proper.

[12] In statistics, a statistic is sufficient with respect to a statistical model and its associated unknown parameter if no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter.

while the PSF in the difference image D of a δ-function in R is

$$\widehat{P_{D_R}} = \frac{F_n \widehat{P_n}}{F_{D_R}\sqrt{\sigma_n^2 F_r^2 |\widehat{P_r}|^2 + \sigma_r^2 F_n^2 |\widehat{P_n}|^2}}. \quad (19)$$

These PSFs are also accompanied by the corresponding zero points, F_{D_N}, F_{D_R} that can be found in Appendix E.

These equations are useful if one would like to search for events that are similar to a delta function (e.g., bad pixels). We note that $\widehat{P_{D_N}}$ and $\widehat{P_{D_R}}$ in many cases can be approximated by a delta function.

To summarize, in order to find a transient source in either the reference or the new image, we can calculate D (Equation (13)) and cross-correlate it with its PSF (Equation (14)). Alternatively, we can calculate directly the statistic S (Equation (12)).

3.2. Construction of the Reference Image

Typically, the reference image is built by coadding multiple images. Here we will show that the best way to produce a reference image for subtraction is using the method described in Zackay & Ofek (2015b).

In the case of multiple reference images, we need to replace Equation (5) with the model for the jth reference image:

$$R_j = F_j P_j \otimes T + \epsilon_j. \quad (20)$$

Here F_j is the flux-based zero point of the jth reference image, P_j is the PSF of the jth reference image, and ϵ_j is the noise of the jth reference image.

As before, the model for N, assuming the null hypothesis, $\mathcal{H}_0$, is given by Equation (6), while if the first hypothesis, $\mathcal{H}_1$, is true then N is given by Equation (7).

As in the previous section, we would like to decide between two simple hypotheses. Therefore, the optimal test statistic is the likelihood ratio test (Neyman & Pearson 1933)

$$\mathcal{L}(q, \alpha) = \frac{\mathcal{P}(N, R_1, \ldots, R_J | \mathcal{H}_0)}{\mathcal{P}(N, R_1, \ldots, R_J | \mathcal{H}_1(q, \alpha))}. \quad (21)$$

As before, we can use the law of conditional probabilities, and the fact that $\mathcal{H}_0$ and $\mathcal{H}_1$ predict the same likelihood for all references. The full derivation is presented in Appendix B, and after some algebra we find that the optimal reference image is given by

$$\widehat{R} = \frac{\sum_j \frac{F_j}{\sigma_j^2} \overline{\widehat{P_j}} \widehat{R_j}}{\sqrt{\sum_j \frac{F_j^2}{\sigma_j^2} |\widehat{P_j}|^2}}. \quad (22)$$

The PSF (normalized to have unit sum) of the reference image is given by

$$\widehat{P_R} = \frac{\sqrt{\sum_j \frac{F_j^2}{\sigma_j^2} |\widehat{P_j}|^2}}{F_r}, \quad (23)$$

where F_r is the flux-based zero point of the reference

$$F_r = \sqrt{\sum_j \frac{F_j^2}{\sigma_j^2}}. \quad (24)$$

Not surprising, this is identical to the optimal coaddition method derived in Zackay & Ofek (2015b). We note that the reason R preserves all the information from the individual references is because, in the computation of each frequency in R, we add random variables scaled by their (conjugate) expectation, divided by the variance. We can identify this operation as the maximal S/N addition of random variables (see Appendix A of Zackay & Ofek 2015a). The reader should refer to Zackay & Ofek (2015b) for analysis and proof of sufficiency of this so-called proper coaddition method.

3.3. Simple, Suboptimal Correction for Source Noise, Astrometric Noise, and Color-refraction Noise

Equation (12) ignores the source noise, and hence the noise level is underestimated in the vicinity of bright stars. The outcome of this will be that bright sources may be flagged as possible transients or variables. Furthermore, this equation ignores any additional important sources of noise like astrometric noise, astrometric scintillation noise, color-refraction noise, flux scintillation noise, and position-dependent flat-fielding errors.

A simple correction to this problem, albeit suboptimal, is to divide S by a correction factor that takes into account the local estimated variance of the extra noise. Derivation of this correction factor is presented in Appendix C. In the image space, the expression for the corrected S is

$$S_{\rm corr} = \frac{S}{\sqrt{V(S_N) + V(S_R) + V_{\rm ast}(S_N) + V_{\rm ast}(S_R) + \ldots}}. \quad (25)$$

Here the terms in the denominator may include any position-dependent contribution to the variance, that is not included in the σ_n^2 and σ_r^2 factors.

In this example, we list two specific contributions from the source noise and from astrometric noise. The first two terms in the denominator are the variance from the source noise in the new and reference images, respectively, while the next two terms are the variance due to astrometric noise. Other sources of noise like color refraction can be added in a similar manner.

Here $V(S_N)$ is the variance of the part of S containing N given by

$$V(S_N) = V(\epsilon_n) \otimes (k_n^2), \quad (26)$$

and $V(S_R)$ is the variance of the part of S containing R given by

$$V(S_R) = V(\epsilon_r) \otimes (k_r^2), \quad (27)$$

and the Fourier transform of k_r is given by

$$\widehat{k_r} = \frac{F_r F_n^2 \overline{\widehat{P_r}} |\widehat{P_n}|^2}{\sigma_r^2 F_n^2 |\widehat{P_n}|^2 + \sigma_n^2 F_r^2 |\widehat{P_r}|^2}, \quad (28)$$

while the Fourier transform of k_n is

$$\widehat{k_n} = \frac{F_n F_r^2 \overline{\widehat{P_n}} |\widehat{P_r}|^2}{\sigma_r^2 F_n^2 |\widehat{P_n}|^2 + \sigma_n^2 F_r^2 |\widehat{P_r}|^2}. \quad (29)$$

The variance of ϵ_n and ϵ_r are simply the variance images. For a single image, the variance map, $V(\epsilon_n)$, is simply the number of electrons in each pixel (including the background), added with the readout noise squared. However, in the case of multiple images, the correct way to construct $V(S_R)$ is to calculate k_r, $V(\epsilon_r)$, and $V(S_R)$ for each reference image and to sum all the individual $V(S_R)$ values up (see Appendix B). However, in many cases, a reasonable approximation is to calculate k_r from

the properly coadded image, and calculate $V(\epsilon_r)$ using a simple addition of all the images (in units of electrons) from which the reference was constructed (i.e., the number of electrons in each pixel including the background) added with the total read noise squared.

Next, the astrometric variance terms are given by

$$V_{\rm ast}(S_N) = \sigma_x^2\left(\frac{dS_N}{dx}\right)^2 + \sigma_y^2\left(\frac{dS_N}{dy}\right)^2, \quad (30)$$

where σ_x and σ_y are the astrometric registration noise in the x and y axes, respectively, while $\frac{dS_N}{dx}$ and $\frac{dS_N}{dy}$ are the gradients of S_N in the x and y directions, respectively. Here, the Fourier transform of S_N is given by

$$\widehat{S_N} = \widehat{k_n}\widehat{N}. \quad (31)$$

In a similar manner

$$V_{\rm ast}(S_R) = \sigma_x^2\left(\frac{dS_R}{dx}\right)^2 + \sigma_y^2\left(\frac{dS_R}{dy}\right)^2. \quad (32)$$

Here the Fourier transform of S_R is given by

$$\widehat{S_R} = \widehat{k_r}\widehat{R}. \quad (33)$$

The origin of these terms is that astrometric noise causes shifts in individual PSFs. The noise induced by these shifts is proportional to the difference between neighboring pixels (i.e., the gradient).

We note that in practice the astrometric registration noise is the rms of the registration fitting process. This term include both registration errors and the astrometric scintillation noise. In some cases, the quality of the registration is position dependent. In this case, it is possible to replace the scalars σ_X and σ_Y by matrices of the position-dependent noise. In Section 3.4, we suggest a more accurate treatment of the astrometric noise component.

3.4. Accurate Treatment of Astrometric Noise and Flux Variability

Astrometric errors and shifts are a major problem for image subtraction. For example, for a bright source with 10^4 electrons and full-width at half maximum (FWHM) of 2 pixels, the astrometric error induced by the Poisson noise will be about a few tens of milli-pixels. This is equivalent to the typical astrometric scintillation noise induced by the Earth turbulent atmosphere (see Section 8.5). Therefore, even in the case of high quality registration, we expect that all bright stars will have subtraction residuals due to astrometric scintillation noise.

Fortunately, due to the closed form and numerical stability of our method, the shape of the subtraction residuals is fully predictable, given the astrometric shift and the flux difference between the star as it appears in the reference and as it appears in the new image. Therefore, we can use this to measure the astrometric shift and flux variability for each star.

For adequately[13] sampled images, this proposed mechanism is accurate, and it allows us to measure astrometric shifts and variability in very crowded fields. The details of this method will be presented in a future publication, but here we provide a brief outline: the astrometric shift and photometric variability kernel is

$$\widehat{P_S}(\alpha_n, \alpha_r, \Delta x, \Delta y) = (\alpha_r - \alpha_n\hat{s})\widehat{P_D}. \quad (34)$$

Here, α_n is the flux of the source in N, α_r is its flux in R, and $\hat{s}$ is the shift operator (including sub-pixel shifts) in Fourier space. This operator is a function of the shifts Δx and Δy. Using Equation (34), we can treat residuals detected in S more carefully than we did in Section 3.3. Specifically, we can now perform hypothesis testing to decide between, e.g., $\mathcal{H}_0$ (changes are consistent with stationary and non-variable source), or $\mathcal{H}_1$ (the star moved or its flux changed). This scheme can be applied to any part of D, for which we identify a significant peak in S (e.g., above 3σ). Apart from using this to eliminate false positives, we can now use this to detect and measure new kinds of signals. For example, we can use it to search for moving objects blindly, even in the presence of complex, constant in time, structure in the background.

3.5. Matching the Local Zero Points, Background Flux, and Astrometric Shift

Our solution so far assumed that the values of the flux-based zero points (F_r and F_n), the background levels, (B_n and B_r), and the relative astrometric shift (Δx and Δy) are known. Careful analysis of Equation (13) shows that, in practice, we only care about the flux zero points ratio

$$\beta \equiv F_n/F_r, \quad (35)$$

the background difference,

$$\gamma \equiv B_n - B_r, \quad (36)$$

and the translation (Δx, Δy).

There are several ways to estimate β and γ (e.g., via relative photometry, Ofek et al. 2011), and here we provide yet another method. By substituting Equations (35) and (36) into D (Equation (13)), and introducing the shift operator, we can get the desired expression we need to minimize in order to find β, γ, Δx, and Δy. This can be done either locally (in small sections of the image), or globally. For simplicity, and since we already discussed astrometric shifts in Section 3.4, here we neglect translations, but toward the end we will mention how this can be incorporated.

In order to find β and γ, we need to compare the two parts of $\widehat{D}$:

$$\widehat{D_n}(\beta) = \frac{\widehat{P_r}\widehat{N}}{\sqrt{\sigma_n^2|\widehat{P_r}|^2 + \beta^2\sigma_r^2|\widehat{P_n}|^2}}, \quad (37)$$

and

$$\widehat{D_r}(\beta) = \frac{\widehat{P_n}\widehat{R}}{\sqrt{\sigma_n^2|\widehat{P_r}|^2 + \beta^2\sigma_r^2|\widehat{P_n}|^2}}. \quad (38)$$

Note that we replaced F_n and F_r by β. All we need to do is to inverse Fourier transform $\widehat{D_n}$ and $\widehat{D_r}$ and to solve the following nonlinear equation for β and γ' (and optionally Δx and Δy):

$$D_n(\beta) = \beta D_r(\beta) + \gamma' \quad (39)$$

[13] By adequately sampled images, we mean that the PSF width is sampled by at least two pixels. This can be referred to as the Nyquist sampling of the PSF by the camera.

where

$$\gamma' = \frac{\gamma}{\sqrt{\sigma_n^2 + \beta^2\sigma_r^2}}. \quad (40)$$

Note that the solution should be performed in the image domain. If we are interested in solving also for small translations, we need to multiply $\widehat{D_r}$ and γ' with the shift operator. If we trust that the images were background subtracted and aligned correctly, then we can set $\gamma = 0$, $\Delta x = 0, \Delta y = 0$ and use the same expression to solve only for the value of β.

Equation (39) is nonlinear in β. Therefore, iterative solutions are required. For example, in the first iteration we set $\beta = 1$ and solve for the new value of β, and use it in the next iteration to find a new value of β, until convergence.[14] Furthermore, it is important to note that one must use robust fitting methods in order to solve Equation (39). The reason is that there may be bad pixels, particle hits, astrometric noise, and saturated pixels in the images. It is also recomended to remove the images-edge pixels prior to fitting β.

3.6. PSF Photometry in the Difference Image

In this section, we present a statistic for measuring the PSF photometry[15] of a source in the difference image. This measurement statistic is unbiased and has maximal S/N among all estimators, which are linear combinations of the input images. In addition, this measurement statistic is also the maximum likelihood estimate for the flux of point-source transients. However, this statistic is optimal only for the background-dominated-noise limit. A full derivation of this statistic is presented in Appendix D.

The best linear estimator for the PSF photometry of a source at position q is

$$\widetilde{\alpha(q)} = \frac{S(q)}{F_S}. \quad (41)$$

Here F_S is the flux normalization of S:

$$F_S = \sum_f \frac{F_n^2|\widehat{P_n}|^2 F_r^2|\widehat{P_r}|^2}{\sigma_r^2 F_n^2|\widehat{P_n}|^2 + \sigma_n^2 F_r^2|\widehat{P_r}|^2}, \quad (42)$$

where f indicates spatial frequencies. The standard deviation of this estimator is

$$\sigma_{\widetilde{\alpha(q)}} = \frac{\sqrt{V(S_N) + V(S_R)}}{F_S}, \quad (43)$$

where $V(S_N)$ and $V(S_R)$ are defined in Equations (26)–(27). Note that Equation (41) can be used to measure the PSF flux of all the transients in the image simultanously.

3.7. Cosmic-ray, Bad Pixel, and Ghost Identification

The image subtraction statistic D can be used to identify cosmic rays and bad pixels. A major advantage of using the proper image subtraction over otherimage differencing techniques is that its pixel noise is uncorrelated and usually it roughly preserves the shape of sources that are similar to δ-functions. This means that in most cases one can identify particle hits by applying edge-detection algorithms (e.g., van Dokkum 2001), without any modifications, directly on D.

An alternative approach is to use a rough model for the shapes of particle hits and bad pixels, and to perform a composite hypothesis testing. The log-likelihood of observing D, if an object at position q is a point-source transient ($\mathcal{H}_{\rm ps}$ hypothesis) with flux α, is given by

$$-\log(\mathcal{P}(D|\mathcal{H}_{\rm ps}(q))) = \sum_x ||D - \alpha F_D \overleftarrow{P_D} \otimes \delta(q)||^2, \quad (44)$$

while the log-likelihood of D if the object at position q is a cosmic ray with flux α and with shape $P_{\rm cr}$ in N, ($\mathcal{H}_{\rm cr}$ hypothesis) is

$$-\log(\mathcal{P}(D|\mathcal{H}_{\rm cr}(q))) = \sum_x ||D - \alpha P_{\rm cr} \otimes \overleftarrow{P_{D_N}} \otimes \delta(q)||^2. \quad (45)$$

Here, x is the subset of pixels that contains the source of interest (e.g., an area with a width twice that of the PSF around the source). The difference between Equations (44) and (45) (using appropriate priors, such as the probability of seeing a transient at a certain magnitude and the probability of seeing a cosmic ray with this flux) is a statistic that can be indicative (after setting the appropriate threshold) for deciding whether the detected transient is a cosmic ray or an astronomical transient. We note that in this case the flux of the source and the intensity and shape of the cosmic ray are free parameters of the model. Therefore, this is a classic case of composite hypothesis testing.

The same approach can be used to identify internal-reflection ghosts. In this case, we need to replace the shape $P_{\rm cr}$ with the shape of a reflection ghost, for example, an extended kernel (e.g., top hat filter) that is wider than the stellar PSF.

4. PROPERTIES OF THE NEW IMAGE SUBTRACTION METHOD

Now that we have an optimal solution for the subtraction problem, we can analyze its properties and compare it to other methods, seeking an intuitive understanding.

4.1. Optimality

Our image subtraction and transient detection formulae were derived using the lemma of Neyman & Pearson (1933). This ensures that whenever our assumptions are correct our method is optimal. Our assumptions are that the images are registered, dominated by uncorrelated Gaussian background noise, and that the PSFs, background, variance, and flux-based zero points are known.

4.2. The Constant-in-time Image T Cancels

For perfectly registered images, both the optimal proper difference image (D) and transient detection image (S) are free of subtraction residuals from the constant-in-time image. This is because the constant-in-time image T algebraically vanishes.

This is not the case in the subtraction methods suggested by Alard & Lupton (1998) and Bramich (2008). In these methods, an optimum for the trade-off between magnifying the image noise and minimizing the constant-in-time residuals of T was explored.

[14] We found that usually β converges in two to three iterations.

[15] PSF photometry refers to (effectively) fitting the source with a PSF.

4.3. Numerical Stability

Inspecting Equations (12)–(14), it is apparent that if the denominator is approaching zero, then the numerator is approaching zero even faster. Therefore, our image subtraction method is numerically stable for all combinations of PSFs for the new and reference images.

We note that the Alard & Lupton (1998) and Bramich (2008) methods are numerically unstable in the general case because these methods effectively perform deconvolution. It is true that if the PSF of the reference image is narrower, in all axes, than the PSF of the new image, then the Alard & Lupton (1998) family of methods are stable. However, even in this case the solution found by these methods is suboptimal (i.e., it does not maximize the S/N of the transients). Furthermore, Equations (12)–(14) are numerically stable even if the PSFs used have large (even order of unity) errors compared to their true values. Therefore, no special accuracy is required in measuring the PSFs because of the division in the Fourier plane. We further demonstrate these points in Section 6.

4.4. Locality

When computing S and D, the only operations performed are convolutions of R and N with small kernels. Therefore, S and D could be calculated independently for every small image patch, (up to a few times the PSF size). This allows the used PSFs for subtraction to vary smoothly across the image. In addition, local artifacts such as bad pixels, particle hits, or saturated stars will affect only their close vicinity.

4.5. The Proper Image Subtraction D Has White Noise

In the expression for $\widehat{D}$ (Equation (13)), in the background-noise dominated limit, the variance of the numerator is equal to the square of the denominator, i.e.,

$$V\left[F_r\widehat{P_r}\widehat{N} - F_n\widehat{P_n}\widehat{R}\right] = \sigma_r^2 F_n^2|\widehat{P_n}|^2 + \sigma_n^2 F_r^2|\widehat{P_r}|^2, \quad (46)$$

which means that all the spatial frequencies of $\widehat{D}$ have equal variance. Furthermore, since we assume that the images have white noise, their Fourier transform has white noise. This means that the spatial frequencies of $\widehat{D}$, as a linear combination of $\widehat{R}$, $\widehat{N}$, has uncorrelated noise. Together, both properties mean that $\widehat{D}$ has white noise, which means that D also has white noise. In other words, the difference image is a proper image (as defined in Zackay & Ofek 2015b). This property is violated by all the other methods for image subtraction.

We note that in the vicinity of bright stars, where the source noise variance is dominant, the proper subtraction image D exhibits correlated noise. Our simulations suggest that if the source variance is at least an order of magnitude higher than the background variance, then correlated noise is detectable by eye in the vicinity of such sources. However, as we stated before, using our method, the source noise is controllable via variance corrections.

4.6. D and P_D are Sufficient for Any Measurement or Decision on the Difference Between the Images

In Appendix E, we provide proof that D and P_D constitute a sufficient statistic for any measurement or hypothesis testing on the difference between the images. The key ingredients for this proof are that any likelihood calculation for any generative model for the difference between the images can be computed by using only these quantities, and the use of the Fisher–Neyman factorization theorem (Fisher 1922; Neyman 1935). We note that there are an infinite number of sufficient statistics with respect to the image subtraction problem (see some examples in Zackay & Ofek 2015b in the context of coaddition). Here we prefer the proper subtraction image D (rather than e.g., S) due to its useful properties.

The sufficiency property has important practical consequences. It means that, in the background-noise-dominated limit, D and P_D contain all the information one needs for any further measurement or hypothesis testing related to the difference between the images. There is no need for other types of difference images for other applications. Examples of practical applications include the identification and removal of particle hits on the detector (see Section 3.7); optimal search for proper motion, astrometric shifts (Section 3.4), and asteroid streaks.

4.7. Symmetry Between the New Image and the Reference Image

The problem of image subtraction is symmetric to the exchange of the reference and the new image (up to the negation of the flux of the transient). Therefore, it is not surprising that the optimal image subtraction statistics (D or S) are symmetric to the exchange of R and N (up to a minus sign). This property is violated by the solutions proposed by Phillips & Davis (1995), Alard & Lupton (1998), and Bramich (2008). We note that the Gal-Yam et al. (2008) method preserves this symmetry. Interestingly, the Gal-Yam et al. (2008) method is identical to the numerator of the proper image subtraction statistic (D).

4.8. The Limit of the Noiseless Reference Image

In the limit of $\sigma_r \to 0$, Equation (12) becomes

$$\lim_{\sigma_r\to 0}\widehat{S} = \frac{F_n F_r^2\overline{\widehat{P_n}}|\widehat{P_r}|^2\widehat{N} - F_r F_n^2\overline{\widehat{P_r}}|\widehat{P_n}|^2\widehat{R}}{\sigma_n^2 F_r^2|\widehat{P_r}|^2} \quad (47)$$

$$= \frac{F_n\overline{\widehat{P_n}}}{\sigma_n^2}\left(\widehat{N} - \frac{F_n\widehat{P_n}}{F_r\widehat{P_r}}\widehat{R}\right). \quad (48)$$

The term $\widehat{P_n}/\widehat{P_r}$ can be identified as the convolution kernel solved for by the methods of Phillips & Davis (1995), Alard & Lupton (1998), and Bramich (2008). Therefore, in this limit, S converges to the Alard & Lupton (1998) family of methods followed by filtering each of the images with the PSF of the new image.

This simple analysis demonstrates that the Alard & Lupton (1998) family of methods, if followed by the correct matched filtering, is a special case of our solution S. Furthermore, Equation (48) provides the prescription for the correct matched filter (only) in the limit of $\sigma_r \to 0$.

To emphasize the importance of correctly accounting for the noise in the reference image R, we want to substitute some numbers into the formulas. For a very good reference images, $\sigma_r \sim 0.1\sigma_n$ (which represents a reference image composed of the coaddition of ∼100 images). Because the ratio $\frac{\widehat{P_n}}{\widehat{P_r}}$ can be much larger than $\frac{\sigma_n}{\sigma_r}$ (for example, if P_n is narrower than P_r, in the high spatial frequencies this ratio can easily get to 10^6), σ_r is

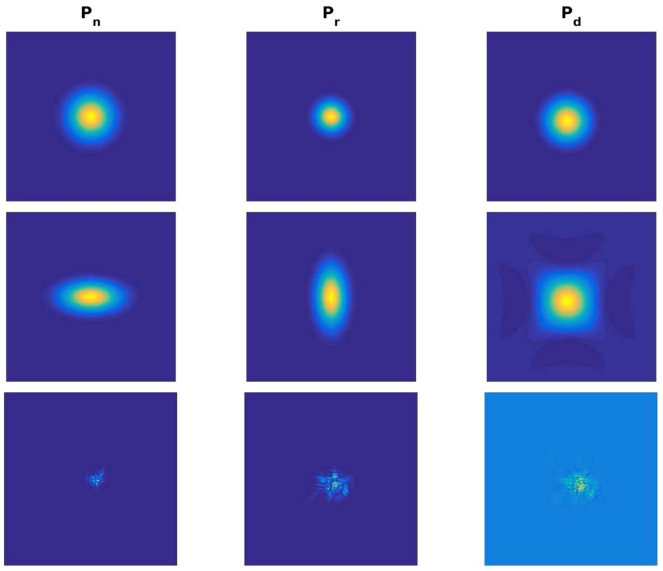

Figure 1. P_n (left column), P_r (middle column), and the corresponding P_D (right column) for three cases. The first row is for the case of symmetric Gaussian PSFs with sigma-widths of 2 and 3 pix for the new and reference, respectively. The second row is for the case of asymmetric Gaussian PSFs with sigma-widths of 2 by 4 pix and 4 by 2 pix for the new and reference, respectively. In the third row, P_n and P_r are simulated speckle images (using the tools in Ofek 2014). In the speckle simulations, we set $D_{\rm tel}/r_0 = 20$, where $D_{\rm tel}$ is the telescope diameter and r_0 is the Fried length.

never small enough to be negligible. This means that the Alard & Lupton (1998) family of solutions is qualitatively close to the correct statistic only if $\frac{|\widehat{P_n}|}{|\widehat{P_r}|} \ll \frac{\sigma_n}{\sigma_r}$. In any other case, these solutions will produce either deconvolution artifacts, or fail to cancel the constant in time image T.

4.9. The PSF of the Difference Image

The PSF, P_D, of the proper subtraction image is a combination of P_n and P_r. In Figure 1, we present P_n, P_r, and the corresponding P_D for three cases, of symmetric Gaussians, asymmetric Gaussians, and speckle images.

4.10. Knowledge of the PSFs

An apparent drawback of our method is that one needs to know the PSFs of the images, while in the Alard & Lupton (1998) family of methods one simply solves for the convolution kernel $\widehat{P_n}/\widehat{P_r}$ without measuring the PSFs.

However, one can write the expression for (a slightly changed) D with $\widehat{P_n}/\widehat{P_r}$, allowing one to incorporate relative knowledge of the PSFs.

$$\widehat{D_{\rm ratio}} = \frac{\widehat{N} - \frac{F_n\widehat{P_n}}{F_r\widehat{P_r}}\widehat{R}}{\sqrt{\sigma_n^2 + \sigma_r^2\frac{F_n^2|\widehat{P_n}|^2}{F_r^2|\widehat{P_r}|^2}}}. \quad (49)$$

In this case, $D_{\rm ratio}$ has the following PSF:

$$\widehat{P_{D_{\rm ratio}}} = \frac{F_rF_n\widehat{P_n}}{F_D\sqrt{\sigma_n^2 + \sigma_r^2\frac{F_n^2|\widehat{P_n}|^2}{F_r^2|\widehat{P_r}|^2}}}. \quad (50)$$

Inspecting these slightly modified definitions of D and P_D still yields $S = D_{\rm ratio} \otimes \overleftarrow{P_{D_{\rm ratio}}}$, and is thus equivalent to the original definition (Equation (13)). It is important to note, that while we are still requiring the estimation of $P_{D_{\rm ratio}}$ (and thus, effectively that of P_n and P_r) for performing transient detection, the required precision in estimating $P_{D_{\rm ratio}}$ is much less demanding than the required precision in estimation of the PSF ratio. In fact, even crude estimation of the general form of $P_{D_{\rm ratio}}$ as a two-dimensional Gaussian may result only in a negligible sensitivity loss. Therefore, if the individual PSFs cannot be estimated from the images, but their ratio can be (for example, if only extended, or complex sources are present in the field of view), it is still possible to optimally subtract the images and detect transients with maximal sensitivity. Therefore, our requirement for the knowledge of the PSFs should not be considered as a drawback, as it could be bypassed. We note that another drawback of solving for $\frac{\widehat{P_n}}{\widehat{P_r}}$ is that astrometric scintillation noise shuffles the relative position of sources in the new and reference images, biasing the solution for $\frac{\widehat{P_n}}{\widehat{P_r}}$ to be wider, and thus less accurate. This effect may degrade the cancellation quality of T, and thus may reduce the achievable contrast between T and the detectable transients. This effect is avoided when estimating P_r and P_n separately because astrometric noise is locally coherent, allowing the approach presented in Section 3.4.

4.11. Registration and Color-refraction Errors

Image subtraction relies on many steps taken prior to the differencing process. Any noise introduced by the pre-processing steps will be propagated into the final subtraction image. Examples for such problems include registration errors, color-refraction systematic errors, and small-scale flat-fielding errors.

Here we suggest two types of treatments for such noise. (1) It is straightforward to introduce these extra sources of noise into the variance image of S and use it to calculate $S_{\rm corr}$ (see Section 6.1 for examples). This correction is suboptimal, but it is resilient to pre-processing errors. (2) An accurate treatment of the problem is to fit any astrometric shift and flux variation for each detected artifact in the difference image D (see Section 3.4). Albeit this is computationally expensive, this kind of solution is very common in astronomy (e.g., `DAOPHOT`, Stetson 1987; `DOPHOT`, Schechter et al. 1993).

4.12. Free Parameters

In principle, our method does not have any free parameters that the user needs to set. We note that the Alard & Lupton (1998), Bramich (2008), and Yuan & Akerlof (2008) methods do have internal degrees of freedom that the user needs to define and that may influence the final outcome. For example, the Bramich (2008) method may be sensitive to the kernel size, while the Alard & Lupton (1998) method depends on the basis functions one chooses to represent the convolution kernel (see, e.g., Becker et al. 2012 and Bramich et al. 2015).

4.13. Computational Complexity

In terms of computational complexity, our subtraction method is fast, as the most demanding operation in our image subtraction method is the FFT operation (or alternatively convolution with a small kernel). Tests indicate that our

algorithm is at least an order of magnitude faster than the inversion algorithms by Alard & Lupton (1998) and Bramich (2008) because they are essentially solving a linear least square problem with a large number of equations and tens to hundreds of unknowns.

5. SUMMARY OF THE ALGORITHM

We recommend to perform the subtraction on small image patches in order to minimize residual astrometric shifts, inhomogeneous transparency, and background. In addition, it allows to use position-dependent PSFs. The image patches should be overlapping by at least two PSF lengths, in each dimension, in order to avoid edge effects of the convolution process. A step-by-step outline of our algorithm is as follows.

INPUT ARGUMENTS

$\boldsymbol{N}$	Background-subtracted new image (registered to R).
$\boldsymbol{R}$	Background-subtracted reference image.
$\boldsymbol{N_b}$	New image including background in electron units.
$\boldsymbol{R_b}$	Reference image including background in electron units.
P_n	PSF of new image normalized to have unit sum.
P_r	PSF of reference image normalized to have unit sum.
$\sigma_{\rm n}$	std of the background of the new image.
$\sigma_{\rm r}$	std of the background of the reference image.
r_n	Read noise of new image in electrons.
r_r	Read noise of reference image in electrons.
σ_x	rms (in pixels) of the astrometric registration solution in the X-axis. This is either a scalar or a matrix.
σ_y	rms (in pixels) of the astrometric registration solution in the Y-axis. This is either a scalar or a matrix.

OUTPUT

D	The proper difference image.
P_D	The PSF of the proper difference image.
$S_{\rm corr}$	The matched filter difference image corrected for source noise and astrometric noise.
P_{D_n}	The PSF of a delta function in N as it appears in D.
P_{D_r}	The PSF of a delta function in R as it appears in D.

ALGORITHM.

1. Optionally construct a reference image (R; Equation (22)), its PSF (P_r; Equation (23)), and flux (F_r; Equation (24)) using the Zackay & Ofek (2015b) proper coaddition method.
2. Solve Equation (39) for the best-fit value of β and optionally γ, Δx, and Δy (need to use Equations (37) and (38)). Since this Equation is nonlinear in β use iterations. Set $\beta = 1$ in the first iteration, update the value of β and continue until convergence. Use robust fitting.[16] Alternatively, find β using other methods (e.g., relative photometry).
3. If applicable, calculate γ (Equation (40)) and subtract γ from N.
4. If applicable, shift P_n by Δx and Δy.
5. Set $F_r = 1$ and $F_n = \beta$.
6. Calculate $\widehat{D}$ (Equation (13)).
7. Calculate $\widehat{P_D}$ (Equation(14)).
8. Calculate $\widehat{S} = \widehat{P_D}\widehat{D}$.
9. Calculate $\widehat{P_{D_n}}$ (Equations (18) and (130)).
10. Calculate $\widehat{P_{D_r}}$ (Equations (19) and (134)).
11. Calculate k_r (Equation (28)).
12. Calculate k_n (Equation (29)).
13. Set $V(\epsilon_n) = N_b + r_n^2$ and calculate $V(S_N)$ (Equation (26)).
14. Set $V(\epsilon_r) = R_b + r_r^2$ and calculate $V(S_R)$ (Equation (27)). If R is composed of multiple images, it is better to sum up the $V(S_{R_j})$ of the individual reference images (see Appendix C and Equation (101)).
15. Calculate $V_{\rm ast}(S_N)$ (Equations (30) and (31)).
16. Calculate $V_{\rm ast}(S_R)$ (Equations (32) and (33)).
17. Calculate $S_{\rm corr}$ (Equation (25)). As a sanity check, the (robust) std of $S_{\rm corr}$ should be ≈ 1.
18. Search for local maxima in $S_{\rm corr}$. The peak value corresponds to the significance of the transient in units of sigmas.
19. As an alternative to steps 15 and 16, we can search all locations in D that correspond to statistically significant sources in $S_{\rm corr}$ (without astrometric contributions) for moving point sources using P_S (Equation (34)), measure their flux and astrometric variability and subtract them.
20. Select remaining sources with significance larger than some threshold, determined from the desired false alarm probability.
21. Calculate the flux of the transient candidates using Equations (41)–(43).

We note that all Fourier transforms in the paper are circular, and any statistic we compute should be regarded only on positions sufficiently far from the borders of the image (i.e., typically a few times the PSF width).

6. TESTS

There are several important challenges in testing any image differencing algorithm. Image subtraction in general is affected by many factors. Therefore, it is desirable to separate between external problems (e.g., non-perfect registration) and issues related to the subtraction itself (e.g., numerical stability). Therefore, we are using both simulations and real data to test our image differencing algorithm.

It is worthwhile to compare the new algorithm with existing methods. However, such a comparison is problematic, as other methods do not specify the matched filter for source detection. Furthermore, some of these methods depend on the selection of basis functions and kernel size. In addition, there are several ways to solve a system of linear equations (e.g., SVD) and

[16] Robust fitting is less sensitive to outliers. An example for a robust fitter is the `robustfit.m` function in MATLAB.

Table 1
Simulated Transients in the New Image

X (pix)	Y (pix)	Flux (electrons)
100	100	1500
120	120	1600
140	140	1800
160	160	2000
180	180	2200
200	200	2400
220	220	2600
240	240	2800
260	260	3000

Note. The position and mean flux of simulated transient sources in the new images in Figures 2–4.

these may influence the final outcome. Therefore, here our comparison with other methods is limited.

In Section 6.1, we present tests based on simulated data, while in Section 6.2 we discuss real images. The code we use is available as part of the Astronomy and Astrophysics package for MATLAB (Ofek 2014) described in Section 7.

6.1. Simulations

An important feature of our algorithm is its numerical stability. The best way to test this is on simulations because the input is fully controlled.

We simulated images of 512 by 512 pixels size, with a background level of 300 electrons, with Poisson noise. In each image, we simulated 100 stars with integrated flux taken from a flat distribution between 0 to 10^5 electrons and Poisson noise. In addition, we added to the new image nine transient sources with position and flux as listed in Table 1. In the first set of simulated images, the PSF of the sources in the images are symmetric Gaussians with sigma-widths of 2 and 3 pixels, for the reference and new images, respectively. Figure 2 shows, left to right (top), new image, the reference image, the proper subtraction image (D); (bottom) the matched-filtered image (S) threshold above 5σ, the Alard & Lupton (1998) subtraction of new minus reference, and the Alard & Lupton (1998) subtraction of reference minus new. The Alard & Lupton (1998) subtractions are based on the ISIS software (Alard & Lupton 1998). This figure demonstrates that while our image subtraction method is symmetric, the Alard & Lupton (1998) algorithm is not symmetric. In this case, it is working well in one direction, but subtraction artifacts are clearly visible (ringing due to deconvolution) in the other direction. Furthermore, thresholding our matched filter image above 5σ reveals only the simulated transients.

Next, we simulated images with the same parameters as in Figure 2, with asymmetric Gaussian PSF with sigma-widths of 2 by 4 pix in the new image and 4 by 2 pix in the reference image. Figure 3, is the same as Figure 2, but for these images. Again, the asymmetry of the Alard & Lupton (1998) family of methods is seen. Furthermore, in this case, the ringing due to deconvolution is seen in both the $N - R$ and $R - N$ subtractions.

One of the most important practical features of our new method is the ability to incorporate other types of noise into the detection process (e.g., source noise, astrometric noise, color-refraction noise). To demonstrate this, we repeated the first simulation (Figure 2), but this time with normally distributed astrometric noise with a standard deviation of 0.3 pix. Figure 4 shows, left to right (top), the new image, the reference image, the proper subtraction image (D); (bottom): the matched-filtered image (S) thresholded above 5σ, the source noise corrected and astrometric noise corrected matched-filtered image ($S_{\rm corr}$) thresholded above 5σ, and the Alard & Lupton (1998) subtraction of the new minus reference. In this case, the subtraction contains a large number of positive–negative residuals, but our $S_{\rm corr}$ image deals well with this astrometric noise, and only the simulated transients are detected.

6.2. Tests on Real Images

We tested the new method on imaging data available from the PTF[17] (Law et al. 2009; Rau et al. 2009) data release 2. The image processing is described in Laher et al. (2014) while the photometric calibration is discussed in Ofek et al. (2012).

Table 1 lists the various images on which we tested our algorithm. Registration, background subtraction and PSF estimation were performed using the code described in Section 7.

Figure 5 presents the image subtraction results of test 1. The top panels from left to right are the new image, the reference image, and the proper difference image D. The bottom panels from left to right are the matched filter corrected image ($S_{\rm corr}$) thresholded at 5σ, the Alard & Lupton (1998) subtraction of the $N - R$, and the Alard & Lupton (1998) subtraction of the $R - N$. Figure 5 also demonstrates that the Alard & Lupton (1998) subtraction is not symmetric to the exchange of R and N, while our method is. Specifically, the $R - N$ image of the Alard & Lupton (1998) has strong, and high amplitude, correlated noise.

On first glance, the Alard & Lupton (1998) $N - R$ image looks cosmetically good. However, on closer inspection we can see that this image has subtraction residuals with large amplitude. For example, Figure 6 shows a profile cut, at the location of the red line in Figure 5, in the proper subtraction image D and the Alard & Lupton (1998) subtraction ($N - R$). The images are normalized such that the standard deviation of the images is one. This figure clearly shows that while our algorithm behaves very well in the presence of stars, the Alard & Lupton (1998) subtraction has very large fluctuations. We note that the fact that the Alard & Lupton (1998) subtraction image is partially filtered is seen by eye (i.e., smoother noise)

Figure 7 is the same as Figure 5, but for the subtraction of images of test 2.

These images contain the bright galaxy M51, and SN 2011dh (Arcavi et al. 2011). We note that in D we clearly see residuals due to the mis-alignment of the images. However, these residuals are gone when we present $S_{\rm corr}$, which takes the source noise and astrometric noise (about $0\farcs2$ rms) into account. We further note that the astrometric residuals are less pronounced in the Alard & Lupton (1998) subtraction simply because these images are partially filtered, and therefore smoother. The transient candidate detected in the $S_{\rm corr}$ image above the 5σ threshold are SN 2011dh, cosmic rays, and bad pixels. S and $S_{\rm corr}$ are presented in Figure 8, just to give an impression of the importance of correcting S before detecting transients. Note that S and $S_{\rm corr}$ contain correlated noise and may mislead the human eye.

[17] http://www.ptf.caltech.edu/iptf

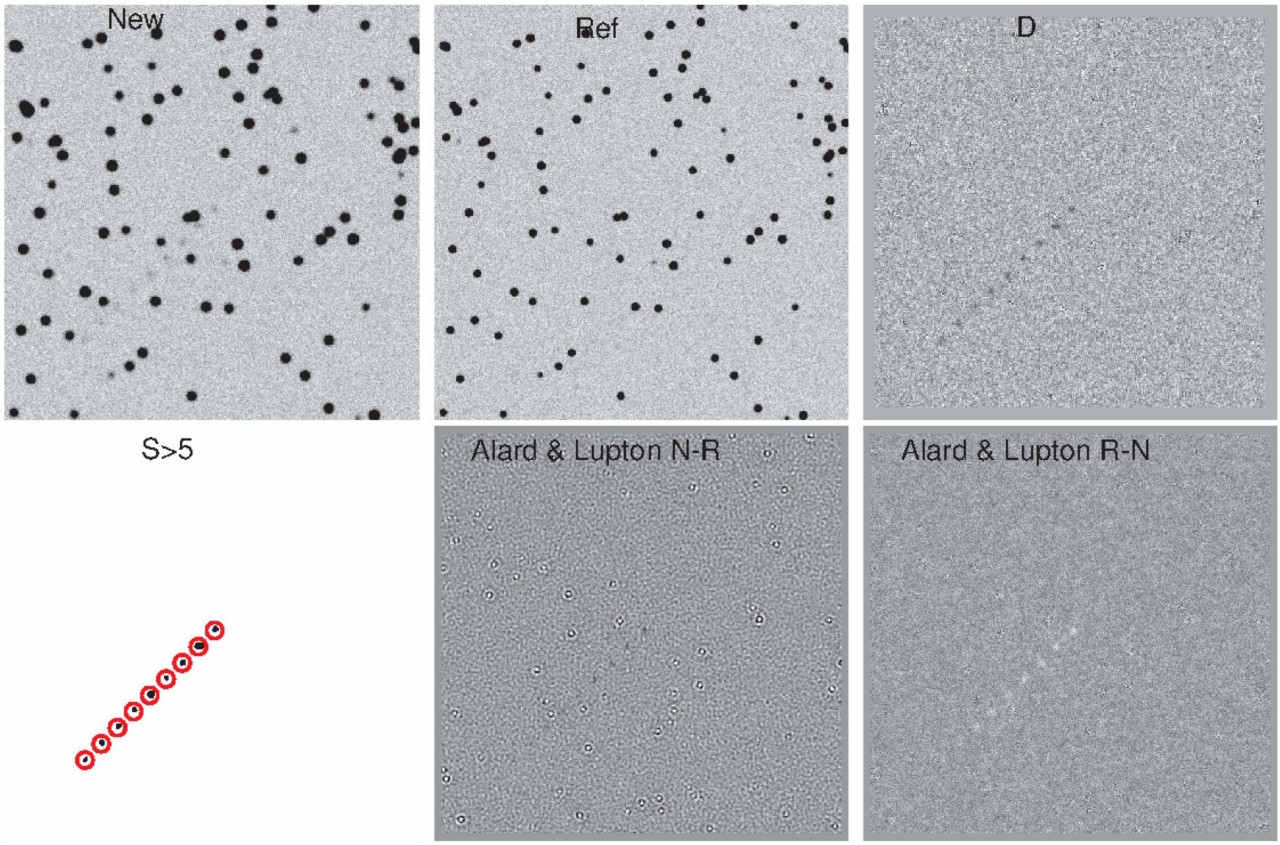

Figure 2. Subtraction of simulated images with symmetric Gaussian PSF with sigma-width of 2 and 3 pixels, for the reference and new images, respectively. Left to right (top): the new image, the reference image, the proper subtraction image (D); (bottom) the matched-filtered image (S) with 5σ threshold, the Alard & Lupton (1998) subtraction of new minus reference, and the Alard & Lupton (1998) subtraction of reference minus new. The position of the simulated transient sources in the thresholded matched-filtered image are marked by red circles. All the images are presented with inverted grayscale map.

7. CODE

We present two sets of codes based on MATLAB and Python. The MATLAB code contains functions to deal with all the image processing steps, including the registration and PSF estimation. The MATLAB code is available as part of the MATLAB Astronomy and Astrophysics package[18] (Ofek 2014). This code is under development and we expect that improved versions will be available in the future. The Python code[19] contains only a simple implementation of our algorithm that requires as input: fully registered images, as well as their PSF, background images, and variance images.

The main high-level MATLAB functions required for image subtraction are described with the online code. Some of these functions are discussed in Zackay & Ofek (2015a, 2015b). The implementation details related to some of these utilities are further discussed in Section 8.

8. IMPLEMENTATION DETAILS

Given background-subtracted images, their variance, PSF, and flux-based zero points ratio, our image subtraction method is presented using a closed-form formula. Therefore, the implementation of this method is simple and rigorous, and does not require special attention. However, like any other method for image subtraction, this technique is sensitive to the steps taken prior to the image subtraction (e.g., registration; flux matching).

Here we discuss some of the details that can greatly influence the successful application of any image subtraction algorithm.

8.1. Background and Variance Estimation

The background and variance in real wide-field-of-view astronomical images cannot be treated as constants over the entire field of view. Therefore, we suggest to estimate them locally and interpolate. To estimate the background and variance, one needs to make sure that the estimators are not biased by stars or galaxies. Following Zackay & Ofek (2015a, 2015b), we suggest to fit a Gaussian to the histogram of the image pixels in small regions[20], and to reject from the fitting process pixels with high values (e.g., the upper 10% of pixel values). Regions containing large galaxies or complex background may require special treatment.

8.2. PSF Estimation and Spatial Variations

We note that Equations (12) and (13) are roughly linear to perturbations in the PSF, compared with the real PSF. Among the complications that may affect the PSF measurement are pixelization, interpolation and the resampling grid. Furthermore, the PSF is likely not constant spatially and it also may change with intensity due to charge self repulsion. This

[18] http://webhome.weizmann.ac.il/home/eofek/matlab/
[19] https://sites.google.com/site/barakzackayhomepage/
[20] We are currently using 256×256 arcsec^2 blocks.

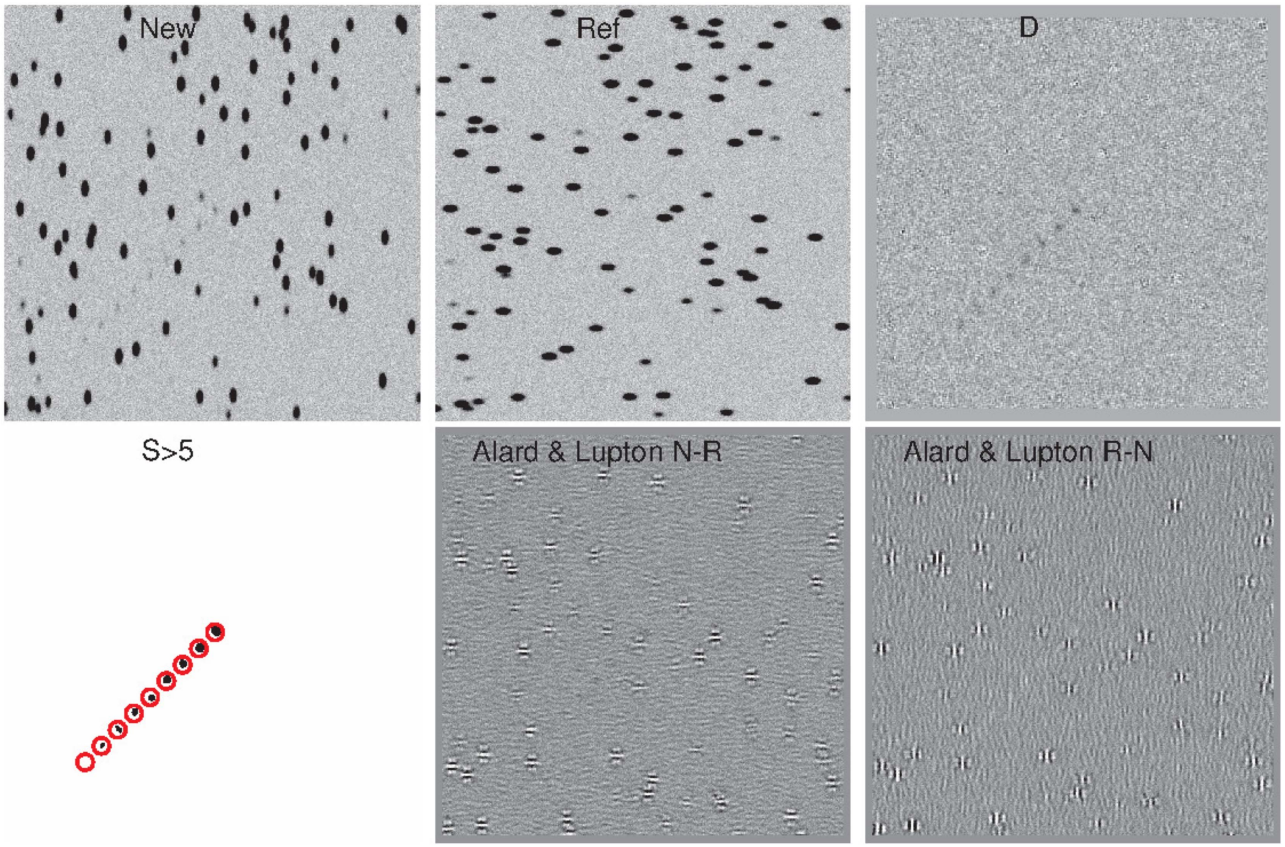

Figure 3. Same as Figure 2, but for the subtraction of simulated images with asymmetric Gaussian PSF with sigma-widths of 2 by 4 pix in the new image and 4 by 2 pix in the reference image.

Table 2
List of Tests on Real Images

Test	Field/CCD	Size (pix)	N	R	FWHM_N (arcsec)	FWHM_R (arcsec)
1	100031/04	560 × 560	2012-12-20.4134	proper	5.4	2.9
2	100031/11	1000 × 1000	2011-08-08.1839	2011-04-12.1865	2.5	2.9

Note. List of tests on real images. "proper" indicates a reference image that was constructed using proper coaddition (Zackay & Ofek 2015b).

specifically may lead to the brighter-fatter effect (e.g., Walter 2015).

In some cases, the PSF may vary over the field of view. The simplest approach is to divide the image to smaller images in which the PSF is approximately constant. These sub-images can be as small as four times the PSF size. Since the convolution operation is local, it is straightforward to incorporate a spatially variable PSF into any subtraction method (e.g., Alard 2000).

8.3. Interpolation

The registration step requires us to interpolate one of the images into a new coordinate grid. If the PSF is Nyquist sampled (band limited), then one can use the Whittaker-Shannon interpolation formula (sometimes called sinc interpolation) without losing information due to the interpolation process. In practice images contain under-sampled radiation hits. Whittaker-Shannon (sinc) interpolation of such images will generate artifacts over the entire image. Therefore, a more practical approach is to use the Lanczos interpolation.

However, if the PSF is undersampled, interpolation will lead to variation in the PSF shape, which depends on the position of the source within the pixel (pixel phase). Such an effect may cause severe problems to any subtraction method. One simple way to deal with this problem is to add a noise term to the denominator of S_{corr} (Equation (25)) that takes into account the extra noise induced by the pixel-phase dependent PSF variations.

8.4. Registration

Registration is a critical step for any image differencing technique. Any leftover registration imperfection residuals between the new and reference image will lead to improper subtraction, subtraction artifacts, and eventually to false detections. In Sections 3.3 and 3.4, we discuss how registration errors, color refraction, and astrometric scintillations can be treated. However, it is still desirable to minimize any registration errors prior to subtraction.

In many cases, affine transformations are not enough to map between the two images. The main reasons include differential

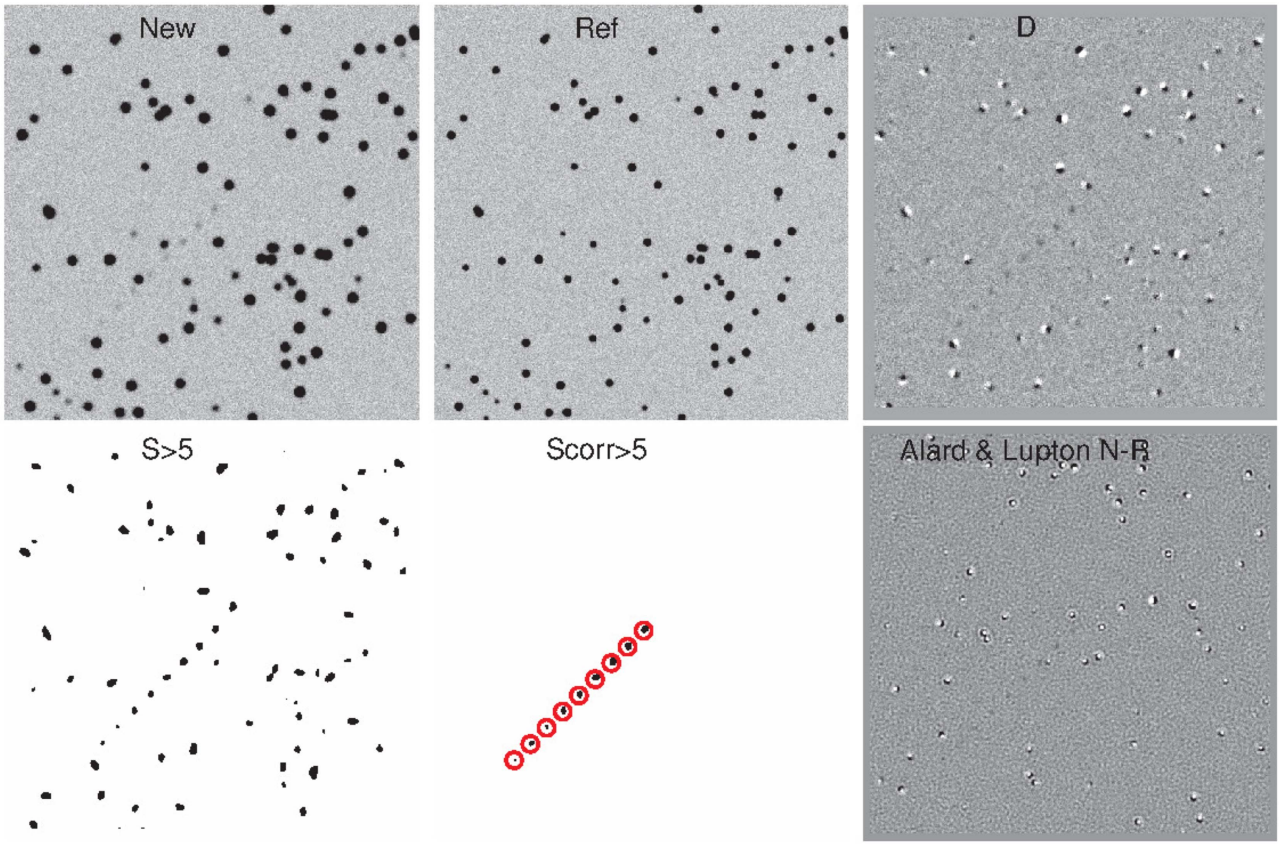

Figure 4. Subtraction of simulated images with 0.3 pix (rms) astrometric noise and symmetric Gaussian PSF with sigma-width of 2 and 3 pixels, for the reference and new images, respectively. Left to right (top): the new image, the reference image, and the proper subtraction image (D). Left to right (bottom): the matched-filtered image (S) threshold above 5σ, the source noise corrected and astrometric noise corrected matched-filtered image ($S_{\rm corr}$) threshold above 5-σ, and the Alard & Lupton (1998) subtraction of the new minus reference. The position of the simulated transient sources in the thresholded matched-filtered image are marked by red circles.

atmospheric refraction, differential aberration of light, and high-order optical distortions.

Usually, when images are taken with the same system and the same on-sky pointing, optical distortions will not play an important role because their effect on the two images is almost identical.

The amplitude of differential atmospheric refraction can be as high as $8''$ deg^{-1}. Figure 9 shows the amplitude of differential atmospheric refraction as a function of altitude. Since the direction of the atmospheric refraction is known very well, the best way to with the distortions caused by the atmosphere is to add to the affine transformation terms that fit the atmospheric refraction amplitude with its known direction (i.e., the parallactic angle). Unfortunately, most astrometric and registration packages do not support distortions of this form, and instead they absorb the refraction correction into high-order polynomials. Furthermore, the current WCS header keywords do not support this kind of transformation. Our code described in Section 7 does support this transformation.

We note that atmospheric refraction distortions are detectable even on small angular scales. For example, this effect can reach $0\farcs1$ arcmin^{-1} at an altitude of 20 deg. In any case, in order to minimize any higher order distortions, it is recommended to divide the image to small sections (say 10 by 10 arcmin).

The typical amplitude of differential aberration of light (due to the Earth motion) is of the order of $\sim 0\farcs2$ deg^{-1}. This is small enough to be ignored in some cases. However, since the effect of aberration is fully predictable it is straightforward to incorporate it into the transformation. As far as we know popular image registration (and astrometric) packages ignore the aberration of light.

8.5. Astrometric Scintillations

Astrometric registration of ground-based imaging is typically limited by astrometric scintillation induced by the Earth atmosphere. An order of magnitude estimate for the amplitude of astrometric scintillation is

$$\sigma_{\rm scint} \sim \frac{\rm FWHM}{\sqrt{t_{\rm int}/t_{\rm scint}}}, \quad (51)$$

where FWHM is the PSF FWHM, $t_{\rm int}$ is the integration time, and $t_{\rm scint}$ is the correlation timescale of the tip/tilt term of the atmospheric scintillations. For example, assuming FWHM$=2''$, $t_{\rm int}=60$ s, and $t_{\rm scint}=0.03$ s, we get $\sigma_{\rm scint}\sim 40$ mas. This can be an order of magnitude larger than the astrometric noise induced by the Poisson noise of bright stars. In practice, this noise depends on the angular scale (see e.g., Lindegren 1980; Shao & Colavita 1992).

This kind of astrometric noise is hard to remove, and therefore we expect that bright stars will always have some leftover residuals in the subtraction process. However, we presented two methods to solve this problem in Sections 3.3 and 3.4.

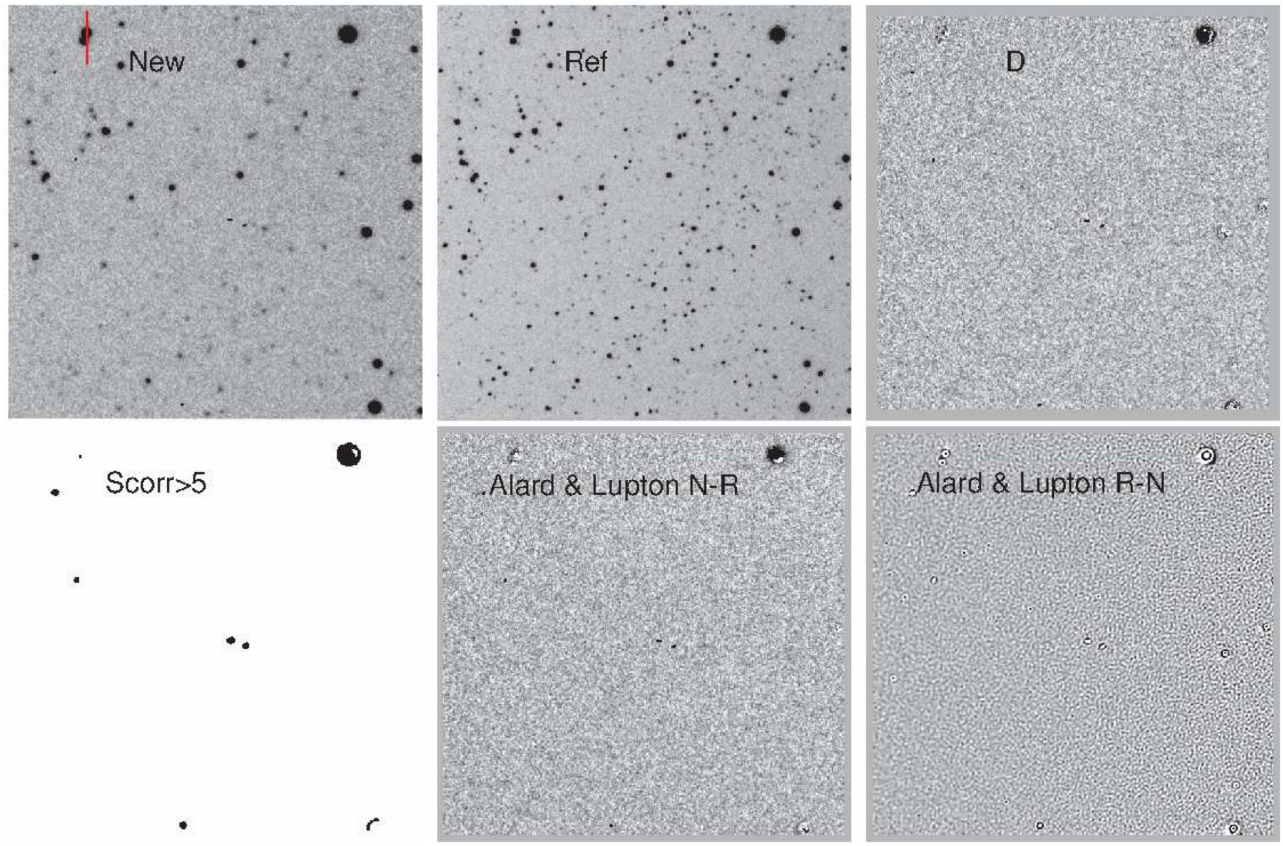

Figure 5. Image subtraction results for test 1 (Table 2). Left to right (top): the new image, the reference image, and the proper subtraction image D; (bottom) the matched filter corrected difference image ($S_{\rm corr}$) filtered at 5σ, the Alard & Lupton (1998) ISIS subtraction of the new minus the reference, and the ISIS subtraction of the reference minus the new. All the images are presented with the inverted grayscale map. The red line (in the new panel) indicates the position of the profile cut we present in Figure 6. In the $S_{\rm corr} > 5$ map, CR1–CR5 indicate the position of cosmic rays detected by our algorithm, while the two bright residuals on the right part of the image are due to saturated stars. The residual at the top left has a significance of 5.7-σ and it is at the interface between two bright stars. The mechanism that generate this particular residual is discussed in Section 8.7.

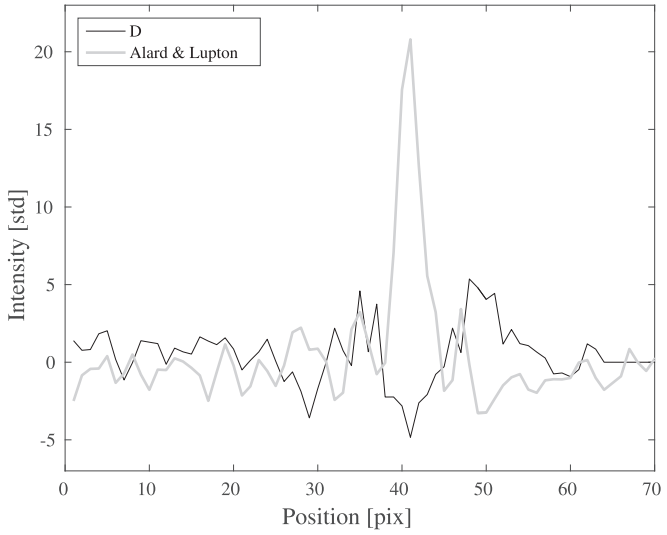

Figure 6. Profile cut, at the position of the red line in Figure 5, in the proper subtraction image D (black line) and the Alard & Lupton (1998) subtraction ($N - R$; gray line). The images are normalized such that the standard deviation of the images is unity. This demonstrates that in the presence of bright stars, the fluctuations in our subtraction image are modest, while the residuals in the Alard & Lupton (1998) subtractions are large. We note that the D image is not filtered, while the Alard & Lupton (1998) subtraction is partially filtered. Therefore, the noise properties of D, relative to the Alard & Lupton (1998) subtraction, are even better than indicated from this plot.

8.6. Color Refraction

The atmospheric refraction is color dependent and hence sources with different spectra will suffer different refraction at the same airmass. Figure 10 presents the relative amplitude of color refraction, in different bands, between an O5V star and an M5V star and between an A0V star and an M5V star, as a function of altitude.

We suggest three solutions to this issue. (1) Construct reference images for several airmass ranges. Since color refraction is symmetric around the meridian, one needs to construct such reference images separately for observations conducted east and west of the meridian. (2) Calculate the variance induced by this effect and introduce it as an extra term in the denominator of $S_{\rm corr}$ (Equation (25)). (3) Fit the astrometric shift for each residual in D using the scheme presented in Section 3.4. The last option is likely the best approach.

8.7. Additional Sources of Noise

There may be additional sources of noise that can influence image subtraction. An example for a rare problem we encountered in our simulations and real images is that if a binary star has uncorrelated astrometric noise[21] this may affect

[21] In reality, this is rare as both registration errors and astrometric scintillation noise (but not the Poisson noise) are correlated on short angular scales.

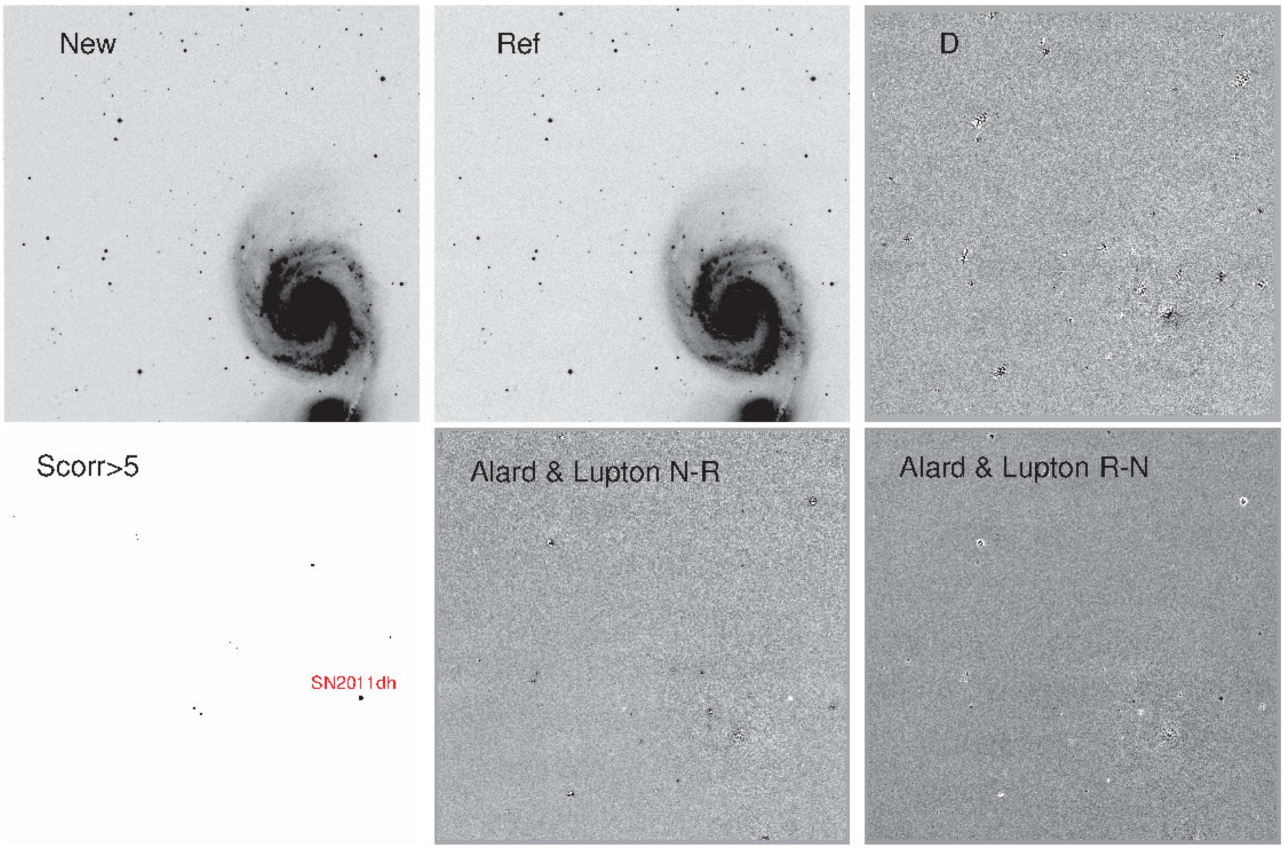

Figure 7. Same as Figure 5, but for the test 2 images, containing the bright galaxy M51 and SN 2011dh. The detected sources in $S_{\rm corr} > 5$ are SN 2011dh, particle hits, and bad pixels.

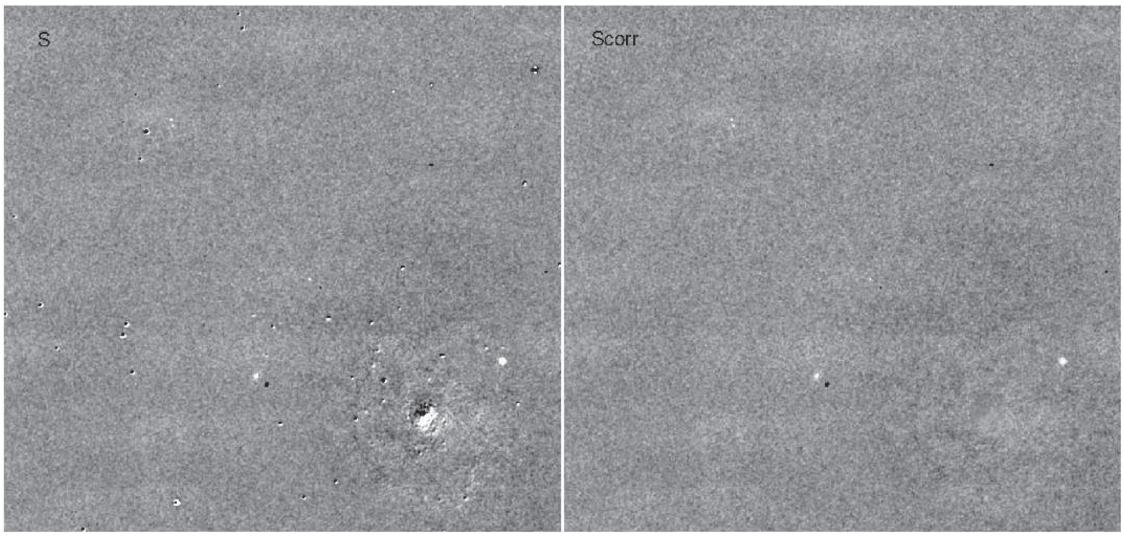

Figure 8. Score map S (left image) and the corrected score map $S_{\rm corr}$ (right image) corresponding to the subtraction shown in Figure 7. Note that albeit a substantial registration error (that could have been done better), the corrected image does not contain false alarms. Also note that in S it is very difficult to distinguish between an image artifact to a real source, while in D it is almost trivial. We note that S and $S_{\rm corr}$ are images with correlated noise. Detecting sources in S and $S_{\rm corr}$ should be done by searching for local maxima and reading their value. The value of the local maxima divided by the (local) standard deviation corresponds to the detection significance in units of sigmas (or more formally to the false alarm probability via the survival function of the normal distribution). Images grayscale is between -7 to $+7$ standard deviations of the images. With the exception of SN 2011dh (see Figure 7), all the significant residuals are due to cosmic rays and bad pixels.

the calculation of the gradient image (Equations (30)–(32); see Figure 5 for example). In principle, such problems can be accounted for in $S_{\rm corr}$; however, one needs to identify these issues. Therefore, successful implementation of this method requires large-scale tests on real data. Such tests are underway, and this may be further discussed in future publications.

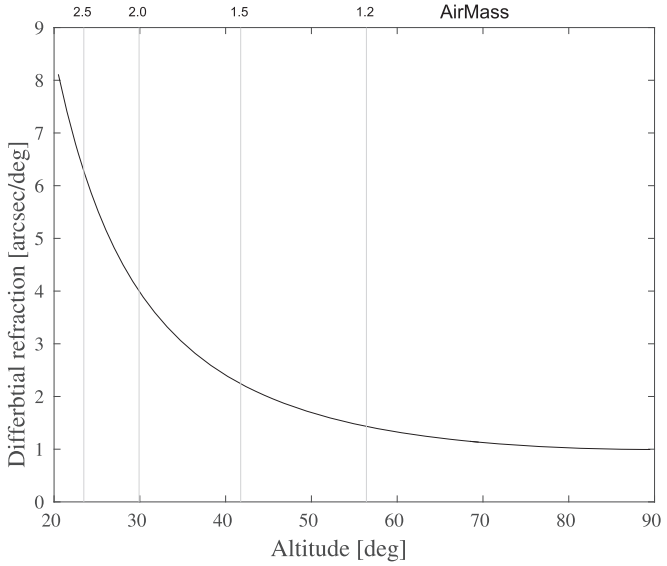

Figure 9. Differential atmospheric refraction (in the altitude direction), in units of arcsec per deg. Calculated using the code in Ofek (2014) and formulae provided in Filippenko et al. (1982), for a wavelength of 5000 Å, temperature of 15° C, pressure of 760 mm Hg and partial water vapor pressure of 8 mm Hg.

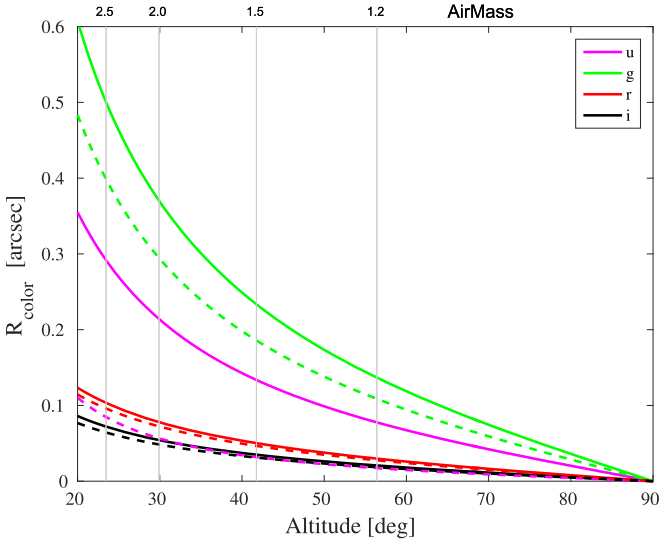

Figure 10. Solid lines represent the difference in color refraction (in the altitude direction) in arcsec, as a function of altitude, between an O5V star and M5V star. The various colors correspond to different filters (see the legend). The dashed lines show the same, but for the difference between an A0V star and M5V star. The calculation includes the atmospheric extinction (at Kitt Peak), and uses stellar spectra (adopted from Pickles 1998). Atmospheric conditions are the same as in Figure 9.

9. SUMMARY

Current popular image subtraction methods have several important limitations, including non-optimality, numerical instability in some cases, some of the methods use a matrix inversion, which is slow to calculate. Most importantly, these methods do not provide a closed-form formula for the calculation of the significance of a transient candidate. Moreover, in some cases, due to the numerical instability of some of the methods it is not possible to calculate, even numerically, the significance of a transient candidate. This undermines any automatic transient detection and classification, and may be a considerable obstacle for future surveys.

We present closed-form transient detection and image subtraction statistics that potentially solve all of the above problems and have the following properties.

1. The transient detection statistic is mathematically proven to be optimal in the background-dominated noise limit.
2. Both statistics are numerically stable for any pair of input images.
3. For accurately registered, adequately sampled images, these statistics do not leave any subtraction residuals or deconvolution artifacts.
4. It is possible to correct the transient detection statistic to be resilient to registration errors, color-refraction errors, and any noise for which a model can be constructed.
5. We can assign credible detection significance for newly found transients.
6. The proper subtraction image has white noise in the background-dominated-noise limit. This makes it attractive for more complex measurements and visualization.
7. The proper subtraction statistic is a sufficient statistic for any further statistical test on the difference image. In particular, it allows to distinguish particle hits and other image artifacts from real transients.
8. Both statistics are symmetric to the exchange of the new and reference images.
9. Both statistics are fast to calculate—at least an order of magnitude faster to compute than popular methods.
10. Both statistics are given in closed form and they are straightforward to implement.
11. The proper subtraction statistic allows us to search for small astrometric changes between the new and reference images, even in arbitrarily crowded regions.
12. The same statistics are also optimal for flux measurements in the background-noise dominated limit.
13. We show that the optimal way to prepare a reference image is the proper image coaddition statistic presented in Zackay & Ofek (2015b).

We demonstrate this method on simulated data and real observations from the Palomar Transient Factory data release 2. A summary of the algorithm and equations are presented in Section 5, while a discussion regarding the implementation is in Section 8. We briefly describe our MATLAB and Python code that implement this method and are available online.

We conclude that this image differencing algorithm has the potential to solve most of the challenges of astronomical image subtraction. However, testing if this method is indeed capable of completely removing the need for post-subtraction processing (e.g., human scanners) requires a considerable research effort and tests on large data sets. Such tests are underway.

We thank Ora Zackay, Zeljko Ivesic, Robert Lupton, and Assaf Horesh for many discussions. This paper is based on observations obtained with the Samuel Oschin Telescope as part of the Palomar Transient Factory project, a scientific collaboration between the California Institute of Technology, Columbia University, Las Cumbres Observatory, the Lawrence Berkeley National Laboratory, the National Energy Research Scientific Computing Center, the University of Oxford, and the Weizmann Institute of Science. B.Z. is grateful for receiving

the Clore fellowship. E.O.O. is incumbent of the Arye Dissentshik career development chair and is grateful for support by grants from the Willner Family Leadership Institute Ilan Gluzman (Secaucus NJ), Israel Science Foundation, Minerva, Weizmann-UK, and the I-Core program by the Israeli Committee for Planning and Budgeting and the Israel Science Foundation (ISF). A.G. acknowledges support from the I-Core program "The Quantum universe," as well as from the Kimmel Award.

APPENDIX A
FULL DERIVATION OF THE IMAGE SUBTRACTION STATISTICS

Let R and N be the background-subtracted reference image and background-subtracted new image, respectively. Denote by T the background-subtracted true constant sky image. Denote by P_r and P_n the PSFs of the reference image and the new image, respectively. P_r and P_n are normalized to have unit sum.

Writing the expression for the reference image

$$R = F_r T \otimes P_r + \epsilon_r, \tag{52}$$

where ϵ_r is the additive noise component of the image R. Given the null hypothesis, $\mathcal{H}_0$, that states there are no new sources in the new image we can write

$$N_{|\mathcal{H}_0} = F_n T \otimes P_n + \epsilon_n. \tag{53}$$

Given the alternative hypothesis, $\mathcal{H}_1(q, \alpha)$, which states that there is a new point source at position q with flux α, we can write

$$N_{|\mathcal{H}_1(q,\alpha)} = F_n T \otimes P_n + F_n \alpha \delta(q) \otimes P_n + \epsilon_n, \tag{54}$$

where $\delta(q)$ denotes a two-dimensional image with one at position q, and zero otherwise. Assuming that the images are background subtracted, and that the dominant source of noise is the background noise, ϵ_r and ϵ_n both satisfy that all pairs of pixels are uncorrelated—i.e., that for all pairs of pixels x_1, x_2 for which $x_1 \neq x_2$:

$$\mathrm{Cov}(\epsilon_r[x_1], \epsilon_r[x_2]) = 0, \mathrm{Cov}(\epsilon_n[x_1], \epsilon_n[x_2]) = 0, \tag{55}$$

and that all pixels have spatially uniform variance[22]:

$$V(\epsilon_r[x]) = \sigma_r^2, V(\epsilon_n[x]) = \sigma_n^2. \tag{56}$$

Because both hypotheses are simple, we can use the Neyman–Pearson lemma (Neyman & Pearson 1933), that states that the most powerful statistic for deciding between two simple hypotheses is the likelihood ratio test:

$$\mathcal{L}(q, \alpha) = \frac{\mathcal{P}(N, R|\mathcal{H}_0)}{\mathcal{P}(N, R|\mathcal{H}_1(q, \alpha))}, \tag{57}$$

where $\mathcal{P}$ denotes probability. A critical point is that we do not have any prior information or assumptions on T. Therefore, we cannot calculate the probabilities $\mathcal{P}(N, R|\mathcal{H}_0)$ and $\mathcal{P}(N, R|\mathcal{H}_1(q, \alpha))$ directly. However, we can calculate their ratio by developing the expression using the law of conditional probabilities

$$\mathcal{L}(q, \alpha) = \frac{\mathcal{P}(N|R, \mathcal{H}_0)\mathcal{P}(R|\mathcal{H}_0)}{\mathcal{P}(N|R, \mathcal{H}_1(q, \alpha))\mathcal{P}(R|\mathcal{H}_1(q, \alpha))}. \tag{58}$$

Using the fact that both $\mathcal{H}_0$ and $\mathcal{H}_1(q, \alpha)$ state the same probabilistic model for R (and therefore will assign the same likelihood for observing R), we can further simplify:

$$\mathcal{L}(q, \alpha) = \frac{\mathcal{P}(N|R, \mathcal{H}_0)}{\mathcal{P}(N|R, \mathcal{H}_1(q, \alpha))}. \tag{59}$$

To calculate $\mathcal{P}(N|R, \mathcal{H}_0)$, we examine the statistical behavior of the Fourier transforms of N and R given both hypotheses, and assume that the images are background-noise dominated. Using the fact that the Fourier transform of white noise is itself a white noise, we know the exact noise properties of the Fourier transform of both R, N given both hypotheses:

$$\widehat{N}_{|\mathcal{H}_0} = F_n \widehat{T} \widehat{P_n} + \widehat{\epsilon_n}, \tag{60}$$

$$\widehat{N}_{|\mathcal{H}_1(q,\alpha)} = F_n (\widehat{T} + \alpha \widehat{\delta(q)}) \widehat{P_n} + \widehat{\epsilon_n}, \tag{61}$$

$$\widehat{R}_{|\mathcal{H}_0} = \widehat{R}_{|\mathcal{H}_1(q,\alpha)} = F_r \widehat{T} \widehat{P_r} + \widehat{\epsilon_r}, \tag{62}$$

where the $\widehat{\ }$ accent denotes Fourier transform and both $\widehat{\epsilon_n}$ and $\widehat{\epsilon_r}$ are complex white Gaussian noise.[23]

Using the fact that $\widehat{R}$ is measured, we can invert its probabilistic model to obtain a model for T:

$$\widehat{T} = \frac{\widehat{R}}{F_r \widehat{P_r}} - \frac{\widehat{\epsilon_r}}{F_r \widehat{P_r}}. \tag{63}$$

We note that the model for $\hat{T}$ is not calculated, and will be used only as a substitution step. This step is valid if $|\hat{P}_r| \neq 0$. In practice, this can be verified either by making sure that the model for $\widehat{P_r}$ does not contain absolute zeros, or by adding a small number (i.e., the computer precision) to the denominator of Equation (13).

Using this expression for $\widehat{T}$, we can write a probabilistic model for N given R and $\mathcal{H}_0$:

$$\widehat{N}_{|\widehat{R}} = \frac{\widehat{R}}{F_r \widehat{P_r}} F_n \widehat{P_n} - \frac{\widehat{\epsilon_r}}{F_r \widehat{P_r}} F_n \widehat{P_n} + \widehat{\epsilon_n}. \tag{64}$$

Given this model for T and assuming the noise is Gaussian, we can calculate the probability to observe N (this is the χ^2 up to a factor of two):

$$\log(\mathcal{P}[\widehat{N}|\widehat{R}, \mathcal{H}_0]) = \sum_f \frac{-\left| \widehat{N} - \frac{F_n \widehat{P_n} \widehat{R}}{F_r \widehat{P_r}} \right|^2}{2V\left(\widehat{\epsilon_n} + \frac{F_n \widehat{P_n} \widehat{\epsilon_r}}{F_r \widehat{P_r}}\right)}, \tag{65}$$

Using the linearity and scalar multiplication properties of the variance and simplifying we get

$$\log(\mathcal{P}[\widehat{N}|\widehat{R}, \mathcal{H}_0]) = \frac{1}{2} \sum_f \frac{-|F_r \widehat{P_r} \widehat{N} - F_n \widehat{P_n} \widehat{R}|^2}{\sigma_n^2 F_r^2 |\widehat{P_r}|^2 + \sigma_r^2 F_n^2 |\widehat{P_n}|^2}, \tag{66}$$

Similarly, given $\mathcal{H}_1$, we can write

$$\log(\mathcal{P}[\widehat{N}|\widehat{R}, \mathcal{H}_1(q, \alpha)]) \tag{67}$$

[22] In practice, this assumption can be relaxed.

[23] The noise in the Fourier transform of an image with white noise, is white except for the obvious symmetry $\widehat{\epsilon_n}(f_1, f_2) = \overline{\widehat{\epsilon_n}(-f_1, -f_2)}$, where the over line denotes complex conjugation. This symmetry is due to the fact that the input images are real.

$$= -\frac{1}{2}\sum_f \frac{|F_r\widehat{P_r}\widehat{N} - F_n\widehat{P_n}\widehat{R} - \alpha F_n F_r \widehat{P_n}\widehat{P_r}\widehat{\delta(q)}|^2}{\sigma_n^2 F_r^2|\widehat{P_r}|^2 + \sigma_r^2 F_n^2|\widehat{P_n}|^2}. \quad (68)$$

Because the hypothesis $\mathcal{H}_1$ has the free parameter α, we cannot use the Neyman–Pearson lemma directly. In order to preserve the optimality proof of the test, we will have to construct a test that is uniformly most powerful with respect to α. The procedure for doing this is as follows. Identify a scalar sufficient statistic of the data $S(q)$ with respect to $\mathcal{H}_1(q, \alpha)$ for every α. Then, show that this statistic satisfies the requirement for the Karlin–Rubin theorem Karlin & Rubin (1956), that for any $\alpha_1 > \alpha_0$, and for any value of $S(q) = x$, the likelihood ratio

$$l(x) = \frac{\mathcal{P}(S(q) = x|\alpha_1)}{\mathcal{P}(S(q) = x|\alpha_0)}, \quad (69)$$

is a non-decreasing function of x. Then, we are guaranteed that the test $S(q) > \eta$ for some threshold η that determines the false positive rate of the test, is uniformly most powerful for the decision problem between $\mathcal{H}_0$ and $\mathcal{H}_1(q, \alpha)$ for any α. To construct the statistic $S(q)$, we open the parentheses of Equation (68) using $|a + b|^2 = |a|^2 + |b|^2 + 2\Re[a\bar{b}]$, where $\Re$ is the real number operator, and removing the $|a|^2$ and $|b|^2$ terms because they do not depend on both α and the data (this is allowed by the Fisher–Neyman factorization criteria for constructing a sufficient statistic, that is presented in full form in Appendix E)

$$\log(\mathcal{L}(q, \alpha)) \quad (70)$$

$$= \sum_f \frac{\Re[(F_r\widehat{P_r}\widehat{N} - F_n\widehat{P_n}\widehat{R})\overline{\alpha F_r F_n \widehat{P_n}\widehat{P_r}\widehat{\delta(q)}}]}{\sigma_n^2 F_r^2|\widehat{P_r}|^2 + \sigma_r^2 F_n^2|\widehat{P_n}|^2}. \quad (71)$$

Noticing that α enters only as a scalar multiplier to the remaining expression, we can identify the sufficient statistic $S(q)$:

$$S(q) \equiv \frac{\log(\mathcal{L}(q, \alpha))}{\alpha}, \quad (72)$$

Trivially, the expression on Equation (70) satisfies the monotonicity requirement for the Karlin–Rubin theorem, and therefore $S(q)$ is uniformly most powerful with respect to testing $\mathcal{H}_1(q, \alpha)$ for any α, and therefore is the optimal transient detection statistic for detecting transients at position q.

In order to express the same score in term of intuitive quantities, we define the proper subtraction image:

$$\widehat{D} = \frac{(F_r\widehat{P_r}\widehat{N} - F_n\widehat{P_n}\widehat{R})}{\sqrt{\sigma_n^2 F_r^2|\widehat{P_r}|^2 + \sigma_r^2 F_n^2|\widehat{P_n}|^2}}. \quad (73)$$

The PSF for transient detection:

$$\widehat{P_D} = \frac{F_r F_n \widehat{P_r}\widehat{P_n}}{F_D\sqrt{\sigma_n^2 F_r^2|\widehat{P_r}|^2 + \sigma_r^2 F_n^2|\widehat{P_n}|^2}}, \quad (74)$$

and the normalization:

$$F_D = \frac{F_n F_r}{\sqrt{F_n\sigma_r^2 + F_r\sigma_n^2}}. \quad (75)$$

We note that F_D can be derived by substituting 1 into $\widehat{P_n}$ and $\widehat{P_r}$ in the expression for $\widehat{P_D}$.

In the background-noise dominated limit, D has white noise (see Section 4.5). The score $S(q)$ can now be expressed by

$$S(q) = \Re\left[F_D\sum_f \widehat{D}\overline{\widehat{P_D}\widehat{\delta(q)}}\right]. \quad (76)$$

Expressing this in real space using the convolution theorem we get

$$S(q) = F_D\Re[D \otimes \overleftarrow{P_D} \otimes \overleftarrow{\delta(q)}](0). \quad (77)$$

Noticing that both D and P_D contain only real numbers, the real operator can be removed. Convolution with a delta function is just the shift operator; therefore, the expression for $S(q)$ can be simplified even further to be

$$S(q) = [F_D D \otimes \overleftarrow{P_D}](q). \quad (78)$$

The expression for its Fourier transform is then expressed by

$$\widehat{S} = \widehat{D}\overline{\widehat{P_D}} = \frac{F_n F_r^2\overline{\widehat{P_n}}|\widehat{P_r}|^2\widehat{N} - F_r F_n^2\overline{\widehat{P_r}}|\widehat{P_n}|^2\widehat{R}}{\sigma_r^2 F_n^2|\widehat{P_n}|^2 + \sigma_n^2 F_r^2|\widehat{P_r}|^2}. \quad (79)$$

This is the final form of the optimal transient detection statistic. An alternative form for this expression can be written as

$$\widehat{S} = \frac{F_n F_r^2\frac{\overline{\widehat{P_n}}}{\sigma_n^2}\frac{|\widehat{P_r}|^2}{\sigma_r^2}\widehat{N} - F_r F_n^2\frac{\overline{\widehat{P_r}}}{\sigma_r^2}\frac{|\widehat{P_n}|^2}{\sigma_n^2}\widehat{R}}{F_n^2\frac{|\widehat{P_n}|^2}{\sigma_n^2} + F_r^2\frac{|\widehat{P_r}|^2}{\sigma_r^2}}. \quad (80)$$

APPENDIX B
CONSTRUCTION OF THE REFERENCE IMAGE

Extending the statistical framework to the situations in which we are given a set of references, we seek to find the optimal transient detection statistic given all of the references. Each reference image out of a total of J images is given by

$$R_j = F_j P_j \otimes T + \epsilon_j. \quad (81)$$

A certain new image N is measured, and we want to determine which of the following is true, $\mathcal{H}_0$:

$$N = F_n P_n \otimes T + \epsilon_n, \quad (82)$$

or $\mathcal{H}_1(q)$:

$$N = F_n P_n \otimes (T + \delta(q)) + \epsilon_n. \quad (83)$$

As in the previous section, we are trying to test between two simple hypotheses. Therefore, the optimal test statistic is the log-likelihood ratio test (Neyman & Pearson 1933)

$$\mathcal{L}(q, \alpha) = \frac{\mathcal{P}(N, R_1, \ldots, R_J|\mathcal{H}_0)}{\mathcal{P}(N, R_1, \ldots, R_J|\mathcal{H}_1(q, \alpha))}. \quad (84)$$

As before, we can use the law of conditional probabilities, and the fact that $\mathcal{H}_0$ and $\mathcal{H}_1$ predict the same likelihood to all references:

$$\mathcal{L}(q, \alpha) = \frac{\mathcal{P}(N|R_1, \ldots, R_J, \mathcal{H}_0)}{\mathcal{P}(N|R_1, \ldots, R_J, \mathcal{H}_1(q, \alpha))}. \quad (85)$$

In order to calculate the conditional probabilities, we need a probabilistic model for N that does not contain T. This could be

achieved by using all references to get the best statistical model for T.

As in the previous section, this can be more easily formulated by stating the hypotheses for the images in the Fourier plane:

$$\widehat{N}_{|\mathcal{H}_0} = \widehat{T}\widehat{P_n} + \widehat{\epsilon_n}, \tag{86}$$

$$\widehat{N}_{|\mathcal{H}_1(q,\alpha)} = (\widehat{T} + \alpha\widehat{\delta(q)})\widehat{P_n} + \widehat{\epsilon_n}, \tag{87}$$

$$\widehat{R_j}_{|\mathcal{H}_0} = \widehat{R_j}_{|\mathcal{H}_1(q,\alpha)} = \widehat{T}\widehat{P_j} + \widehat{\epsilon_j}. \tag{88}$$

Following Appendix A, we can continue to develop this in the long way into the correct difference image and the correct transient detection statistic. However, we can take a shortcut. The key observation we make, is that we can cast all the information in the reference images into a statistical model for $\widehat{T}$. Using the result from the appendix of Zackay & Ofek (2015a; paper I in the series on coaddition), the choice that maximizes the S/N is the weighted addition of all the sources of information on $\widehat{T}(f)$:

$$\widehat{T} = \frac{\sum_j \frac{F_j\overline{\widehat{P_j}}}{\sigma_j^2}\widehat{R_j}}{\sum_j \frac{F_j^2|\widehat{P_j}|^2}{\sigma_j^2}} + \widehat{\epsilon_T}. \tag{89}$$

Where we have denoted the noise contribution from all the reference images by $\widehat{\epsilon_T}$. Calculating its variance we get

$$V[\widehat{\epsilon_T}] = \frac{1}{\sum_j \frac{F_j^2|\widehat{P_j}|^2}{\sigma_j^2}} \equiv \frac{1}{F_r^2|\widehat{P_r}|^2}, \tag{90}$$

where we have defined

$$F_r = \sqrt{\sum_j \frac{F_j^2}{\sigma_j^2}}, \quad \widehat{P_r} = \frac{1}{F_r}\sqrt{\sum_j \frac{F_j^2}{\sigma_j^2}|\widehat{P_j}|^2}. \tag{91}$$

Given these choices and the template of Equation (63), we find the formula for the coaddition of the reference images:

$$\widehat{R} = \frac{\sum_j F_j \frac{\overline{\widehat{P_j}}}{\sigma_j^2}\widehat{R_j}}{\sqrt{\sum_j F_j^2 \frac{|\widehat{P_j}|^2}{\sigma_j^2}}}. \tag{92}$$

Here $\widehat{\sigma_R} = 1$. Since R, P_r, and T satisfy Equation (52), we have a single reference image that complies with the requirements of the statistical model. Interestingly, Equation (92) is identical to the proper coaddition image presented in Zackay & Ofek (2015b; paper II in the series of coaddition).

Substituting Equation (92) into $\widehat{D}$, we get

$$\widehat{D} = \frac{\sqrt{\sum_j \frac{F_j^2|P_j|^2}{\sigma_j^2}}\widehat{N} - F_n\widehat{P_n}\left(\frac{\sum_j \frac{\overline{F_j\widehat{P_j}}}{\sigma_j^2}\widehat{R_j}}{\sqrt{\sum_j \frac{F_j^2|\widehat{P_j}|^2}{\sigma_j^2}}}\right)}{\sqrt{\sigma_n^2\left(\sum_j \frac{F_j^2|P_j|^2}{\sigma_j^2}\right) + \sigma_r^2 F_n^2|\widehat{P_n}|^2}}. \tag{93}$$

Writing the source detection statistic in explicit form we get

$$\widehat{S} = \frac{\frac{F_n\overline{\widehat{P_n}}}{\sigma_n^2}\left(\sum_j \frac{F_j^2|\widehat{P_j}|^2}{\sigma_j^2}\right)\widehat{N} - \frac{F_n^2|\widehat{P_n}|^2}{\sigma_n^2}\left(\sum_j \frac{F_j\overline{\widehat{P_j}}\widehat{R_j}}{\sigma_j^2}\right)}{\frac{F_n^2|\widehat{P_n}|^2}{\sigma_n^2} + \left(\sum_j \frac{F_j^2|\widehat{P_j}|^2}{\sigma_j^2}\right)}. \tag{94}$$

Thus, we arrive at an optimal solution with a closed formula for optimal transient detection given a set of references. We note that there are other choices that can be used instead of R. However, we prefer the proper coaddition image due to its uncorrelated noise (see Zackay & Ofek 2015b). Finally, N can also be composed of multiple images. In this case, the optimal solution for the subtraction is to perform the optimal transient detection with both N and R being the proper coaddition of all the images in their corresponding sets.

APPENDIX C
CORRECTION FOR SOURCE NOISE OF BRIGHT OBJECTS

The assumption that the noise distribution is independent of position, and of the true image itself, is of course not true. Specifically, for near bright stars the dominant source of noise is the Poisson fluctuations of the source itself, which is obviously position dependent. Therefore, in the vicinity of bright sources the variance is underestimated, and random fluctuations in the noise can cause false transient detections in these positions. Since only a negligible part of the sky behaves in such a way, we do not wish to change the statistic S in places away from bright sources.

Therefore, the approach we currently recommend is the following. Calculate separately the two parts of Equation (94):

$$\widehat{S_N} = \frac{\frac{F_n\overline{\widehat{P_n}}}{\sigma_n^2}\left(\sum_j \frac{F_j^2|\widehat{P_j}|^2}{\sigma_j^2}\right)}{\frac{F_n^2|\widehat{P_n}|^2}{\sigma_n^2} + \left(\sum_j \frac{F_j^2|\widehat{P_j}|^2}{\sigma_j^2}\right)}\widehat{N} \equiv \widehat{k_n}\widehat{N}, \tag{95}$$

and

$$\widehat{S_{R_j}} = \frac{\frac{F_n^2|\widehat{P_n}|^2}{\sigma_n^2}\frac{F_j\overline{\widehat{P_j}}}{\sigma_j^2}}{\frac{F_n^2|\widehat{P_n}|^2}{\sigma_n^2} + \left(\sum_j \frac{F_j^2|\widehat{P_j}|^2}{\sigma_j^2}\right)}\widehat{R_j} \equiv \widehat{k_j}\widehat{R_j}. \tag{96}$$

Next, apply inverse Fourier transform to get to the image domain:

$$S = S_N - \sum_j S_{R_j}. \tag{97}$$

Then calculate the corrected score for the existence of transient sources:

$$S_{\rm corr} = \frac{S_N - \sum_j S_{R_j}}{\sqrt{V(S_N) + \sum_j V(S_{R_j})}}, \tag{98}$$

where $V(S_N)$ and $V(S_{R_j})$ are the variance maps of S_N and S_{R_j}. Essentially, these can be computed analytically by following all the operations done on R_j and N, and applying the corresponding corrections to $V(S_{R_j})$ and $V(S_N)$ respectively.

Using the fact that for a zero expectancy noise source ϵ,

$$\mathrm{V}(\epsilon \otimes P) = V(\epsilon) \otimes (P^2). \tag{99}$$

we can derive a closed formula solution for $V(S_N)$ and $V(S_{R_j})$:

$$V(S_N) = V(\epsilon_n) \otimes (k_n^2), \tag{100}$$

$$V(S_{R_j}) = V(\epsilon_j) \otimes (k_j^2), \tag{101}$$

where k_n and k_j are defined in Equations (95) and (96), respectively. We note that the squaring of the convolution kernel happens in the image domain.

In the presence of bright stars, the noise is correlated, this means that we need to store, or sum up the individual $V(S_{R_j})$. Using the proper coaddition image and its effective k_r will not recover all the information. However, using R and k_r may serve as an approximation to this process.

The proposed correction (Equation (98)) does not change the score image away from bright sources (other than moving the detection statistic to units of standard deviations). The reason for this is that the variance map is spatially uniform in places away from strong sources. We note that this correction is suboptimal near bright sources, but at least it is a score with known statistical properties, that we can use to prevent false positives and to retain some sensitivity.

This method of correcting S by the variance can be extended to any additional sources of noise for which we can construct a model. For example, in Section 3.3, we present also the variance due to astrometric errors.

APPENDIX D
OPTIMAL PSF PHOTOMETRY OF TRANSIENT POINT SOURCES

In general, in the statistical community, there is no consensus on how to derive the best measurement. Therefore, in this section, we will search for a measurement statistic that is unbiased and has maximal S/N, and is a linear function of the input images. Not surprisingly, the resulting statistics is simply S (Equation (12)) normalized by some factor. This analysis also presents another formalism in which our transient detection statistic is optimal—it is the maximum S/N linear statistic composed out of R and N that cancels the constant in time image T. As a side, we note that the same solution arises when calculating the maximum likelihood estimator by maximizing Equation (68) with respect to α.

We start by stating again the statistical model we use:

$$R = P_r \otimes T + \epsilon_r, \tag{102}$$

$$N = P_n \otimes (T + \alpha\delta(q)) + \epsilon_n, \tag{103}$$

where α is the flux of the new source at position q, and $\delta(q)$ is an image with 0 everywhere except position q where it is value is 1. We continue to work under the assumption that the background noise is the most significant source of noise, which allows us to write

$$V[\epsilon_r] = \sigma_r^2, \quad V[\epsilon_n] = \sigma_n^2. \tag{104}$$

We write the statistic that we are looking for in its most general linear form:

$$C = k_n \otimes N + k_r \otimes R, \tag{105}$$

where k_n and k_r are some kernels, and we require that

$$F_n k_n \otimes P_n = -F_r k_r \otimes P_r. \tag{106}$$

Writing C in Fourier space we get

$$\widehat{C} = \widehat{k_n}\widehat{N} + \widehat{k_r}\widehat{R} = \alpha\widehat{\delta(q)}F_n\widehat{P_n}\widehat{k_n} + \widehat{\epsilon_c}, \tag{107}$$

where ϵ_c absorbs all noise sources in both images.

Here, we will use a well known result (also given in Appendix B of Zackay & Ofek 2015a) that the maximal S/N measurement of a parameter θ given a set of statistics X_j such that

$$X_j = \mu_j\theta + \epsilon_j, \tag{108}$$

where μ_j are scaling factors and ϵ_j has variance $V[\epsilon_j] = \sigma_j$, is

$$\tilde{\theta} = \frac{\sum_j \frac{\overline{\mu_j}}{\sigma_j^2} X_j}{\sum_j \frac{|\mu_j|^2}{\sigma_j^2}}. \tag{109}$$

In our case $\mu = \widehat{\delta(q)}F_n\widehat{P_n}\widehat{k_n}$. Applying this to $\widehat{C}$, we get the maximum S/N statistic for α:

$$\tilde{\alpha} = \frac{\sum_f \frac{\overline{\delta(q)F_n\widehat{P_n}\widehat{k_n}}\widehat{C}}{\sigma_r^2|\widehat{k_r}|^2 + \sigma_n^2|\widehat{k_n}|^2}}{\sum_f \frac{|\delta(q)F_n\widehat{P_n}\widehat{k_n}|^2}{\sigma_r^2|\widehat{k_r}|^2 + \sigma_n^2|\widehat{k_n}|^2}}, \tag{110}$$

substituting $\widehat{C} = \widehat{k_n}\widehat{N} + \widehat{k_r}\widehat{R}$, and

$$\widehat{k_r} = -\widehat{k_n}\frac{F_n\widehat{P_n}}{F_r\widehat{P_r}}, \tag{111}$$

and simplifying (notice the cancellation of k_n in the ratio, and the use of $|\widehat{\delta(q)}| = 1$) we get

$$\begin{aligned}\tilde{\alpha} &= \frac{\sum_f \frac{\overline{\delta(q)F_n\widehat{P_n}\widehat{k_n}}(\widehat{k_n}\widehat{N} + \widehat{k_r}\widehat{R})}{\sigma_r^2|\widehat{k_r}|^2 + \sigma_n^2|\widehat{k_n}|^2}}{\sum_f \frac{|\delta(q)F_n\widehat{P_n}\widehat{K_n}|^2}{\sigma_r^2|\widehat{k_r}|^2 + \sigma_n^2|\widehat{k_n}|^2}} \\ &= \frac{\sum_f \frac{\overline{\delta(q)F_n\widehat{P_n}}(\widehat{N} - \frac{F_n\widehat{P_n}}{F_r\widehat{P_r}}\widehat{R})}{\sigma_r^2\left|\frac{F_n\widehat{P_n}}{F_r\widehat{P_r}}\right|^2 + \sigma_n^2}}{\sum_f \frac{F_n^2|\widehat{P_n}|^2}{\sigma_r^2\left|\frac{F_n^2\widehat{P_n}}{F_r^2\widehat{P_r}}\right|^2 + \sigma_n^2}},\end{aligned} \tag{112}$$

$$\tilde{\alpha} = \frac{\sum_f \frac{\overline{\delta(q)}(F_r^2F_n|\widehat{P_r}|^2\overline{\widehat{P_n}}\widehat{N} - F_n^2F_r|\widehat{P_n}|^2\overline{\widehat{P_r}}\widehat{R})}{\sigma_r^2F_n^2|\widehat{P_n}|^2 + \sigma_n^2F_r^2|\widehat{P_r}|^2}}{\sum_f \frac{F_n^2F_r^2|\widehat{P_n}|^2|\widehat{P_r}|^2}{\sigma_r^2F_n^2|\widehat{P_n}|^2 + \sigma_n^2F_r^2|\widehat{P_r}|^2}}. \tag{113}$$

Last, we see that we can calculate all the fluxes for all the transient sources simultaneously by noticing that the numerator in the expression for $\tilde{\alpha}$ is the q'th position in the previously defined transient detection image S (Equation (79)). That is,

$$\tilde{\alpha} = \frac{S}{\sum_f \frac{F_n^2F_r^2|\widehat{P_n}|^2|\widehat{P_r}|^2}{\sigma_r^2F_n^2|\widehat{P_n}|^2 + \sigma_n^2F_r^2|\widehat{P_r}|^2}}. \tag{114}$$

This means that the same statistic can be computed both for detection and measurement. Therefore, in order to get a flux measurement from S, all we need is to normalize it by F_S—the denominator of Equation (114):

$$F_S = \sum_f \frac{F_n^2F_r^2|\widehat{P_n}|^2|\widehat{P_r}|^2}{\sigma_r^2F_n^2|\widehat{P_n}|^2 + \sigma_n^2F_r^2|\widehat{P_r}|^2}. \tag{115}$$

Via the same process as for the detection, the standard deviation of the flux measurement S at position q can be

estimated via inspection of S_N and S_R. We find that the standard deviation of F can be calculated by

$$\sigma_{\tilde{\alpha}} = \frac{\sqrt{V(S_N) + V(S_R)}}{F_S}. \quad (116)$$

If the reference image is constructed from many reference images, then

$$V(S_R) = \sum_j V(S_{R_j}). \quad (117)$$

Note that Equation (116) is valid for both faint (i.e., in background-dominated-noise areas) and bright transients (source-dominated-noise areas). We further note that Equation (114) is equivalent to PSF photometry as each pixel is weighted by the appropriate value of the PSF.

APPENDIX E
D, P_D, F_D ARE SUFFICIENT FOR ANY STATISTICAL MEASUREMENT OR DECISION ON THE DIFFERENCE BETWEEN THE IMAGES

In order to show that D, P_D, F_D are sufficient statistics, we will use the Fisher–Neyman factorization theorem. This theorem states that if the probability density function is $\mathcal{P}_\theta(X)$, then T is sufficient for the parameter θ if and only if nonnegative functions g and h can be found, such that

$$\mathcal{P}_\theta(X) = h(X) g_\theta(T(X)). \quad (118)$$

In our case, we would like to show that for any generative model $A_n(\theta)$ for the difference between the images, with parameter θ, the probability of observing the data (R and N) factorizes into

$$\mathcal{P}(R, N|A_n(\theta)) = \mathcal{P}(D|A_n(\theta)) g(R, N). \quad (119)$$

This will prove that D is a sufficient statistic.

We note that the meaning of sufficient statistics is profound —it means that any measurement or decision performed on D will return the same numerical value as if it was performed using all the data. Examples for such measurements or decisions are arbitrary shape measurements or identifying particle hits.

In this appendix, we show that D, along with P_D, P_{D_N}, P_{D_R}, are together sufficient for any likelihood calculation (up to some multiplicative, model independent factor, as allowed from the Fisher–Neyman criterion) for any instance of a generative model for $A_n(\theta)$, regardless of the constant-in-time image T. We state the family of statistical models $A_n(\theta, q)$ for which we want D to be sufficient to

$$R = F_r T \otimes P_r + \epsilon_r, \quad (120)$$

$$N = F_n T \otimes P_n + A_n(\theta) \otimes \delta(q) + \epsilon_n, \quad (121)$$

where $A_n(\theta)$ is the change made in the new image, located in position q, and T is the constant-in-time (unknown) image. Note that we did not convolved $A_n(\theta)$ with the PSF of the images, as this will allow us to deal with the signal that was not convolved by the PSF (e.g., bad pixels, small astrometric shifts). However, such a PSF can be included in $A_n(\theta)$.

Using the law of conditional probability, the probability we would like to calculate is

$$\mathcal{P}(R, N|A_n(\theta)) = \mathcal{P}(N|R, A_n(\theta)) \mathcal{P}(R|A_n(\theta)) \quad (122)$$

$$= \mathcal{P}(N|R, A_n(\theta)) \mathcal{P}(R). \quad (123)$$

Since the probability of R is independent of the model parameter θ (as it only influences the model for N), it suffices for us to calculate $\log(\mathcal{P}(N|R, A_n(\theta)))$. As we did in previous sections, we can project our knowledge of R to a statistical model for T:

$$\widehat{T} = \frac{\widehat{R}}{F_r \widehat{P_r}} - \frac{\widehat{\epsilon_r}}{F_r \widehat{P_r}} \equiv \frac{\widehat{R}}{F_r \widehat{P_r}} + \widehat{\epsilon_T}. \quad (124)$$

We can then use it to calculate the probability of observing N given $A_n(\theta)$:

$$-log(\mathcal{P}(N|R, A_n(\theta))) = \sum_f \frac{||\widehat{N} - F_n \widehat{P_n} \widehat{T} - \widehat{A_n(\theta)} \widehat{\delta(q)}||^2}{2V[\widehat{\epsilon_n} + F_n \widehat{P_n} \widehat{\epsilon_T}]}. \quad (125)$$

Opening the absolute value, we get the summation of three terms. The first term, $\sum_f \frac{||\widehat{N} - F_n \widehat{P_n} \widehat{T}||^2}{2V[\widehat{\epsilon_n} + F_n \widehat{P_n} \widehat{\epsilon_T}]}$, does not depend on $A_n(\theta)$ and therefore can be removed (can be absorbed in the Fisher–Neyman h). The second term is

$$\sum_f 2\Re \left[\frac{(\widehat{N} - F_n \widehat{P_n} \widehat{T}) \overline{\widehat{A_n(\theta)} \widehat{\delta(q)}}}{2V[\widehat{\epsilon_n} + F_n \widehat{P_n} \widehat{\epsilon_T}]} \right] = ... = \quad (126)$$

$$= \frac{(F_r^2 |\widehat{P_r}|^2 \widehat{N} - F_n F_r \widehat{P_n} \overline{\widehat{P_r}} \widehat{R}) \overline{\widehat{A_n(\theta)}}}{\sigma_n^2 F_r^2 |\widehat{P_r}|^2 + \sigma_r^2 F_n^2 |\widehat{P_n}|^2}. \quad (127)$$

In the last expression, we can identify a matched filter operation between the proper subtraction image D

$$\widehat{D} = \frac{(F_r \widehat{P_r} \widehat{N} - F_n \widehat{P_n} \widehat{R})}{\sqrt{\sigma_n^2 F_r^2 |\widehat{P_r}|^2 + \sigma_r^2 F_n^2 |\widehat{P_n}|^2}}, \quad (128)$$

and the PSF for delta function in N (the PSF of A_n)

$$\widehat{P_{D_N}} = \frac{F_r \widehat{P_r}}{F_{D_N} \sqrt{\sigma_n^2 F_r^2 |\widehat{P_r}|^2 + \sigma_r^2 F_n^2 |\widehat{P_n}|^2}}, \quad (129)$$

with thezero point

$$F_{D_N} = \frac{F_r}{\sqrt{\sigma_n^2 F_r^2 + \sigma_r^2 F_n^2}}. \quad (130)$$

Finally, we need to show that the third term in Equation (125) can be calculated only using D and its set of PSFs and zero points.

$$\sum_f \frac{|\widehat{A_n(\theta)}|^2}{V[\widehat{\epsilon_n} + F_n \widehat{P_n} \widehat{\epsilon_T}]} = \frac{F_r^2 |\widehat{P_r}|^2 |\widehat{A_n(\theta)}|^2}{\sigma_n^2 F_r^2 |\widehat{P_r}|^2 + \sigma_r^2 F_n^2 |\widehat{P_n}|^2} \quad (131)$$

$$= F_{D_N}^2 |\widehat{A_n(\theta)}|^2 |\widehat{P_{D_N}}|^2. \quad (132)$$

Symmetrically, every statistical change in R can be calculated in the same fashion using D and P_{D_R}.

$$\widehat{P_{D_R}} = \frac{F_n \widehat{P_n}}{F_{D_R} \sqrt{\sigma_n^2 F_r^2 |\widehat{P_r}|^2 + \sigma_r^2 F_n^2 |\widehat{P_n}|^2}}, \quad (133)$$

with thezero point

$$F_{D_R} = \frac{F_n}{\sqrt{\sigma_n^2 F_r^2 + \sigma_r^2 F_n^2}}. \quad (134)$$

As expected, a change in either N or R, that experiences the same PSF (and transparency) as the true image (e.g., a supernovae, variable star or small solar system body) will have the effective PSF P_D, and zero point F_D.

This analysis means that the subtraction product D, is the optimal statistic for any, even yet unspecified, measurement or hypothesis testing we wish to perform on the data.

REFERENCES

Alard, C. 2000, A&AS, 144, 363
Alard, C., & Lupton, R. H. 1998, ApJ, 503, 325
Arcavi, I., Gal-Yam, A., Yaron, O., et al. 2011, ApJL, 742, L18
Baltay, C., Rabinowitz, D., Hadjiyska, E., et al. 2013, PASP, 125, 683
Becker, A. C., Homrighausen, D., Connolly, A. J., et al. 2012, MNRAS, 425, 1341
Bellm, E. C., Kulkarni, S. R. & ZTF Collaboration 2015, in AAS Meeting 225 Abstracts, #328
Bloom, J. S., Richards, J. W., Nugent, P. E., et al. 2012, PASP, 124, 1175
Bramich, D. M. 2008, MNRAS, 386, L77
Bramich, D. M., Horne, K., Alsubai, K. A., et al. 2015, arXiv:1512.04655
Cenko, S. B., Kulkarni, S. R., Horesh, A., et al. 2013, ApJ, 769, 130
Cenko, S. B., Urban, A. L., Perley, D. A., et al. 2015, ApJL, 803, L24
Filippenko, A. V. 1982, PASP, 94, 715
Fisher, R. A. 1922, Philosophical Transactions of the Royal Society, 222, 309
Gal-Yam, A., Arcavi, I., Ofek, E. O., et al. 2014, Natur, 509, 471
Gal-Yam, A., Kasliwal, M. M., Arcavi, I., et al. 2011, ApJ, 736, 159
Gal-Yam, A., Maoz, D., Guhathakurta, P., & Filippenko, A. V. 2008, ApJ, 680, 550
Goldstein, D. A., D'Andrea, C. B., Fischer, J. A., et al. 2015, arXiv:1504.02936
Ivezic, Z., Tyson, J. A., Abel, B., et al. 2008, arXiv:0805.2366
Karlin, S., & Rubin, H. 1956, Ann. Math. Statist, 27, 272
Laher, R. R., Surace, J., Grillmair, C. J., et al. 2014, PASP, 126, 674
Law, N. M., Kulkarni, S. R., Dekany, R. G., et al. 2009, PASP, 121, 1395
Lindegren, L. 1980, AAP, 89, 41
Neyman, J. 1935, Giorn. Ist. Ital. Att., 6, 320
Neyman, J., & Pearson, E. S. 1933, Philosophical Transactions of the Royal Society, 231, 694
Ofek, E. O. 2014, MATLAB package for astronomy and astrophysics, Astrophysics Source Code Library, ascl:1407.005
Ofek, E. O., Frail, D. A., Breslauer, B., et al. 2011, ApJ, 740, 65
Ofek, E. O., Laher, R., Law, N., et al. 2012, PASP, 124, 62
Ofek, E. O., Sullivan, M., Shaviv, N. J., et al. 2014, ApJ, 789, 104
Phillips, A. C., & Davis, L. E. 1995, adass IV, 77, 297
Rau, A., Kulkarni, S. R., Law, N. M., et al. 2009, PASP, 121, 1334
Shao, M., & Colavita, M. M. 1992, A&A, 262, 353
Smith, A. M., Lynn, S., Sullivan, M., et al. 2011, MNRAS, 412, 1309
Stetson, P. B. 1987, PASP, 99, 191
van Dokkum, P. G. 2001, PASP, 113, 1420
Walter, C. W. 2015, Journal of Instrumentation, 10, C05015
Wright, D. E., Smartt, S. J., Smith, K. W., et al. 2015, MNRAS, 449, 451
Yuan, F., & Akerlof, C. W. 2008, ApJ, 677, 808
Zackay, B., & Ofek, E. O. 2015a, arXiv:1512.06872
Zackay, B., & Ofek, E. O. 2015b, arXiv:1512.06879