

Spatio-temporal Hotelling observer for signal detection from image sequences

Luca Caucci^{1,2}, Harrison H. Barrett^{1,2}, and Jeffrey J. Rodríguez³

¹ College of Optical Sciences, University of Arizona, Tucson, AZ 85721

² Center for Gamma-Ray Imaging, University of Arizona, Tucson, AZ 85724

³ Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721

caucci@email.arizona.edu

Abstract: Detection of signals in noisy images is necessary in many applications, including astronomy and medical imaging. The optimal linear observer for performing a detection task, called the Hotelling observer in the medical literature, can be regarded as a generalization of the familiar prewhitening matched filter. Performance on the detection task is limited by randomness in the image data, which stems from randomness in the object, randomness in the imaging system, and randomness in the detector outputs due to photon and readout noise, and the Hotelling observer accounts for all of these effects in an optimal way. If multiple temporal frames of images are acquired, the resulting data set is a spatio-temporal random process, and the Hotelling observer becomes a spatio-temporal linear operator. This paper discusses the theory of the spatio-temporal Hotelling observer and estimation of the required spatio-temporal covariance matrices. It also presents a parallel implementation of the observer on a cluster of Sony PLAYSTATION 3 gaming consoles. As an example, we consider the use of the spatio-temporal Hotelling observer for exoplanet detection.

© 2009 Optical Society of America

OCIS codes: (000.5490) Probability theory, stochastic processes, and statistics; (030.4280) Noise in imaging systems; (110.1080) Active or adaptive optics; (110.2970) Image detection systems; (110.3000) Image quality assessment; (110.4155) Multiframe image processing; (200.4960) Parallel processing; (330.1880) Detection.

References and links

1. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (John Wiley and Sons, Inc., New York, NY, 1966).
2. H. L. Van Trees, *Detection, Estimation, and Modulation Theory. Part III, Radar-sonar Signal Processing and Gaussian Signals in Noise* (John Wiley and Sons, Inc., New York, NY, 2001).
3. K. J. Myers, H. H. Barrett, M. C. Borgstrom, D. D. Patton, and G. W. Seeley, "Effect of Noise Correlation on Detectability of Disk Signals in Medical Imaging," *J. Opt. Soc. Am. A* **2**, 1752–1759 (1985).
4. S. J. Starr, C. E. Metz, L. B. Lusted, and D. J. Goodenough, "Visual detection and localization of radiographic images," *Radiology* **116**, 533–538 (1975).
5. D. W. Hughes, "Nonsolar Planets and Their Detection," *Nature* **279**, 579 (1979).
6. M. Tamura, "Extra-solar Planet Detection," *Viva Origino* **30**, 157–161 (2002).
7. J. O. Berger, *Statistical Decision Theory, Foundations, Concepts, and Methods* (Springer-Verlag, New York, NY, 1980).
8. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. (John Wiley and Sons, Inc., New York, NY, 2001).
9. J. L. Melsa and D. L. Cohn, *Decision and Estimation Theory* (McGraw-Hill, New York, NY, 1978).

10. R. N. McDonough and A. D. Whalen, *Detection of Signal in Noise* (Academic Press, San Diego, CA, 1995).
11. S. Park, E. Clarkson, M. A. Kupinski, and H. H. Barrett, "Efficiency of the human observer detecting random signals in random backgrounds," *J. Opt. Soc. Am. A* **22**, 3–16 (2005).
12. M. P. Hobson and C. McLachlan, "A Bayesian Approach to Discrete Object Detection in Astronomical Data Sets," *Mon. Not. R. Astron. Soc.* **338**, 765–784 (2003).
13. N. J. Kasdin and I. Braems, "Linear and Bayesian Planet Detection Algorithms for the *Terrestrial Planet Finder*," *Astrophys. J.* **646**, 1260–1274 (2006).
14. H. W. Babcock, "The Possibility of Compensating Astronomical Seeing," *Publ. Astron. Soc. Pac.* **65**, 229–236 (1953).
15. R. K. Tyson, *Adaptive Optics Engineering Handbook* (Marcel Dekker, New York, NY, 2000).
16. R. Akbarpour, S. N. Friedman, J. H. Siewerdsen, J. D. Neary, and I. A. Cunningham, "Signal and noise transfer in spatiotemporal quantum-based imaging systems," *J. Opt. Soc. Am. A* **24**, B151–B164 (2007).
17. G. K. Yadava, S. Rudinac, A. T. Kuhls-Gilcristab, and D. R. Bednarek, "Generalized Objective Performance Assessment of a New High-Sensitivity Microangiographic Fluoroscopic (HSMF) Imaging System," in *Medical Imaging 2008: Physics of Medical Imaging*, E. S. Jiang Hsieh, ed., vol. 6913, p. 69130U (Proc. SPIE, 2008).
18. H. H. Barrett, C. K. Abbey, and E. Clarkson, "Objective Assessment of Image Quality. III. ROC Metrics, Ideal Observers, and Likelihood-generating Functions," *J. Opt. Soc. Am. A* **15**, 1520–1535 (1998).
19. J. A. Hanley and B. J. McNeil, "The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve," *Radiology* **143**, 29–36 (1982).
20. S. H. Park, J. M. Goo, and C.-H. Jo, "Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists," *Korean J. Radiol.* **5**, 11–18 (2004).
21. H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley-Interscience, Hoboken, NJ, 2004).
22. H. Hotelling, "The Generalization of Student's Ratio," *Ann. Math. Stat.* **2**, 360–378 (1931).
23. H. H. Barrett, K. J. Myers, N. Devaney, and J. C. Dainty, "Objective Assessment of Image Quality: IV. Application to Adaptive Optics," *J. Opt. Soc. Am. A* **23**, 3080–3105 (2006).
24. E. Clarkson, "Estimation receiver operating characteristic curve and ideal observers for combined detection/estimation tasks," *J. Opt. Soc. Am. A* **24**, B91–B98 (2007).
25. L. Caucci, H. H. Barrett, N. Devaney, and J. J. Rodríguez, "Application of the Hotelling and ideal observers to detection and localization of exoplanets," *J. Opt. Soc. Am. A* **24**, B13–B24 (2007).
26. D. Burke, N. Devaney, S. Gladysz, H. H. Barrett, M. K. Whitaker, and L. Caucci, "Optimal linear estimation of binary star parameters," in *Adaptive Optics Systems*, N. Hubin, C. E. Max, and P. L. Wizinowich, eds., vol. 7015, p. 70152J (Proc. SPIE, 2008).
27. H. H. Barrett, K. J. Myers, B. D. Gallas, E. Clarkson, and H. Zhang, "Megalopinkophobia: Its symptoms and cures," in *Medical Imaging 2001: Physics of Medical Imaging*, L. E. Antonuk and M. J. Yaffe, eds., vol. 4320, pp. 299–307 (Proc. SPIE, 2001).
28. D. J. Tylavsky and G. R. L. Sohie, "Generalization of the Matrix Inversion Lemma," *Proc. IEEE* **74**, 1050–1052 (1986).
29. M. S. Bartlett, "An Inverse Matrix Adjustment Arising in Discriminant Analysis," *Ann. Math. Stat.* **22**, 107–111 (1951).
30. J. Sherman and W. J. Morrison, "Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix," *Ann. Math. Stat.* **21**, 124–127 (1950).
31. H. V. Henderson and S. R. Searle, "On Deriving the Inverse of a Sum of Matrices," *SIAM Rev.* **23**, 53–60 (1981).
32. J. A. Kahle, M. N. Day, H. P. Hofstee, C. R. Johns, T. R. Maeurer, and D. Shippy, "Introduction to the Cell Multiprocessor," *IMB J. Res. Dev.* **49**, 589–604 (2005).
33. D. Pham, T. Aipperspach, D. Boerstler, M. Bolliger, R. Chaudhry, D. Cox, P. Harvey, P. Harvey, H. Hofstee, C. Johns, J. Kahle, A. Kameyama, J. Keaty, Y. Masubuchi, M. Pham, J. Pille, S. Posluszny, M. Riley, D. Stasiak, O. Suzuoki, M. Takahashi, J. Warnock, S. Weitzel, D. Wendel, and K. Yazawa, "Overview of the architecture, circuit design, and physical implementation of a first-generation Cell Processor," *IEEE J. Solid-State Circuits* **41**, 179–196 (2006).
34. P. Hofstee, "Introduction to the Cell Broadband Engine," Tech. rep., IBM Corporation, Riverton, NJ (2005).
35. S. Williams, J. Shalf, L. Oliker, S. Kamil, P. Husbands, and K. Yelick, "The potential of the Cell Processor for scientific computing," in *Proceedings of the 3rd conference on Computing Frontiers*, pp. 9–20 (ACM, New York, NY, 2006).
36. D. A. Bader, V. Agarwal, K. Madduri, and S. Kang, "High performance combinatorial algorithm design on the Cell Broadband Engine processor," *Parallel Comput.* **33**, 720–740 (2007).
37. I. W. C. Benthin, M. Scherbaum, and H. Friedrich, "Ray tracing on the Cell Processor," in *IEEE Symposium on Interactive Ray Tracing*, pp. 15–23 (2006).
38. M. Sakamoto and M. Murase, "Parallel implementation for 3-D CT image reconstruction on Cell Broadband Engine™," in *IEEE International Conference on Multimedia and Expo*, pp. 276–279 (2007).
39. M. Kachelrieß, M. Knaup, and O. Bockenbach, "Hyperfast parallel-beam and cone-beam backprojection using the Cell general purpose hardware," *Med. Phys.* **34**, 1474–1486 (2007).
40. R. R. Parenti and R. J. Sasiela, "Laser-guide-star systems for astronomical applications," *J. Opt. Soc. Am. A* **11**,

- 288–309 (1994).
41. D. L. Fried, “Statistics of a Geometric Representation of Wavefront Distortion,” *J. Opt. Soc. Am* **55**, 1427–1435 (1965).
 42. M. P. Fitzgerald and J. R. Graham, “Speckle Statistics in Adaptively Corrected Images,” *Astrophys. J.* **637**, 541–547 (2006).
 43. G. L. Turin, “An introduction to matched filters,” *IRE Trans. Inf. Theory* **6**, 311–329 (1960).
 44. D. Middleton, *An Introduction to Statistical Communication Theory* (IEEE Press, Piscataway, NJ, 1960).
 45. B. D. Gallas, “One-shot estimate of MPMC Variance: AUC,” *Acad. Radiol.* **13**, 353–362 (2006).
 46. E. Bertin and S. Arnouts, “SEXTRACTOR: Software for source extraction,” *Astron. Astrophys. Suppl. Ser.* **117**, 393–404 (1996).
 47. L. Caucci, “Point Detection and Hotelling Discriminant: An Application in Adaptive Optics,” Master’s thesis, University of Arizona (2007).
 48. C. K. Abbey and H. H. Barrett, “Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability,” *J. Opt. Soc. Am. A* **18**, 473–488 (2001).
-

1. Introduction

Signal detection [1, 2] is a fundamental task in image science and a common problem in many fields, from medicine (for example, the detection of tumors [1, 3, 4]) to astronomy (detection of extrasolar planets [5, 6] or near-earth objects). As a consequence, an extensive theory and numerous algorithms have been developed to address the problem of signal detection from images [1, 2, 7–13]. Some systems, such as adaptive optics systems [14, 15], can deliver sequences of spatially correlated images. Spatio-temporally correlated data also appear in medical applications, such as [16, 17]. In many cases, the large amount of spatio-temporal data makes it hard to examine them and apply the detection algorithm directly to the data. Reduction in the temporal dimension is usually performed by adding together some or all of the frames. In adaptive optics, for example, images of the same object are taken over time and summed together either on the readout chip or in an external computer. The resulting single-frame image is used to perform the task of interest, but the information loss that results from the summation can reduce the detection performance. In medical applications, by contrast, it is usually not valid to assume that the object is constant over a long exposure time, and indeed the temporal dynamics of the object may be what defines the signal to be detected. Short-exposure images are inherently noisy, but long exposures wash out the signal. Full spatio-temporal processing is required for optimal signal detection.

The performance of a detection algorithm is mathematically quantified using receiver operating characteristic (ROC) analysis [18–20] and the area under the ROC curve (AUC) [19, 21]. With respect to that measure, the likelihood ratio is the optimal detector [21]. However, the likelihood ratio requires knowledge of the probability density functions under the hypotheses signal present and signal absent. Such probability density functions are often unknown or hard to estimate in practical cases. A more viable solution is the Hotelling observer [22], which requires only the knowledge of the data mean vector and covariance matrix. The Hotelling observer is linear, and it is optimal with respect to the class of linear observers [21] and a certain detectability measure to be defined below.

In this paper, the optimal-linear Hotelling observer is applied to spatio-temporal imagery. For this reason, we talk about the spatio-temporal Hotelling observer. By construction, such an observer is able to use both the spatial and temporal correlations between pixels in an optimal way, with respect to all linear observers. Methods for the estimation of the mean data vector and covariance matrix are described as well. Computational methods are described, and a parallel algorithm is implemented on a cluster of Sony PLAYSTATION 3 game consoles.

Table 1: Possible Outcomes of a Binary Discrimination Problem

Decision \ Signal	Present ($f \in \Gamma_1$)	Absent ($f \in \Gamma_0$)
“Yes” (H_1)	“hit” or “true positive”	“false alarm” or “false positive”
“No” (H_0)	“miss” or “false negative”	“correct rejection” or “true negative”

2. The spatio-temporal Hotelling observer

Let $\{\mathbf{g}^{(1)}, \dots, \mathbf{g}^{(J)}\}$ be a collection of J images of the same object f (assumed to be independent of time) taken over time. Each $\mathbf{g}^{(j)}$ is, in turn, a collection of M pixel intensities $\{g_1^{(j)}, \dots, g_M^{(j)}\}$. The data set $\{\mathbf{g}^{(1)}, \dots, \mathbf{g}^{(J)}\}$ represents the intensities of a total of MJ pixels, which we raster-scan and represent in compact form as the $MJ \times 1$ vector \mathbf{G} .

The task of interest is binary discrimination: given a noisy data set \mathbf{G} , an observer must decide whether the object f that produced \mathbf{G} belongs to either the “signal-absent” class Γ_0 or to the “signal-present” class Γ_1 . The observer can also be said to decide between hypothesis H_0 where the signal is absent, or hypothesis H_1 where the signal is present. In any case, the observer evaluates a real-valued non-random function t on the random data \mathbf{G} and compares $t(\mathbf{G})$ to a threshold τ . If $t(\mathbf{G}) > \tau$, hypothesis H_1 is assumed. Otherwise, hypothesis H_0 is concluded. All of the possible outcomes and associated terminology are summarized in Table 1.

Recall that \mathbf{G} is a random vector, so $t(\mathbf{G})$ is a random variable. For any fixed value of τ , the decision taken depends on $t(\mathbf{G})$, so it is a random variable as well. We can consider its probability density functions $\text{pr}(t|H_0)$ and $\text{pr}(t|H_1)$ given, respectively, hypothesis H_0 or H_1 . These two densities allow us to formally define the true positive fraction (TPF) and the false positive fraction (FPF) as

$$\text{TPF}(\tau) = \int_{\tau}^{\infty} \text{pr}(t|H_1) dt, \quad \text{FPF}(\tau) = \int_{\tau}^{\infty} \text{pr}(t|H_0) dt, \quad (1)$$

in which it is made clear that TPF and FPF are functions of τ .

If we change the value of τ , different values of $\text{TPF}(\tau)$ and $\text{FPF}(\tau)$ are obtained; a plot of $\text{TPF}(\tau)$ versus $\text{FPF}(\tau)$ as τ is varied over the real line is called a receiver operating characteristic (ROC) curve [18–20], and the area under the ROC curve (AUC) [19, 21] is a meaningful figure of merit for a binary classification task. The AUC is defined as [21]

$$\text{AUC}_{t(\mathbf{G})} = - \int_{-\infty}^{\infty} \text{TPF}(\tau) \frac{d\text{FPF}(\tau)}{d\tau} d\tau.$$

Another figure of merit for the same task is the signal-to-noise ratio (SNR) on the test statistic $t(\mathbf{G})$:

$$\text{SNR}_{t(\mathbf{G})} = \frac{\langle t(\mathbf{G}) \rangle_{\mathbf{G}|H_1} - \langle t(\mathbf{G}) \rangle_{\mathbf{G}|H_0}}{\sqrt{\frac{1}{2} \text{Var}\{t(\mathbf{G})|H_1\} + \frac{1}{2} \text{Var}\{t(\mathbf{G})|H_0\}}}, \quad (2)$$

in which the notation $\langle t(\mathbf{G}) \rangle_{\mathbf{G}|H_i}$ denotes the statistical expectation of the random variable $t(\mathbf{G})$ conditioned to the knowledge that hypothesis H_i is true. Similarly, $\text{Var}\{t(\mathbf{G})|H_i\}$ is the variance of $t(\mathbf{G})$ under the hypothesis H_i . In (2), hypotheses H_0 and H_1 are assumed equiprobable. The $\text{AUC}_{t(\mathbf{G})}$ is a known, monotonic function of $\text{SNR}_{t(\mathbf{G})}$ if $t(\mathbf{G})$ is normally distributed [21].

Many classical monographs discuss statistical decision theory; some of them are [2] and [9]. These texts show that, for binary classifications and for any ROC-related figure of merit (in-

cluding the AUC), the optimal observer is the likelihood ratio [1, 2, 9, 21]:

$$\Lambda(\mathbf{G}) = \frac{\text{pr}(\mathbf{G}|H_1)}{\text{pr}(\mathbf{G}|H_0)},$$

or, equivalently, its logarithm $\lambda(\mathbf{G}) = \ln \Lambda(\mathbf{G})$. However, we note that $\Lambda(\mathbf{G})$ requires knowledge of the multivariate densities $\text{pr}(\mathbf{G}|H_i)$, which are usually unknown or difficult to estimate. A more viable alternative can be found by restricting attention to linear observers, i.e., observers of the form $t(\mathbf{G}) = \mathbf{W}^T \mathbf{G}$, for an appropriate template vector \mathbf{W} of the same size of \mathbf{G} . Here, the symbol T denotes the transpose of a vector or matrix, so $\mathbf{W}^T \mathbf{G}$ is a scalar product. The optimal template vector can be derived by substituting $t(\mathbf{G}) = \mathbf{W}^T \mathbf{G}$ in (2) and maximizing $\text{SNR}_{t(\mathbf{G})}^2$ with respect to \mathbf{W} . The resulting template vector, which we call the Hotelling template vector [21, 22], is of the form

$$\mathbf{W}_{\text{Hot}} = [\mathbf{K}_{\mathbf{G}}]^{-1} \mathbf{S}, \quad (3)$$

in which $\mathbf{K}_{\mathbf{G}}$ is the covariance matrix of the data vector \mathbf{G} and $\mathbf{S} = \overline{\overline{\overline{\mathbf{G}}}}_1 - \overline{\overline{\overline{\mathbf{G}}}}_0$ is the image of the signal. The triple overbar represents an average over object randomness, system randomness, and measurement noise as discussed below. The linear observer that uses \mathbf{W}_{Hot} is the Hotelling observer, defined as $t_{\text{Hot}}(\mathbf{G}) = \mathbf{W}_{\text{Hot}}^T \mathbf{G}$. The Hotelling observer is also called a prewhitening matched filter [21], and the prewhitening operation is both correcting for the correlation in a single image and also undoing the frame-to-frame correlation. Note that the Hotelling observer $t_{\text{Hot}}(\mathbf{G})$ defined above is applied to spatio-temporal data, as opposed to the classical use of the Hotelling observer in the case of purely spatial data.

3. Analysis of the data covariance matrix

In a general case, we have three sources of randomness: detector noise, point spread function variability, and object variability. This leads to the decomposition [23]

$$\mathbf{K}_{\mathbf{G}} = \overline{\overline{\overline{\mathbf{K}}}}_{\mathbf{G}}^{\text{noise}} + \overline{\overline{\overline{\mathbf{K}}}}_{\mathbf{G}}^{\text{PSF}} + \overline{\overline{\overline{\mathbf{K}}}}_{\mathbf{G}}^{\text{object}}, \quad (4)$$

with which the Hotelling template vector is written as

$$\mathbf{W}_{\text{Hot}} = [\overline{\overline{\overline{\mathbf{K}}}}_{\mathbf{G}}^{\text{noise}} + \overline{\overline{\overline{\mathbf{K}}}}_{\mathbf{G}}^{\text{PSF}} + \overline{\overline{\overline{\mathbf{K}}}}_{\mathbf{G}}^{\text{object}}]^{-1} \mathbf{S}. \quad (5)$$

The expression in (4) can be formally derived from the definition of covariance matrix for \mathbf{G} :

$$\mathbf{K}_{\mathbf{G}} = \left\langle \left\langle \left\langle [\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}] [\mathbf{G} - \overline{\overline{\overline{\mathbf{G}}}}]^T \right\rangle_{\mathbf{G}|\mathbf{P},f} \right\rangle_{\mathbf{P}|f} \right\rangle_f, \quad (6)$$

in which \mathbf{P} represents the sequence $\{\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(J)}\}$ of point spread functions (PSFs). We will allow the PSFs to be random, as in the adaptive optics problem. The PSFs will be assumed known statistically (PSF-known-statistically or PKS [21]), and their contribution $\overline{\overline{\overline{\mathbf{K}}}}_{\mathbf{G}}^{\text{PSF}}$ to the data covariance matrix will be estimated by means of simulated data. It is important to note that full knowledge of the statistical properties of the PSFs is not needed. Instead, the Hotelling observer requires the knowledge of only the mean signal \mathbf{S} and the data covariance matrix $\mathbf{K}_{\mathbf{G}}$. As long as these two quantities can be estimated with sufficient accuracy, the Hotelling observer will deliver high performance. No moments higher than the second are needed.

Expression (6) contains three averaging (or statistical expectation) steps. The innermost expectation is on the noise and for given \mathbf{P} and f . The resulting quantity is averaged over \mathbf{P} given

f , and, finally, we average over f . From (6) and adding and subtracting terms appropriately [23], (4) is derived, provided that:

$$\begin{aligned}\bar{\mathbf{K}}_{\mathbf{G}}^{\text{noise}} &= \langle \langle \mathbf{K}_{\mathbf{G}}^{\text{noise}} \rangle_{\mathbf{P}|f} \rangle_f = \left\langle \left\langle \left\langle [\mathbf{G} - \bar{\mathbf{G}}] [\mathbf{G} - \bar{\mathbf{G}}]^T \right\rangle_{\mathbf{G}|\mathbf{P},f} \right\rangle_{\mathbf{P}|f} \right\rangle_f, \\ \bar{\mathbf{K}}_{\bar{\mathbf{G}}}^{\text{PSF}} &= \langle \mathbf{K}_{\bar{\mathbf{G}}}^{\text{PSF}} \rangle_f = \left\langle \left\langle \left\langle [\bar{\mathbf{G}} - \bar{\bar{\mathbf{G}}}] [\bar{\mathbf{G}} - \bar{\bar{\mathbf{G}}}]^T \right\rangle_{\mathbf{P}|f} \right\rangle_{\bar{\mathbf{G}}|f} \right\rangle_f, \\ \mathbf{K}_{\bar{\bar{\mathbf{G}}}}^{\text{object}} &= \left\langle [\bar{\bar{\mathbf{G}}} - \bar{\mathbf{G}}] [\bar{\bar{\mathbf{G}}} - \bar{\mathbf{G}}]^T \right\rangle_f,\end{aligned}$$

in which \mathbf{G} with a variable number of bars denotes average with respect to noise, PSFs, and object. By construction, random vectors \mathbf{G} , $\bar{\mathbf{G}}$, and $\bar{\bar{\mathbf{G}}}$ are uncorrelated [23].

For simplicity, we will assume that the signal we want to detect is known in brightness and location. In this case, we refer to a signal-known-exactly (SKE) problem [21]. The more realistic problem of unknown signal location can be handled by scanning the observer template [24–26]. The SKE hypothesis provides valuable information that the observer can use to improve detection performance (quantified by an increase in AUC) with respect to the detection and localization problem of [25]. We also assume that the object background is known exactly (background-known-exactly or BKE in the terminology of [21]).

The noise covariance matrix is usually very easy to study. Indeed, if we assume that only photon (Poisson) and readout (Gaussian) noises are present in the sequence \mathbf{G} , then the noise in distinct elements of the detector are usually uncorrelated and so we can write

$$[\bar{\mathbf{K}}_{\mathbf{G}}^{\text{noise}}]_{m,m'}^{(j,j')} = \begin{cases} \sigma_m^2 + \bar{\bar{g}}_m^{(j)} & \text{if } m = m' \text{ and } j = j', \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where σ_m^2 is the readout noise variance for the m -th pixel of the detector, and

$$\bar{\bar{g}}_m^{(j)} = \Pr(H_0) \bar{g}_{m|H_0}^{(j)} + \Pr(H_1) \bar{g}_{m|H_1}^{(j)}$$

is the variance of the photon noise for the same pixel when $\bar{\bar{g}}_m^{(j)}$ is the expected number of photons (from objects in the field of view and background) collected for the m -th pixel of the j -th image. The appropriateness of the Poisson model for the photon noise is justified by invoking the so-called Poisson postulates [21]. The readout noise variance σ_m^2 at each detector pixel is usually known, as provided by the detector's manufacturer, or it can be measured. Matrix $\bar{\mathbf{K}}_{\mathbf{G}}^{\text{noise}}$ above is diagonal with no zero terms on the diagonal, which guarantees the invertibility of $\mathbf{K}_{\mathbf{G}}$.

4. Estimation of the Hotelling template vector

In this section, we describe how quantities discussed in the previous section can be estimated. We will rely on simulation code and consider L_1 realization of the PSF sequence \mathbf{P} and L_2 realization of a random signal. Consider $L_1 L_2$ simulated noiseless data sets $\bar{\mathbf{G}}_1^{(\ell_1, \ell_2)}$, $\ell_1 = 1, \dots, L_1$, $\ell_2 = 1, \dots, L_2$ for the hypothesis H_1 and, similarly, L_1 noiseless data sets $\bar{\mathbf{G}}_0^{(\ell_1)}$, for $\ell_1 = 1, \dots, L_1$ for the hypothesis H_0 .

The noise covariance matrix is estimated as:

$$[\hat{\bar{\mathbf{K}}}_{\mathbf{G}}^{\text{noise}}]_{m,m'}^{(j,j')} = \begin{cases} \sigma_m^2 + \hat{\bar{\bar{g}}}_m^{(j)} & \text{if } m = m' \text{ and } j = j', \\ 0 & \text{otherwise.} \end{cases}$$

where we denoted estimated quantities using the hat symbol and we set

$$\hat{\bar{\mathbf{g}}}_m^{(j)} = \left[\hat{\bar{\mathbf{G}}} \right]_m^{(j)} = \frac{1}{L_1} \sum_{\ell_1=1}^{L_1} \left[\bar{\mathbf{G}}_0^{(\ell_1)} \right]_m^{(j)}.$$

The PSF covariance matrix $\bar{\mathbf{K}}_{\mathbf{G}}^{\text{PSF}}$ is estimated from simulated data as well:

$$\left[\hat{\bar{\mathbf{K}}}_{\mathbf{G}}^{\text{PSF}} \right]_{m,m'}^{(j,j')} = \frac{1}{L_1-1} \sum_{\ell_1=1}^{L_1} \left[\Delta \bar{\mathbf{G}}_0^{(\ell_1)} \right]_m^{(j)} \left[\Delta \bar{\mathbf{G}}_0^{(\ell_1)} \right]_{m'}^{(j')}, \quad (8)$$

where

$$\left[\Delta \bar{\mathbf{G}}_0^{(\ell_1)} \right]_m^{(j)} = \left[\bar{\mathbf{G}}_0^{(\ell_1)} \right]_m^{(j)} - \frac{1}{L_1} \sum_{\ell'_1=1}^{L_1} \left[\bar{\mathbf{G}}_0^{(\ell'_1)} \right]_m^{(j)}. \quad (9)$$

Expressions (8) and (9) show that $\hat{\bar{\mathbf{K}}}_{\mathbf{G}}^{\text{PSF}}$ is the sample estimate of $\bar{\mathbf{K}}_{\mathbf{G}}^{\text{PSF}}$, estimated from the noiseless simulated data.

Finally, if we consider randomness in the signal to be detected or the background on which it is superimposed, the object term in the covariance matrix expression can be estimated as:

$$\left[\hat{\bar{\mathbf{K}}}_{\mathbf{G}}^{\text{object}} \right]_{m,m'}^{(j,j')} = \frac{1}{L_2-1} \sum_{\ell_2=1}^{L_2} \left[\Delta \bar{\mathbf{G}}_1^{(\ell_2)} \right]_m^{(j)} \left[\Delta \bar{\mathbf{G}}_1^{(\ell_2)} \right]_{m'}^{(j')},$$

where

$$\left[\Delta \bar{\mathbf{G}}_1^{(\ell_2)} \right]_m^{(j)} = \frac{1}{L_1} \sum_{\ell_1=1}^{L_1} \left[\Delta \bar{\mathbf{G}}_1^{(\ell_1, \ell_2)} \right]_m^{(j)}.$$

It might be tempting to try to estimate the whole data covariance matrix $\mathbf{K}_{\mathbf{G}}$ in (6) from noisy simulated data, without using the decomposition (4). A necessary (but not sufficient) condition for such estimate $\hat{\mathbf{K}}_{\mathbf{G}}$ to be nonsingular is that the number $L = L_1 L_2$ of simulated noisy sequences must be greater than MJ , the order of $\mathbf{K}_{\mathbf{G}}$ itself. If, for example, each image $\mathbf{g}^{(j)}$ is of size 64×64 and there are 25 of them in each sequence \mathbf{G} , then $MJ = 64^2 \cdot 25 \approx 10^5$. Simulating such a huge number of image sequences is clearly prohibitive. Instead, the decomposition (4), along with (7), guarantees the invertibility of $\hat{\mathbf{K}}_{\mathbf{G}}$. This can be proved by noting that $\hat{\mathbf{K}}_{\mathbf{G}}$ is symmetric and (strictly) positive definite. Indeed:

$$\mathbf{x}^T \hat{\mathbf{K}}_{\mathbf{G}} \mathbf{x} = \mathbf{x}^T \hat{\bar{\mathbf{K}}}_{\mathbf{G}}^{\text{noise}} \mathbf{x} + \mathbf{x}^T \left[\hat{\bar{\mathbf{K}}}_{\mathbf{G}}^{\text{PSF}} + \hat{\bar{\mathbf{K}}}_{\mathbf{G}}^{\text{object}} \right] \mathbf{x} \geq \mathbf{x}^T \hat{\bar{\mathbf{K}}}_{\mathbf{G}}^{\text{noise}} \mathbf{x} = \sum_{m=0}^M \left(\sigma_m^2 + \hat{\bar{\mathbf{g}}}_m^{(j)} \right) x_m^2 > 0,$$

for any vector $\mathbf{x} \neq \mathbf{0}$.

Finally, the average contribution of the signal to the image data is estimated as:

$$\hat{\mathbf{S}} = \frac{1}{L_1 L_2} \sum_{\ell_1=1}^{L_1} \sum_{\ell_2=1}^{L_2} \bar{\mathbf{G}}_1^{(\ell_1, \ell_2)} - \frac{1}{L_1} \sum_{\ell_1=1}^{L_1} \bar{\mathbf{G}}_0^{(\ell_1)}.$$

Armed with an estimate of the signal to be detected and estimates of the covariance matrices that appear on the right-hand side of (4), we can formally write an expression for the Hotelling template vector estimate: in analogy with (5), we define

$$\hat{\mathbf{W}}_{\text{Hot}} = \left[\hat{\mathbf{K}}_{\mathbf{G}} \right]^{-1} \hat{\mathbf{S}} = \left[\hat{\bar{\mathbf{K}}}_{\mathbf{G}}^{\text{noise}} + \hat{\bar{\mathbf{K}}}_{\mathbf{G}}^{\text{PSF}} + \hat{\bar{\mathbf{K}}}_{\mathbf{G}}^{\text{object}} \right]^{-1} \hat{\mathbf{S}}. \quad (10)$$

For the SKE/BKE case, (10) reduces to:

$$\hat{\mathbf{W}}_{\text{Hot}} = [\hat{\mathbf{K}}_{\mathbf{G}}]^{-1} \hat{\mathbf{S}} = [\hat{\mathbf{K}}_{\mathbf{G}}^{\text{noise}} + \hat{\mathbf{K}}_{\mathbf{G}}^{\text{PSF}}]^{-1} \hat{\mathbf{S}}. \quad (11)$$

Although (10) and (11) make sense because $[\hat{\mathbf{K}}_{\mathbf{G}}]^{-1}$ exists, the matrix we need to invert is huge, so computing the inverse by means of standard algorithms (such as Gaussian elimination) is computationally prohibitive [27]. Even if we could invert $\hat{\mathbf{K}}_{\mathbf{G}}$, storing it will require an incredible amount of disk space. However, we recognize the particular structure of the PSF covariance matrix [see (8)] and consider an algorithm that takes advantage of it. Indeed, if we introduce the matrix \mathbf{R} whose elements are

$$[\mathbf{R}]_{\ell,m}^{(j)} = \frac{1}{\sqrt{L-1}} [\Delta \bar{\mathbf{G}}_0^{(\ell)}]_m^{(j)},$$

then (8) is rewritten as

$$\hat{\mathbf{K}}_{\mathbf{G}}^{\text{PSF}} = \mathbf{R} \mathbf{R}^T.$$

The $MJ \times L$ matrix \mathbf{R} contains in the ℓ -th column the $MJ \times 1$ vector obtained by raster-scanning the pixels in $\Delta \bar{\mathbf{G}}_0^{(\ell)}$. The Woodbury matrix-inversion lemma [28–31] allows us to rewrite the inverse in (10) as a computationally tractable expression. In abstract form, the matrix-inversion lemma can be stated as follows:

$$[\mathbf{A} - \mathbf{UBV}]^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1} [\mathbf{I}_N - \mathbf{BVA}^{-1} \mathbf{U}]^{-1} \mathbf{BVA}^{-1},$$

in which \mathbf{I}_N is the identity matrix of order N . If

$$\mathbf{A} = \hat{\mathbf{K}}_{\mathbf{G}}^{\text{noise}}, \quad \mathbf{B} = -\mathbf{I}_L, \quad \mathbf{U} = \mathbf{R}, \quad \mathbf{V} = \mathbf{R}^T,$$

then

$$[\hat{\mathbf{K}}_{\mathbf{G}}^{\text{noise}} + \hat{\mathbf{K}}_{\mathbf{G}}^{\text{PSF}}]^{-1} = [\hat{\mathbf{K}}_{\mathbf{G}}^{\text{noise}}]^{-1} \left\{ \mathbf{I}_{MJ} - \mathbf{RQ}^{-1} \mathbf{R}^T [\hat{\mathbf{K}}_{\mathbf{G}}^{\text{noise}}]^{-1} \right\}, \quad (12)$$

where

$$\mathbf{Q} = \mathbf{I}_L + \mathbf{R}^T [\hat{\mathbf{K}}_{\mathbf{G}}^{\text{noise}}]^{-1} \mathbf{R}.$$

The matrix $\hat{\mathbf{K}}_{\mathbf{G}}^{\text{noise}}$ is diagonal, so computing its inverse does not pose computational problems. The matrix \mathbf{Q} is of size $L \times L$ and invertible. Note that L is usually much smaller than MJ , which implies that \mathbf{Q}^{-1} can be calculated in much shorter time than $[\hat{\mathbf{K}}_{\mathbf{G}}]^{-1}$. Standard Gaussian elimination with pivoting is a fast and numerically stable way to invert \mathbf{Q} . Overall, using (12) is a more tractable and stable problem than computing (11) directly.

5. Implementation

For this research, we took advantage of the computational capabilities available at the Center for Gamma-Ray Imaging, University of Arizona. In particular, we used a Sony PLAYSTATION 3 cluster consisting of 30 units. Each unit was equipped with a Cell Broadband Engine (Cell BE) processor, 256 MB of RAM memory, 60 GB of disk space, and was running Linux Fedora 7, kernel version 2.6.23. The IBM's Cell BE Software Development Kit 3.0 was installed on all of the units and used to generate code suitable for the Cell BE microprocessor. All machines in the cluster were connected by means of a 1-Gbit/s local area network (LAN). All algorithms were coded using the C programming language, and source files were compiled using IBM XL C/C++ compiler, version 9.0. Communication between different nodes of the computer cluster was achieved using the Message Passing Interface (MPI) standard. The Cell BE

architecture [32–34] has recently received enormous attention from the computer science community. The possibility of using up to nine processing cores and the fact that SIMD instructions are supported make the Cell BE processor particularly appealing as a low-cost solution for high-performance scientific computing [35–39].

An algorithm for the computation of the template vector $\hat{\mathbf{W}}_{\text{Hot}}$ according to (10) and (12) was implemented and run on the PLAYSTATION 3 cluster. The algorithm is composed of two different programs: one to be run as a master process and one to be run as a slave process. Pseudocode for the master process is shown below:

```

 $J \leftarrow$  number of images in each sequence
 $L \leftarrow$  number of simulated sequences
 $N \leftarrow$  number of slave processes
for all  $n \in \{1, \dots, N\}$  do
    send “initialize” to slave process  $n$ 
end for
for all  $\ell \in \{1, \dots, L\}$  do
    read  $\bar{\mathbf{G}}_0^{(\ell)}$  from the disk
    for all  $n \in \{1, \dots, N\}$  do
        send  $\bar{\mathbf{G}}_0^{(\ell)}$  to slave process  $n$ 
    end for
end for
for all  $\ell \in \{1, \dots, L\}$  do
    read  $\bar{\mathbf{G}}_1^{(\ell)}$  from the disk
    for all  $n \in \{1, \dots, N\}$  do
        send  $\bar{\mathbf{G}}_1^{(\ell)}$  to slave process  $n$ 
    end for
end for
 $j \leftarrow 1$ 
 $m \leftarrow 0$  {number of tasks completed}
while  $m < J$  do
    while  $m < J$  and there is an idle slave process do
         $n \leftarrow$  index of an idle slave process
        send “compute  $[\hat{\mathbf{W}}_{\text{Hot}}]^{(j)}$ ” to slave process  $n$ 
         $j \leftarrow j + 1$ 
    end while
    if  $m < J$  then
         $n \leftarrow$  slave process that has just computed  $[\hat{\mathbf{W}}_{\text{Hot}}]^{(j')}$ 
        receive  $[\hat{\mathbf{W}}_{\text{Hot}}]^{(j')}$  from slave process  $n$ 
        save  $[\hat{\mathbf{W}}_{\text{Hot}}]^{(j')}$  to disk
         $m \leftarrow m + 1$ 
    end if
end while
for all  $n \in \{1, \dots, N\}$  do
    send “end of computation” to slave process  $n$ 
end for

```

Pseudocode for the slave processes is shown below:

```

 $J \leftarrow$  number of images in each sequence
 $L \leftarrow$  number of simulated sequences
 $\sigma_m^2 \leftarrow$  readout noise variance for all  $m \in \{1, \dots, M\}$ 
for all  $\ell \in \{1, \dots, L\}$  do
    receive  $\bar{\mathbf{G}}_0^{(\ell)}$  from master process
    save  $\bar{\mathbf{G}}_0^{(\ell)}$  to disk
end for
for all  $\ell \in \{1, \dots, L\}$  do
    receive  $\bar{\mathbf{G}}_1^{(\ell)}$  from master process
    save  $\bar{\mathbf{G}}_1^{(\ell)}$  to disk
end for
 $\hat{\mathbf{G}}_0 \leftarrow \frac{1}{L} \sum_{\ell=1}^L \bar{\mathbf{G}}_0^{(\ell)}$ 

```

```

 $\hat{\hat{\mathbf{G}}}_1 \leftarrow \frac{1}{L} \sum_{\ell=1}^L \bar{\mathbf{G}}_1^{(\ell)}$ 
 $\hat{\mathbf{S}} \leftarrow \hat{\hat{\mathbf{G}}}_1 - \hat{\mathbf{G}}_0$ 
for all  $j \in \{1, \dots, J\}$  do
     $[\hat{\hat{\mathbf{K}}}_{\mathbf{G}}^{\text{noise}}]^{(j)} \leftarrow \text{diag} \left( \sigma_1^2 + \hat{\hat{g}}_{0,1}^{(j)}, \dots, \sigma_M^2 + \hat{\hat{g}}_{0,M}^{(j)} \right)$ 
end for
 $[\mathbf{R}]_{\ell}^{(j)} \leftarrow \frac{1}{\sqrt{L-1}} [\Delta \bar{\mathbf{G}}_0^{(\ell)}]^{(j)}$ 
 $\mathbf{Q} \leftarrow \mathbf{I}_L + \mathbf{R}^T [\hat{\hat{\mathbf{K}}}_{\mathbf{G}}^{\text{noise}}]^{-1} \mathbf{R}$ 
 $\mathbf{Q}^{-1} \leftarrow \text{inverse}(\mathbf{Q})$ 
while not message “end of computation” received do
    receive “compute  $[\hat{\mathbf{W}}_{\text{Hot}}]^{(j)}$ ” from master process
     $[\hat{\mathbf{W}}_{\text{Hot}}]^{(j)} \leftarrow \mathbf{0}_{M \times M}$ 
    for all  $j' \in \{1, \dots, J\}$  do
         $\mathbf{T}^{(j,j')} \leftarrow (j, j')\text{-th block of } [\hat{\hat{\mathbf{K}}}_{\mathbf{G}}^{\text{noise}}]^{-1} \left\{ \mathbf{I}_{MJ} - \mathbf{RQ}^{-1} \mathbf{R}^T [\hat{\hat{\mathbf{K}}}_{\mathbf{G}}^{\text{noise}}]^{-1} \right\}$ 
         $[\hat{\mathbf{W}}_{\text{Hot}}]^{(j)} \leftarrow [\hat{\mathbf{W}}_{\text{Hot}}]^{(j)} + \mathbf{T}^{(j,j')} [\hat{\mathbf{S}}]^{(j')}$ 
    end for
    send  $[\hat{\mathbf{W}}_{\text{Hot}}]^{(j)}$  to master process
end while

```

In the implementation, we took advantage of the Cell BE processor and its SIMD capabilities. Floating-point values were stored in double precision, and SIMD instructions were used for the computation of the blocks of $[\hat{\mathbf{K}}_{\mathbf{G}}]^{-1}$ as they were needed. Splitting the work load among all of the processing cores available on each PLAYSTATION 3 allowed more than a 15-fold reduction in the computation time with respect to an implementation that uses only one core and ignores their SIMD capabilities. Had we used single-precision values, the speed-up would have been even larger.

6. Simulation results

As an example, we developed the spatio-temporal Hotelling observer for adaptive optics (AO) images [14, 15]. The simulation code we used is AOTools (webpage: <http://www.tosc.com/software/software.html>). The atmospheric thickness was set to 24km and, in order to compute the aberrated wavefront, we assumed the frozen flow hypothesis of atmospheric turbulence. The turbulence was computed according to the modified Hufnagel-Valley C_n^2 profile, given by [40]:

$$C_n^2(h) = 8.16 \cdot 10^{-54} h^{10} e^{-h/1000} + 3.02 \cdot 10^{-17} e^{-h/1500} + 1.90 \cdot 10^{-15} e^{-h/100},$$

in which the altitude h is expressed in meters and the units of C_n^2 are meters to the $-2/3$ power. We assumed $C_n^2(h)$ trascurable when $h \geq 24$ km. The Fried parameter r_0 [41] for the phase screens was 0.30m at the wavelength $\lambda = 500$ nm. Our simulation took into account effects such as scintillation and anisoplanetism [15] as well. The wind speed was 1.25 m/s. The telescope we simulated had a circular aperture of diameter 5 m, and the diameter of the central obscuration due to the secondary mirror was 0.50m. The secondary mirror was supported by three arms. For the estimation of the template vector $\hat{\mathbf{W}}_{\text{Hot}}$ in (10), we simulated $L = 512$ sequences containing $J = 25$ images each of size 64×64 (pixel size $5.96 \mu\text{m}$). The wavefront sensor apparatus of many AO systems includes a lenslet array. In our simulation, we simulated a 32×32 array of side length 0.02m. The total power entering the telescope was equally split between the wavefront sensor and the science camera by a 50/50 beamsplitter. The system was assumed idea, with an efficiency of 1 (i.e., no losses) and the AO loop was running at a speed of 1 kHz. The quality of the AO correction can be quantified with an average Strehl ratio of about 0.67.

Table 2: Values of the AUC, σ , and the SNR for the Observers $t_{\text{Hot}}(\mathbf{G})$, $t_{\text{Hot}}(\mathbf{g})$, and $t_{\text{mat}}(\mathbf{g})$

Observer	AUC	σ	SNR
$t_{\text{Hot}}(\mathbf{G})$	0.9305	0.001696	2.0957
$t_{\text{Hot}}(\mathbf{g})$	0.8140	0.002966	1.2609
$t_{\text{mat}}(\mathbf{g})$	0.7812	0.003213	1.0918

We applied the spatio-temporal Hotelling observer to the detector of dim planets orbiting a star (assuming the star and the planet in the same isoplanetic patch), and we simulated data sets for both hypotheses H_0 and H_1 . The apparent magnitude of the star was $m = 6$ and the difference in apparent magnitude between the planet and the star was $\Delta m = 16.74$. More specifically, we simulated $L = 512$ sequences for each hypothesis and, for each sequence, we used $J = 25$ frames containing $M = 64^2$ pixels each. The simulation was set up in order to mimic an astronomical observation. For example, we used typical values for the readout noise as found in [42]. The exposure time was 0.1 ms. The data used to estimate the template vector $\hat{\mathbf{W}}_{\text{Hot}}$ was noiseless and no noise was considered in the wavefront sensor. On the other hand, data on which the detection task was applied were noisy and were obtained with different phase screens.

We considered the SKE/BKE/PKS case and we compared the performance of the spatio-temporal Hotelling observer to that of the purely spatial Hotelling observer $t_{\text{Hot}}(\mathbf{g})$ and the purely spatial matched filter [43, 44] observer $t_{\text{mat}}(\mathbf{g}) = \mathbf{s}^T \mathbf{g}$, where \mathbf{s} is the signal to be detected. In an effort to mimic current practice in astronomical detection, both $t_{\text{Hot}}(\mathbf{g})$ and $t_{\text{mat}}(\mathbf{g})$ were applied to long-exposure data \mathbf{g} obtained by on-chip integration of many short-exposure frames. The observers were run on simulated noisy data. We generated test noise-free image sequences for both the planet-absent and planet-present hypotheses and degraded them with Poisson photon noise and Gaussian readout noise to generate $n = 10,000$ noisy sequences of short-exposure images for the planet-absent hypothesis and n noisy sequences short-exposure images for the planet-present hypothesis. The corresponding long-exposure images were generated as well. These test data were supplied to the observers considered in this comparison, and the corresponding values of the test statistics $t_0^{(1)}, \dots, t_0^{(n)}$ and $t_1^{(1)}, \dots, t_1^{(n)}$ were collected. Binning these values would provide approximated plots of the densities $\text{pr}(t|H_0)$ and $\text{pr}(t|H_1)$ for a particular observer t . For each observer, we estimated the values of the TPF and FPF [see (1)] as:

$$\text{TPF}(\tau) = \frac{|\{i \text{ such that } t_1^{(i)} > \tau, i = 1, \dots, n\}|}{n},$$

$$\text{FPF}(\tau) = \frac{|\{i \text{ such that } t_0^{(i)} > \tau, i = 1, \dots, n\}|}{n},$$

in which the notation $|S|$ stands for the number of elements of the set S , and we varied the value of τ to obtain ROC curves. The ROC curves are reported in Fig. 1, and the corresponding values of the AUC, standard deviation σ on the AUC, and SNR are reported in Table 2. The SNR for the three observers was computed according to (2), in which conditional means $\langle \dots \rangle$ and variances $\text{Var}\{\dots\}$ were replaced by the sample means and sample variances of the $t_0^{(i)}$ and $t_1^{(i)}$. The values of σ were computed as described in [45].

The results reported in Fig. 1 and Table 2 confirm the superiority of the spatio-temporal Hotelling observer with respect to the spatial Hotelling observer and the matched filter observer. The results of Fig. 1 complete the ones reported in [25]. Indeed, in [25], we compared the spatial Hotelling observer with current techniques used in astronomy for point-source detection, and we noted that the spatial Hotelling observer outperforms popular detection algorithms, such as [46]. In this paper, we showed that short-exposure images retain temporal information,

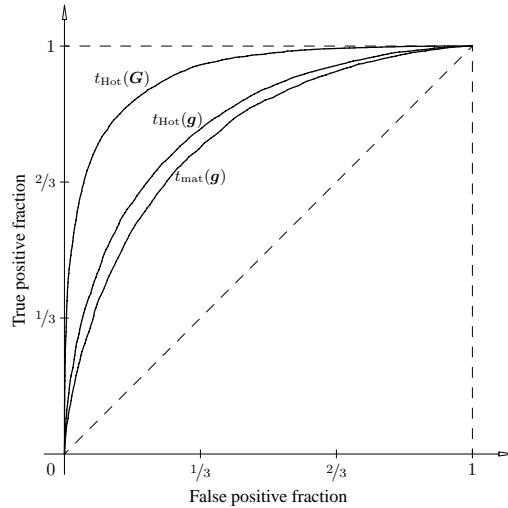


Fig. 1: Plots of the ROC for the observers $t_{\text{Hot}}(\mathbf{G})$, $t_{\text{Hot}}(\mathbf{g})$, and $t_{\text{mat}}(\mathbf{g})$.

which increases detection performance. We see that the spatio-temporal Hotelling observer outperforms current long-exposure detection algorithms as well.

7. Conclusions

In this paper, statistical decision theory was rigorously applied to the problem of signal detection for spatio-temporal data. Three sources of randomness were initially considered: randomness in the object, randomness in the residual point spread function, and randomness due to measurement noise in the detector array. However, for simplicity, we assumed the signal in the object to be nonrandom and at a known location, and the background constant and known. We noted that, in many applications of interest, a complete description of the statistics of the data is not available. Only the first and second moments of the data are required to compute the Hotelling observer. We remarked the importance of the temporal correlations between pixels in the temporal data. Indeed, the Hotelling observer was applied to the whole sequence of temporally-related images, rather than to their average.

In some cases, such as adaptive optics systems, a complete analytical study of the first two moments of the data might be complicated. Therefore, we proposed to estimate means and covariance matrices using simulated data. We implemented one algorithm for the computation of the spatio-temporal Hotelling observer, and we ran the algorithm on a Sony PLAYSTATION 3 cluster. Our implementation took advantage of the computational capabilities of the Cell Broadband Engine Architecture (Cell BE) processor, which equips all of the Sony PLAYSTATION 3 units of our cluster. Thanks to the matrix-inversion lemma, the problem of computing the product between the inverse of a large covariance matrix and the signal was recast to the problem of computing matrix multiplications involving the inverse of a much smaller matrix.

Research concerning an analytical expression for the PSF covariance matrix is currently underway for the case of an ideal thin lens with a weak Gaussian phase perturbation in the pupil. This model might be appropriate for high-performance adaptive optics systems, for which the residual perturbation can be shown to be Gaussian and weak. This study would be of great importance, as it would eliminate the need for simulation in order to estimate the PSF covariance matrix. If an analytical expression for the data covariance matrix is available, other methods—based, for example, on the Landweber algorithm or on the Neumann series expansion—for the

computation of the Hotelling template vector could be investigated as well [21]. In addition, it would be interesting to find an expression for the data's probability density function. With such density, the likelihood ratio could be computed and compared with the Hotelling observer. We expect these two methods to deliver similar detection performance, because, for large mean values, Poisson random variables are very well approximated with Gaussian random variables.

In this study, we relied on simulation code to estimate the mean and covariance of the data. Differences between the actual mean and covariance present in real data and the mean and covariance used in the detection task can arise from two sources: errors in the simulation code or sampling errors because the mean and covariance are estimated from a finite number of simulated sample images. In this work and related previous studies [25,47], we have investigated the latter point in great detail. The former point, effect of model errors on detection performance, has not been investigated for the spatiotemporal Hotelling observer (implemented for the first time in this paper), but it has been studied in the medical literature for purely spatial Hotelling observers [21,48]. The general conclusion is that even crude modeling of the covariance model affords a demonstrable improvement in detection performance, though of course more accurate models are always preferable. This issue will be discussed in a separate paper.

Acknowledgments

The authors would like to acknowledge the Center for Gamma-Ray Imaging at the University of Arizona and NIH grant R37EB000803 for their support.