Optics Communications 411 (2018) 187–194

Contents lists available at ScienceDirect

Optics Communications

journal homepage: www.elsevier.com/locate/optcom

Improvement of correlation-based centroiding methods for point source Shack–Hartmann wavefront sensor

Xuxu Li [a,b], Xinyang Li [a,*], Caixia wang [a]

[a] The Key Laboratory on Adaptive Optics, Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, 610209, China
[b] University of Chinese Academy of Sciences, Beijing, 100039, China

ARTICLE INFO

Keywords:
Shack–Hartmann wavefront sensor
Centroiding methods
Correlation methods
Point source
Fast search algorithms

ABSTRACT

This paper proposes an efficient approach to decrease the computational costs of correlation-based centroiding methods used for point source Shack–Hartmann wavefront sensors. Four typical similarity functions have been compared, i.e. the absolute difference function (ADF), ADF square (ADF^2), square difference function (SDF), and cross-correlation function (CCF) using the Gaussian spot model. By combining them with fast search algorithms, such as three-step search (TSS), two-dimensional logarithmic search (TDL), cross search (CS), and orthogonal search (OS), computational costs can be reduced drastically without affecting the accuracy of centroid detection. Specifically, OS reduces calculation consumption by 90%. A comprehensive simulation indicates that CCF exhibits a better performance than other functions under various light-level conditions. Besides, the effectiveness of fast search algorithms has been verified.

1. Introduction

Shack–Hartmann wavefront sensors have been widely used in adaptive optics (AO) systems for wavefront detection [1–4]. Depending on the target features, charge-coupled devices can obtain extended sub-images or point source spots [3,5]. In both cases, the accuracy of displacement estimation is crucial for improving the performance of wavefront detection. As the centre of gravity (CoG) is invalid when features spread across the whole sub-aperture, solar AO systems use the correlation method. The absolute difference function (ADF) and ADF square (ADF^2) are recommended and typically adopted in solar AO systems [6]. In this study, however, we concentrate on point source wavefront detection under a low-light-level condition, which is dominant for an AO system in applications for star observation and laser communication. Although improved gravity-based methods have been proposed, e.g. threshold CoG (TCoG), weighted CoG, intensity weighted CoG, and iteratively weighted CoG, their performances highly depend on the choice of threshold, weighting function, or number of recursive calculations [7].

Correlation-based methods have a potential immunity to noise. Further, Poyneer et al. [8,9] have provided the first results of point source spots centroiding using a correlation method with the cross correlation function (CCF). However, the principle of selecting a similarity function has been neglected. Therefore, we compare four typical similarity functions using the Gaussian spot model in Section 2. Details about Gaussian interpolation are presented in Section 3. It is seldom discussed but has potential advantages for point source spots.

More importantly, we propose an effective approach to reduce the computational costs of correlation-based methods by employing block-matching fast search algorithms (Section 4). Block-matching is widely used for motion estimation in fields like image registration [10] and video coding [11]. Four typical fast search strategies have been proposed successively: three-step search (TSS) [11], two-dimensional logarithmic search (TDL) [10], cross search (CS) [12], and orthogonal search (OS) [13]. By calculating less matching points, the pixel-level offsets can be obtained much faster.

The paper is organised as follows. In Section 2, similarity functions are analysed and compared in terms of computational costs and robustness. In Section 3, we introduce the interpolation methods used in this paper. In Section 4, the search procedures of four fast search algorithms are explained. Further, in Section 5, we present a simulation to compare the correlation-based methods and the TCoG method as well as the performances of fast search algorithms. Finally, in Section 6, conclusions are drawn.

* Corresponding author.
E-mail address: xyli@ioe.ac.cn (X. Li).

https://doi.org/10.1016/j.optcom.2017.11.059
Received 22 May 2017; Received in revised form 16 November 2017; Accepted 20 November 2017
Available online 20 December 2017

2. Analysis and comparison of similarity functions

2.1. Discrete sampling model

A Shack–Hartmann sensor splits the wavefront into an array of spots by applying micro mirrors. Each spot represents the sampling of a local wavefront. Its displacement is directly proportional to the local gradient, which is used for wavefront reconstruction. Typically, sub-pixel precision is required. However, this is not an easy task—especially under low-light-level conditions because of the presences of photon and readout noise. As a matched filter, correlation-based methods are potentially more resistible to noise then compared with gravity-based methods.

Correlation-based methods firstly calculate the similarity function between a displaced and a reference image. Then, the methods seek the pixel-level offset with the largest similarity value, i.e. the best matching point (BMP). Sub-pixel displacement is typically obtained by an interpolation of BMP neighbours. There exist several candidate functions: ADF, ADF^2, square difference function (SDF), CCF, and normalised cross correlation (NCC). They are defined as Eqs. (1)–(5):

$$S_{\mathrm{ADF}}(i,j) = -\sum_{m=0}^{M-1}\sum_{n=0}^{M-1} |R(m,n) - I(i+m,j+n)|, \tag{1}$$

$$S_{\mathrm{ADF}^2}(i,j) = -\left[\sum_{m=0}^{M-1}\sum_{n=0}^{M-1} |R(m,n) - I(i+m,j+n)|\right]^2, \tag{2}$$

$$S_{\mathrm{SDF}}(i,j) = -\sum_{m=0}^{M-1}\sum_{n=0}^{M-1} [R(m,n) - I(i+m,j+n)]^2, \tag{3}$$

$$S_{\mathrm{CCF}}(i,j) = \sum_{m=0}^{M-1}\sum_{n=0}^{M-1} R(m,n)I(i+m,j+n), \tag{4}$$

$$S_{\mathrm{NCC}}(i,j) = \frac{\sum_{m=0}^{M-1}\sum_{n=0}^{M-1} R(m,n)I(i+m,j+n)}{\sqrt{\sum_{m=0}^{M-1}\sum_{n=0}^{M-1} R(m,n)^2}\sqrt{\sum_{m=0}^{M-1}\sum_{n=0}^{M-1} I(i+m,j+n)^2}}. \tag{5}$$

The functions ADF, ADF^2, and SDF are difference functions, where the negative signs in Eqs. (1)–(3) make it favourable to determine the maximum of each function. The function $R\,(m, n)$ is the reference, $I\,(m, n) = R\,(m + x_0, n + y_0)$ the obtained image with a displacement of (x_0, y_0), where x_0 and y_0 are fractions, while m and n are integers. Assuming that the sizes of R and I are $M \times M$ and $N \times N$ pixels, respectively, all the similarity or difference functions have the size $(N - M + 1)(N - M + 1)$. Meanwhile, the detectable displacement range is $i, j \in [-w, w]$, where $w = (N - M)/2$.

Table 1 shows the exact calculation costs of five similarity functions for $N = 16$ and $M = 8$. Evidently, ADF is far more computationally efficient than all the other functions because there is no multiplication. Löfdahl [6] recommended to add a square operation after calculating each similarity value (ADF^2) as an enhancement of the ADF. Function CCF has less computational costs compared with SDF and NCC.

2.2. Successive-function analysis

In order to study the impact of amplitude difference and equivalent Gaussian width (EGW) difference, which are two main factors when choosing a reference, we introduce a successive-function model. We use $R(x)$ as a reference curve and $I(x)$ as a shifted curve. All equations use a one-dimensional Gaussian function because a Gaussian spot is radially symmetric.

2.2.1. Amplitude difference

The amplitudes of two Gaussian curves are defined as A_{R} and A_{I} to study the impact of the amplitude difference. Meanwhile, they keep the same EGW: σ.

$$R(x) = A_{\mathrm{R}} \exp(-x^2/2\sigma^2) \tag{6}$$

$$I(x) = A_{\mathrm{I}} \exp(-(x - x_0)^2/2\sigma^2) \tag{7}$$

The successive expressions of the four similarity functions ADF, ADF^2, SDF, and CCF have been derived and are shown in Eqs. (8)–(11). Fig. 1 shows the theoretical curves with and without amplitude difference. The x- and y-axes represent the successive displacement between two curves and amplitude of the similarity functions, respectively.

$$\begin{aligned} S_{\mathrm{ADF}}(x) &= -\sqrt{2\pi}\sigma \Big| A_{\mathrm{R}} erf\left[\left(2\sigma^2 \ln\left(A_{\mathrm{R}}/A_{\mathrm{I}}\right) + (x-x_0)^2\right)/2\sqrt{2}\sigma(x-x_0)\right] \\ &\quad - A_{\mathrm{R}} erf\left[\left(2\sigma^2 \ln\left(A_{\mathrm{R}}/A_{\mathrm{I}}\right) - (x-x_0)^2\right)/2\sqrt{2}\sigma(x-x_0)\right]\Big| \end{aligned} \tag{8}$$

$$\begin{aligned} S_{\mathrm{ADF}^2}(x) &= -2\pi\sigma^2 \Big\{ A_{\mathrm{R}} erf\left[\left(2\sigma^2 \ln\left(A_{\mathrm{R}}/A_{\mathrm{I}}\right) + (x-x_0)^2\right)/2\sqrt{2}\sigma(x-x_0)\right] \\ &\quad - A_{\mathrm{I}} erf\left[\left(2\sigma^2 \ln\left(A_{\mathrm{R}}/A_{\mathrm{I}}\right) - (x-x_0)^2\right)/2\sqrt{2}\sigma(x-x_0)\right]\Big\}^2 \end{aligned} \tag{9}$$

$$S_{\mathrm{SDF}}(x) = -\sqrt{\pi}\sigma\left\{(A_{\mathrm{R}}^2 + A_{\mathrm{I}}^2) - 2A_{\mathrm{R}}A_{\mathrm{I}}\exp\left[-(x-x_0)^2/4\sigma^2\right]\right\} \tag{10}$$

$$S_{\mathrm{CCF}}(x) = \sqrt{\pi}\sigma A_{\mathrm{R}}A_{\mathrm{I}}\exp\left[-(x-x_0)^2/4\sigma^2\right] \tag{11}$$

From Fig. 1, we can see that ADF has a cone-shaped distribution near the peak without amplitude difference ($A_{\mathrm{R}} = A_{\mathrm{I}} = 1$). This implies that an equiangular line fitting (ELF) [14] would be the best interpolation method. Because ADF^2 tends to follow a parabolic distribution, a plausible choice would be a parabolic interpolation. The functions SDF and CCF should use a Gaussian interpolation (more details can be found in Section 3).

As the amplitude difference increases, a "flat region" near the peak (highlighted in a rectangle) becomes dominant for ADF. A similar, less evident effect happens to ADF^2. The effect makes the sub-pixel-acquiring process more challenging—especially, when the EGW is large and discrete sampling is inevitable. The reason is that it might end up with more than one local maximum, thereby making the interpolation procedure void. Contrary to this, both SDF and CCF show Gaussian curves with a large amplitude difference. Therefore, the normalisation process in Eq. (5) is not very helpful. Further, it increases computational costs significantly. A larger amplitude for reference is recommended when using CCF.

2.2.2. EGW difference

To analyse the influence of EGW difference, σ_{R} and σ_{I} are defined as the EGWs of two Gaussian curves, respectively, as shown in Eqs. (12) and (13). Furthermore, the successive similarity functions are shown in Eqs. (14)–(17).

$$R(x) = A\exp(-x^2/2\sigma_{\mathrm{R}}^2) \tag{12}$$

$$I(x) = A\exp(-(x-x_0)^2/2\sigma_{\mathrm{I}}^2) \tag{13}$$

$$\begin{aligned} S_{\mathrm{ADF}}(x) = -\sqrt{2\pi}A\Bigg\{ &|\sigma_{\mathrm{I}} - \sigma_{\mathrm{R}}| \\ &+ \Bigg|\sigma_{\mathrm{R}}\left[erf\left(\frac{\sqrt{2}}{2}\frac{x-x_0}{\sigma_{\mathrm{I}}-\sigma_{\mathrm{R}}}\right) + erf\left(\frac{\sqrt{2}}{2}\frac{x-x_0}{\sigma_{\mathrm{I}}+\sigma_{\mathrm{R}}}\right)\right] \\ &+ \sigma_{\mathrm{I}}\left[erf\left(\frac{\sqrt{2}}{2}\frac{x-x_0}{\sigma_{\mathrm{I}}+\sigma_{\mathrm{R}}}\right) - erf\left(\frac{\sqrt{2}}{2}\frac{x-x_0}{\sigma_{\mathrm{I}}-\sigma_{\mathrm{R}}}\right)\right]\Bigg|\Bigg\} \end{aligned} \tag{14}$$