Research Article Vol. 31, No. 4 / 13 Feb 2023 / Optics Express 5990

Optics EXPRESS

Elimination of optical phase sensitivity in a shift, scale, and rotation invariant hybrid opto-electronic correlator via off-axis operation

JULIAN GAMBOA,[1,*] TABASSOM HAMIDFAR,[1] XI SHEN,[1] AND SELIM M. SHAHRIAR[1,2,3]

[1] Department of ECE, Northwestern University, Evanston, IL 60208, USA
[2] Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA
[3] shahriar@northwestern.edu
[*] JulianGamboa2023@u.northwestern.edu

Abstract: The hybrid opto-electronic correlator (HOC) uses a combination of optics and electronics to perform target recognition. Achieving a stable output from this architecture has previously presented a significant challenge due to a high sensitivity to optical phase variations, limiting the real-world feasibility of the device. Here we present a modification to the architecture that essentially eliminates the dependence on optical phases, and demonstrate verification of the proposed approach. Experimental results are shown to agree with the theory and simulations, for scale, rotation and shift invariant image recognition. This approach represents a major innovation in making the HOC viable for real-world applications.

1. Introduction

Ultra-fast image recognition is of significant interest in many fields. Recent advances in machine learning and other implementations of artificial intelligence promise to increase the reliability and precision of image detection schemes, but these typically require powerful processing stages that are still limited in speed. An improved version of the so-called YOLO (you only look once) algorithm, for example, is currently recognized as the fastest such computational technique, yielding inference times close to 12 ms [1]. This speed may be sufficient for some cases, but often falls short for real-time processing or for applications with large datasets consisting of millions of images.

Optical image processing techniques have received much attention in recent years as an alternative to these electronic methods [2–6] thanks in large part to the inherent speed-of-light advantage. One branch of optical image processing has focused on the use of Fourier optics and holography to produce the convolution or cross-correlation of images [7–15]. In these schemes, the Fourier transforms (FT) of the target images are obtained through lenses, multiplied with each other, and finally FT'd again to obtain the final signals. Unfortunately, the multiplication process is not trivial when dealing with optical signals, and various architectures have been proposed with distinct approaches [6,7,16,17]. As an example, the traditional Vander-Lugt correlator uses a hologram to store the amplitude and phase information of the FT of a reference image, which is then used as a filter to multiply the FT of a query image [17]. While this setup has the advantage of simplicity, it is not useful in practice because of the fact that an individual filter has to be made for each reference image. As an alternative, the joint-transform correlator performs the writing and reading of the holograms simultaneously in a single-step process, allowing for more dynamic operation of the device [16]. However, in this case the nonlinear holographic material that is required tends to be quite fragile and difficult to operate [18], again severely limiting the maximum real-world operational speed.

#484149 https://doi.org/10.1364/OE.484149
Journal © 2023 Received 21 Dec 2022; revised 13 Jan 2023; accepted 18 Jan 2023; published 3 Feb 2023

Optics EXPRESS

We previously proposed and demonstrated a correlator architecture [13,14,19] that replaces the analog holographic approach with digital techniques in order to perform the multiplication step. We have estimated that the HOC architecture that incorporates a holographic memory disc (HMD) for rapid retrieval of reference images could reach a maximum real-world operating speed on the order of a few microseconds with current technologies [15]. We have also shown that by incorporating the polar Mellin transform (PMT), the HOC can be operated to produce shift, scale, and rotation invariant (SSRI) correlations [20], requiring only an additional opto-electronic PMT pre-processor (OPP) stage [21].

While the HOC has been successfully demonstrated in different modalities, it has thus far required highly stable optical phases along two critical beam paths. This necessitated the use of an actively isolated optical table, an enclosure that minimizes air currents, and an optical phase scanning stage. For optimal operation of the HOC, it was necessary to maximize the correlation signal by scanning the relevant phase difference. If it is possible to keep the optical pathlengths extremely stable during the whole search process, this approach would in principle work. However, implementing a system with the requisite degree of stability is extremely difficult in practice. In this paper, we describe and demonstrate a technique that enables the HOC to operate optimally without requiring such phase stability. As such, this represents a major innovation in making the HOC viable for real-world applications.

The rest of the paper is organized as follows. Section 2 presents a brief review of the HOC architecture. The off-axis technique that reduces the sensitivity to optical phase instability is presented in Section 3 along with simulation results. Details of the experimental implementation are given in Section 4. Results and a discussion on the implications of the data follow in Section 5, and Section 6 concludes with a summary and outlook.

2. HOC architecture

2.1. Design

The optical portion of the HOC consists of an input stage and an output stage, as shown in Fig. 1. The input stage is itself composed of three arms: the phase scanning (PS) arm, the query image arm, and the reference image arm. The query arm projects an image, H_q, into the optical domain using an SLM (spatial light modulator), while the reference arm uses a volume HMD for the reference image, H_r. In principle, both arms could use SLMs, but even state of the art devices are limited to frame times on the order of milliseconds. In contrast, the speed of an HMD is only limited by how quickly the reading angle or physical position can be changed, which can be accomplished in microsecond speeds using acousto-optic modulators and high-speed rotation stages, allowing us to scan many reference images during a single query frame of the SLM. With the exception of the projection method, both image arms function identically: a lens is placed one focal distance away from the SLM or HMD, projecting the FT, which we label $M_{r,q} = FT\{H_{r,q}\}$, at the opposite focal plane. A beam splitter (BS) reflects part of the $M_{r,q}$ beam towards an FPA (focal plane array), denoted as ${FPA^B}_{r,q}$, at the Fourier plane, thus detecting the intensity of the FT, which we label $B_{r,q}$. The transmitted portion of the $M_{r,q}$ beam continues towards another BS that is used to combine the FT with an auxiliary plane wave (APW), $C_{r,q}$, finally reaching ${FPA^A}_{r,q}$, which is also placed at the Fourier plane, thus detecting the interference between the FT and the APW, which we label $A_{r,q}$. Finally, a third BS is used in conjunction with ${FPA^C}_{r,q}$ to capture the intensity of the APW itself. These signals can be electronically processed to obtain the product of the FTs, as described in Section 2.2 below. In all three arms of the input stage of the HOC, half-wave plates and polarizers are used to control the relative intensities of the beams while maintaining a shared polarization axis.

The output stage consists of SLM_{out}, which is placed one focal distance away from a lens, and FPA_{out}, which is placed at the opposite focal plane of the lens. The SLM projects the result of the electronic processing, and its FT is captured by the FPA. Because the electronic signal contains

Optics EXPRESS

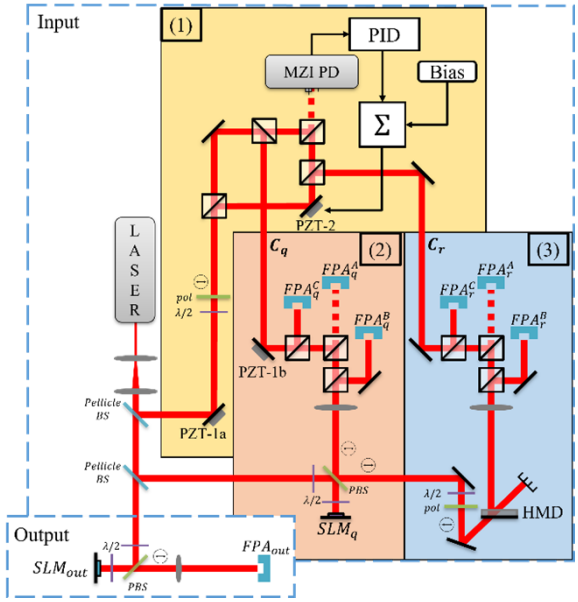

Fig. 1. Optical section of the HOC architecture using an SLM for query images and an HMD for reference images. Not shown: electronic processing stages. The input and output stages are bordered by a blue dashed line. (1, yellow): Phase scanning arm. (2, orange): Query image arm. (3, blue): Reference image arm.

the multiplication of the FTs of the original images, the FPA measurement effectively contains the two-dimensional convolution and correlation signals.

This architecture is inherently shift invariant but requires additional processing in order to achieve rotation and scale invariance. To this end, the input images are pre-processed using the PMT to generate a signature that is rotation and scale invariant. The HOC then operates on these images, thus enabling SSRI correlation. More details on this process can be found in [14,20].

The architecture of the HOC is designed specifically for the task of recognizing objects. As such, it is expected that all images would be converted from the digital domain to the optical domain prior to the use of the HOC. This affords us the flexibility of using the same linear polarization at every stage. However, for some applications, such as in polarization-based optical image encryption [22], it may be necessary to deal with images that have spatially varying polarizations. In order to implement such protocols using the hybrid opto-electronic approach, the system architecture as well as the processing steps have to be modified significantly, with the details depending on the specific task at hand. An analysis of such modifications is beyond the scope of this paper.

2.2. Detected signals and output

In total, the HOC detects three signals for each arm:

$$A_{r,q}(\vec{\rho}) = |M_{r,q}(\vec{\rho}) + C_{r,q}(\vec{\rho})|^2 \quad B_{r,q}(\vec{\rho}) = |M_{r,q}(\vec{\rho})|^2 \quad |C_{r,q}(\vec{\rho})|^2 \tag{1}$$

where $\vec{\rho}=(x,y)$ is the spatial plane at the corresponding FPA. For brevity, the functional dependence on $\vec{\rho}$ will not be explicitly indicated in the remaining equations unless needed. To get rid of the unwarranted terms in the $A_{r,q}$ signals, we compute the following:

$$S_{r,q} = A_{r,q} - B_{r,q} - |C_{r,q}|^2 = M_{r,q}C^*_{r,q} + M^*_{r,q}C_{r,q} \tag{2}$$

Optics EXPRESS

This can be performed rapidly on a pixel-by-pixel basis on an FPGA. Additionally, the same FPGA can multiply the S_r and S_q signals to get an output that contains the products of the FTs:

$$S = S_r \bullet S_q = \alpha^* M_r M_q + \alpha M_r^* M_q^* + \beta^* M_r M_q^* + \beta M_r^* M_q \tag{3}$$

where we have defined $\alpha = C_r C_q$ and $\beta = C_r C^*_q$. This signal is projected back into the optical domain by SLM_{out} and passed through a lens to obtain its FT:

$$\begin{aligned} S_f &= FT\{S\} \approx \alpha^* FT\{M_r M_q\} + \alpha\ FT\{M_r^* M_q^*\} + \beta^* FT\{M_r M_q^*\} + \beta\ FT\{M_r^* M_q\} \\ &= \alpha^* T_1 + \alpha\ T_2 + \beta^* T_3 + \beta\ T_4 \end{aligned}$$

$$\begin{aligned} T_1 &= H_r(\vec{\rho}) \otimes H_q(\vec{\rho}) \\ T_2 &= H_r(-\vec{\rho}) \otimes H_q(-\vec{\rho}) = T_1(-\vec{\rho}) \\ T_3 &= H_q(\vec{\rho}) \odot H_r(\vec{\rho}) \\ T_4 &= H_r(\vec{\rho}) \odot H_q(\vec{\rho}) \end{aligned} \tag{4}$$

where $\otimes$ and $\odot$ represent the 2D convolution and cross-correlation operations, respectively. Here we have assumed that the spatial profile of α and β, and thus the spatial profile of C_r and C_q, are constant, which is the case for perfect plane waves.

The final S_f signal is obtained optically and captured by FPA_{out}, and so only the intensity of the signal can be measured. Therefore, the final output signal is given by:

$$\begin{aligned} |S_f|^2 = {} & |\alpha|^2|T_1|^2 + |\alpha|^2|T_2|^2 + |\beta|^2|T_3|^2 + |\beta|^2|T_4|^2 + (\alpha^2)^* T_1 T_2^* + \alpha^2 T_1^* T_2 \\ & + \alpha^*\beta T_1 T_3^* + \alpha\beta^* T_1^* T_3 + \alpha^*\beta^* T_1 T_4^* + \alpha\beta T_1^* T_4 + \alpha\beta T_2 T_3^* + \alpha^*\beta^* T_2^* T_3 \\ & + \alpha\beta^* T_2 T_4^* + \alpha^*\beta T_2^* T_4 + (\beta^2)^* T_3 T_4^* + \beta^2 T_3^* T_4 \end{aligned} \tag{5}$$

We note that the T terms are the convolution and cross-correlation of real images, and so must themselves be real-valued. Therefore, it follows that:

$$\begin{aligned} |S_f|^2 = {} & |\alpha|^2 T_1^2 + |\alpha|^2 T_2^2 + |\beta|^2 T_3^2 + |\beta|^2 T_4^2 + 2T_1T_2 Re\{\alpha^2\} + 2T_3T_4 Re\{\beta^2\} \\ & + 2Re\{\alpha^*\beta\}(T_1T_3 + T_2T_4) + 2Re\{\alpha\beta\}(T_1T_4 + 2T_2T_3) \end{aligned} \tag{6}$$

It is important to observe that the T^2 terms only depend on the magnitude of α and β, and so are unaffected by the optical phases of C_r and C_q. In contrast, the terms that contain $T_n T_m$, where n and m are different indices between 1-4, directly depend on the phase of α and β and so are affected by the optical phases of C_r and C_q This phase dependence complicates the operation of the HOC, as it is very difficult to stabilize optical phases, especially in free space.

3. Off-axis correlation technique

As noted earlier, we have developed an off-axis correlation technique that makes it possible to detect selectively only the terms that are (a) relevant for measuring the cross-correlation signals and (b) independent of the relative phase between the two plane waves. To explain how this process works, we start by considering the domain of the convolution and correlation signals in the spatial plane for different positions of input images. For conciseness, we will only explicitly derive the case of the T_1 (convolution) and T_4 (correlation) terms. The results for the T_2 and T_3 terms were obtained through the same calculations and will be presented in the end.

Optics EXPRESS

Consider two signals, H_r and H_q, that are centered about (0,0) such that:

$$H_{r,cen}(x,y) = \begin{cases} image_r & (-x_{rw}<x<x_{rw}, \quad -y_{rw}<y<y_{rw}) \\ 0 & otherwise \end{cases}$$
$$H_{q,cen}(x,y) = \begin{cases} image_q & (-x_{qw}<x<x_{qw}, \quad -y_{qw}<y<y_{qw}) \\ 0 & otherwise \end{cases} \tag{7}$$

where x_{rw}, y_{rw}, x_{qw}, and y_{qw} are half of the width and height of the images, respectively. Here the images are assumed to have a finite domain, as is the case with real images. The results of the convolution for this centered case can be written as:

$$\begin{aligned} T_{1,cen}(x,y) &= H_{r,cen}(x,y) \otimes H_{q,cen}(x,y) \\ &= FT^{-1}\{M_{r,cen} \bullet M_{q,cen}\} \end{aligned} \tag{8}$$

where $M_{r,cen}$ and $M_{q,cen}$ are the FTs of $H_{r,cen}$ and $H_{q,cen}$, respectively. Consider now the case where the reference image is shifted by x_{rs} and y_{rs}. Here, the domain of the image will also be shifted:

$$\begin{aligned} H_{r,s} &= \begin{cases} image_r & (-x_{rw}+x_{rs}<x<x_{rw}+x_{rs}, \quad -y_{rw}+y_{rs}<y<y_{rw}+y_{rs}) \\ 0 & otherwise \end{cases} \\ &= H_{r,cen}(x-x_{rs}, y-y_{rs}) \end{aligned} \tag{9}$$

In this scenario, we can find the relationship between the new shifted convolution term and the previous centered convolution term.

$$\begin{aligned} T_{1,rs}(x,y) &= H_{r,cen}(x-x_{rs}, y-y_{rs}) \otimes H_{q,cen}(x,y) \\ &= FT^{-1}\{\exp(-i\,\hat{\omega}_x\,x_{rs}) \bullet \exp(-i\,\hat{\omega}_y\,y_{rs}) \bullet M_{r,cen} \bullet M_{q,cen}\} \\ &= T_{1,cen}(x-x_{rs}, y-y_{rs}) \end{aligned} \tag{10}$$

Additionally, if the H_q image is also shifted by x_{qs} and y_{qs}, then the T_1 term becomes:

$$\begin{aligned} T_{1,rs,qs}(x,y) &= FT^{-1}\{\exp(-i\,\hat{\omega}_x\,(x_{rs}+x_{qs})) \bullet \exp(-i\,\hat{\omega}_y\,(y_{rs}+y_{qs})) \bullet M_{r,cen} \bullet M_{q,cen}\} \\ &= T_{1,cen}(x-x_{rs}-x_{qs}, y-y_{rs}-y_{qs}) \end{aligned} \tag{11}$$

Here, it is clear that a shift in either H_r or H_q will equally result in a shift in the T_1 convolution term. Repeating the derivation for T_2, we obtain a similar result, but now the resulting output shift will be in the opposite direction compared to the same for H_r or H_q.

$$T_{2,rs,qs}(x,y) = T_{2,cen}(x+x_{rs}+x_{qs}, y+y_{rs}+y_{qs}) \tag{12}$$

Following the same steps for the T_4 term, we find that, unlike the convolution terms, the directionality of the output shift now depends on which image is being shifted. This is due to the presence of a complex conjugate in the Fourier-domain expression for the correlation.

$$\begin{aligned} T_{4,cen}(x,y) &= H_{r,cen}(x,y) \odot H_{q,cen}(x,y) \\ &= FT^{-1}\{M^*_{r,cen} \bullet M_{q,cen}\} \end{aligned} \tag{13}$$

In this form, it is clear to see that in the case of shifted input images, the direction of the shift in the conjugated term will be reversed in the output.

$$\begin{aligned} T_{4,rs,qs}(x,y) &= FT^{-1}\{\exp(-i\,\hat{\omega}_x\,(-x_{rs}+x_{qs})) \bullet \exp(-i\,\hat{\omega}_y\,(-y_{rs}+y_{qs})) \bullet M^*_{r,cen} \bullet M_{q,cen}\} \\ &= T_{4,cen}(x+x_{rs}-x_{qs}, y+y_{rs}-y_{qs}) \end{aligned} \tag{14}$$

Optics EXPRESS

Here, a shift in H_r will move the correlation in the opposite direction compared to the same for a shift in H_q. Finally, the shifted T_3 term can be written as:

$$T_{3,rs,qs}(x, y) = T_{3,cen}(x - x_{rs} + x_{qs}, y - y_{rs} + y_{qs}) \tag{15}$$

From Eqs. (11), (12), (14), and (15) it is evident that each T term moves in a different direction in response to both images shifting in x and y. We will denote this as the separation property of the T terms.

While each T term responds differently to a shift in the input images, it may not be evident as to what domain the terms will have in the output plane. To find this, we can write the integral form of the T_1 convolution term when both images are centered about *(0,0)*:

$$T_{1,cen}(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_{r,cen}(x - \tilde{x}, y - \tilde{y}) \bullet H_{q,cen}(\tilde{x}, \tilde{y})\, d\tilde{x}\, d\tilde{y} \tag{16}$$

From this expression, and because $H_{r,cen}$ and $H_{q,cen}$ are zero outside of (($-x_{rw}$, x_{rw}),($-y_{rw}$, y_{rw})) and (($-x_{qw}$, x_{qw}),($-y_{qw}$, y_{qw})), respectively, we can conclude that $T_{1,cen}$ will always be zero outside of the extended range (($-x_{tw}$, x_{tw}),($-y_{tw}$, y_{tw})), where $x_{tw} = x_{rw} + x_{qw}$ and $y_{tw} = y_{rw} + y_{qw}$. This will also be the case for $T_{2,cen}$, $T_{3,cen}$, and $T_{4,cen}$. Furthermore, when the images are shifted, this range will shift in the same direction as the corresponding T term. These properties are summarized in Table 1. By setting $x_{rs} = -x_{qs} = x_{tw}$ and $y_{rs} = -y_{qs} = y_{tw}$, the distribution of the T terms is simplified greatly, as shown in Table 2. This is equivalent to shifting both images in opposite directions by the sum of half their sizes in each axis.

Figure 2 shows a diagram of the regions of the T terms shown in Table 2. Here, the T_1 and T_2 convolution terms are centered about zero, while the T_3 and T_4 correlation terms are separated by a fixed distance and are only non-zero within well-defined regions that do not overlap any other term. Because of this, the product of a correlation term with any other term will always yield a null value. Furthermore, as described in Section 2.2 above, these terms and their products appear in Eq. (6), which we may now simplify for the shifted case by taking into account the lack of overlap from the correlation terms:

$$|S_f|^2{}_{shifted} = |\alpha|^2 T_{1,rs,qs}^2 + |\alpha|^2 T_{2,rs,qs}^2 + |\beta|^2 T_{3,rs,qs}^2 + |\beta|^2 T_{4,rs,qs}^2 + 2T_{1,rs,qs}T_{2,rs,qs}Re\{\alpha^2\} \tag{17}$$

Finally, we are not interested in the convolution terms or their products, which are centered about zero. As such, we may implement an oval DC blocking filter in front of FPA_{out} with a radius of x_{tw} and y_{tw} in each respective axis. This filter will block the convolution terms altogether, allowing us to simplify Eq. (17) further:

$$|S_f|^2_{shifted,\ \ DC\ blocked} = |\beta|^2 T_{3,rs,qs}^2 + |\beta|^2 T_{4,rs,qs}^2 \tag{18}$$

In this final expression, it is clear that the output of the HOC under this off-axis technique will only contain the two complementary cross-correlation terms that are themselves independent of the complex phase of the α and β terms, which contain the optical phase information of the C_r and C_q APWs.

Figures 3 and 4 show simulation results using Eq. (6) with unshifted (on-axis) and shifted (off-axis) inputs, respectively, as well as the dependence on the optical phase of the APWs for both cases. In Fig. 3(B), the cross-correlation and convolution terms overlap in the output plane, and so cannot be distinguished. Additionally, the total power of the output signal depends sinusoidally on the phase difference between the two APWs, as shown in Fig. 3 (C). In contrast, Fig. 4(B) shows how the shift in the images results in a broad separation of the cross-correlation and convolution terms, allowing the former to be independently measured. Figure 4(C) confirms that the power of the isolated correlation terms is independent of the phases of the APWs, essentially eliminating the need for high phase stability and additional phase scanning. If the convolution terms in the red area are blocked, these results are equivalent to those obtained using Eq. (18).

Optics EXPRESS

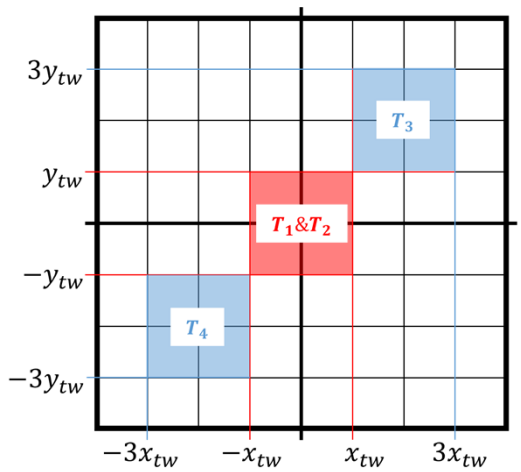

Fig. 2. Distribution of the convolution and cross-correlation terms after shifting the input images by x_{rs} = -x_{qs} = x_{tw} and y_{rs} = -y_{qs} = y_{tw}, as shown in Table 2. Red: Area where the T_1 and T_2 convolution terms are non-zero. Blue: Area where the T_3 and T_4 cross-correlation terms are non-zero.

Table 1. Location and domain of each T term for shifted input images.

Term	Center Coordinate	Non-zero Range
$T_{1,rs,qs}$	$(+x_{rs}+x_{qs},\ \ +y_{rs}+y_{qs})$	$x\|(-x_{tw}+x_{rs}+x_{qs},\ \ x_{tw}+x_{rs}+x_{qs})$ $y\|(-y_{tw}+y_{rs}+y_{qs},\ \ y_{tw}+y_{rs}+y_{qs})$
$T_{2,rs,qs}$	$(-x_{rs}-x_{qs},\ \ -y_{rs}-y_{qs})$	$x\|(-x_{tw}-x_{rs}-x_{qs},\ \ x_{tw}-x_{rs}-x_{qs})$ $y\|(-y_{tw}-y_{rs}-y_{qs},\ \ y_{tw}-y_{rs}-y_{qs})$
$T_{3,rs,qs}$	$(+x_{rs}-x_{qs},\ \ +y_{rs}-y_{qs})$	$x\|(-x_{tw}+x_{rs}-x_{qs},\ \ x_{tw}+x_{rs}-x_{qs})$ $y\|(-y_{tw}+y_{rs}-y_{qs},\ \ y_{tw}+y_{rs}-y_{qs})$
$T_{4,rs,qs}$	$(-x_{rs}+x_{qs},\ \ -y_{rs}+y_{qs})$	$x\|(-x_{tw}-x_{rs}+x_{qs},\ \ x_{tw}-x_{rs}+x_{qs})$ $y\|(-y_{tw}-y_{rs}+y_{qs},\ \ y_{tw}-y_{rs}+y_{qs})$

Table 2. Location and domain of each T term for x_{rs} = -x_{qs} = x_{tw} and y_{rs} = -y_{qs} = y_{tw}.

Term	Center Coordinate	Non-zero Range
$T_{1,rs,qs}$	$(0,\ \ 0)$	$x\|(-x_{tw},\ \ x_{tw})$ $y\|(-y_{tw},\ \ y_{tw})$
$T_{2,rs,qs}$	$(0,\ \ 0)$	$x\|(-x_{tw},\ \ x_{tw})$ $y\|(-y_{tw},\ \ y_{tw})$
$T_{3,rs,qs}$	$(+2x_{tw},\ \ +2y_{tw})$	$x\|(x_{tw},\ \ 3x_{tw})$ $y\|(y_{tw},\ \ 3y_{tw})$
$T_{4,rs,qs}$	$(-2x_{tw},\ \ -2y_{tw})$	$x\|(-3x_{tw},\ \ -x_{tw})$ $y\|(-3y_{tw},\ \ -y_{tw})$

Optics EXPRESS

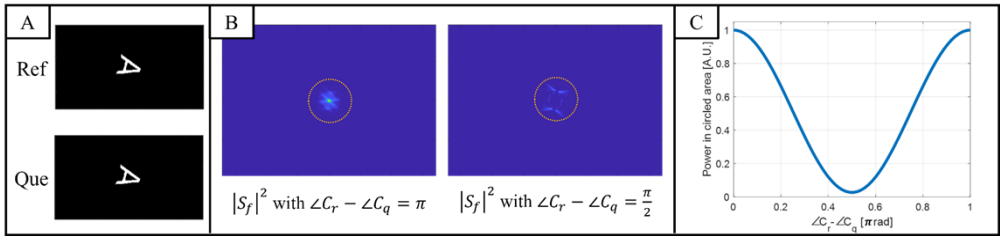

Fig. 3. Simulation results for an HOC with unshifted images. (A): Input images. (B): $|S_f|^2$ Output for an APW phase difference of π and π/2. The yellow circles denote the area within which both the cross-correlation and convolution terms are confined. (C): APW phase difference vs. the total power of the encircled area.

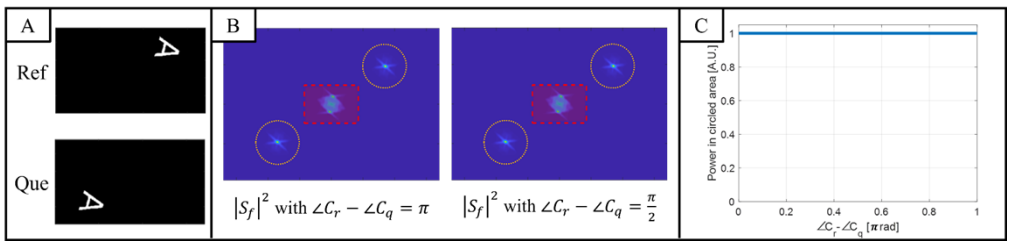

Fig. 4. Simulation results for an HOC with images shifted in opposite directions by ($\pm x_{tw}, \pm y_{tw}$). (A): Input images. (B): $|S_f|^2$ Output for an APW phase difference of π and π/2. Here, the yellow circles denote the areas within which the correlation terms are confined. The red squares show the ($-x_{tw} : x_{tw}$, $-y_{tw} : y_{tw}$) range within which the convolution terms are confined. (C): APW phase difference vs. the total power of the encircled area.

4. Experimental Implementation

We note that, for a general situation with arbitrary images, neither the sizes of any of the query images nor any inherent shift they may possess are known a priori. As such, it is not possible to precisely determine the requisite values of $x_{rs} = -x_{qs} = x_{tw}$ and $y_{rs} = -y_{qs} = y_{tw}$. However, the query image will exist within the physical dimensions of the SLM, which will define both the non-zero range and center position of an unshifted image. Similarly, for the reference arm, the image will be constrained by the HMD from which it is read. Thus, we may redefine x_{qw}, y_{qw}, x_{rw}, and y_{rw} to be half of the width and height of the active area of the projection device, rather than the particular image, simultaneously updating the definition of x_{tw} and y_{tw} to use these new values. It is easy to see that this does not affect the definition of $H_{r,cen}$ and $H_{q,cen}$ from Eq. (7), and so the rest of the derivation remains the same. By considering the dimensions of the active area rather than the images, we are able to induce a shift of $\pm x_{tw}$ and $\pm y_{tw}$ by moving the SLM and HMD at the input planes of the image arms. Of course, an image may yet be shifted within the active area, but it will never extend beyond it. Thus, the convolution and correlation terms will still move in response to a shift in the image, but they will be confined to a predetermined region of the output plane.

In the experiments reported here, the HMD was written using an SLM of the same size as the one in the query arm, which measured 6.9×3.9 mm, yielding values of $x_{rw} = x_{qw} = 3.45\ mm$, $y_{rw} = y_{qw} = 1.95\ mm$, $x_{tw} = 6.9\ mm$, and $y_{tw} = 3.9\ mm$. In practice, it is difficult to shift the SLM and HMD by exactly $\pm x_{tw}$ and $\pm y_{tw}$. However, it is not strictly necessary to shift the images by these exact values. From the equations shown in Table 1, it is easy to see that as long as H_r and H_q are shifted in opposite directions such that the distance between them is greater than

Optics EXPRESS

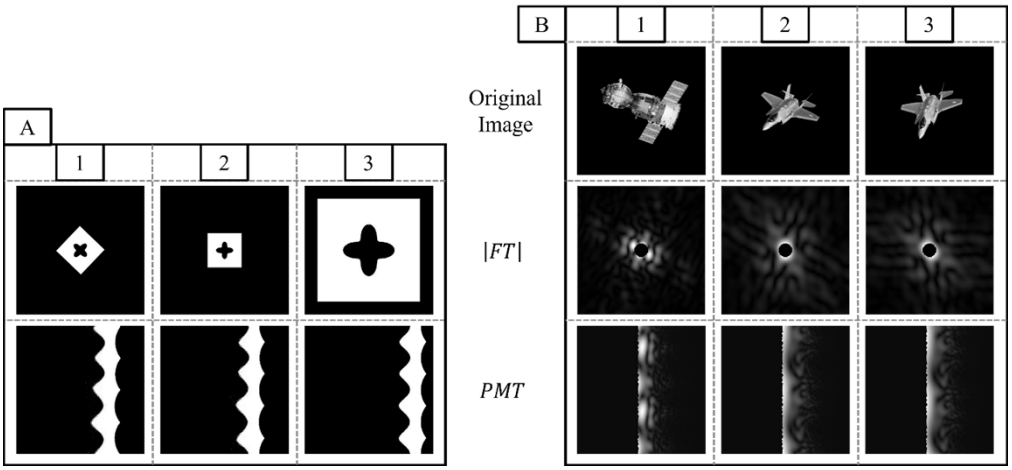

Fig. 5. (A): PMTs of artificial FTs. (B): PMTs of real images. Bottom row (both A and B): PMTs used for the experiments presented here. Middle row (both A and B): Magnitude of the FTs used to generate the PMTs. Top row (B only): original images that correspond to the FTs in the middle row. (Column A.1): Reference FT and the corresponding PMT. (Column A.2): FT and PMT of the same image as (A.1) with 45° of rotation. (Column A.3): FT and PMT of the same image as (A.1) with 45° of rotation and scaled by a factor of 3.75. (Column B.1): Original Image, FT, and PMT of a picture of a Soyuz capsule. (Column B.2): Original Image, FT, and PMT of a picture of an F-22 fighter jet. (Column B.3) Original Image, FT, and PMT of the same image as (B.2) with 30° of rotation.

$2x_{tw}$ and $2y_{tw}$, the convolution and correlation terms will never overlap. Notably, if this shift is asymmetric, which may often be the case in the real world, then the convolution terms will not be centered about *(0,0)* but will instead exist in opposite quadrants to the correlation terms. As such, if we want to block the unwanted signals, the DC block in Fig. 4 can be replaced by an ellipse. In practice, a better approach is to detect only one of the cross-correlations signals (since the two cross-correlation signals contain the same information) in one of the quadrants (depending on the choice of off-axes shifts employed) and block the signals outside the region of interest.

We have previously shown how the PMT can be used to allow the HOC to recognize images with variations in shift, scale, and rotation [14,20]. Figure 5 shows the PMTs that were used as input images for the experiments presented here. In Fig. 5(A), we have used artificially generated FTs with shapes for which the PMT images are easy to interpret. Since these FTs are real-valued, there is no real-valued image pattern that would produce these FTs. As such, this case is meant to be for illustrative purposes only. It should be noted [14,20] that when carrying out the PMT process, it is essential to exclude a small area around the center of the FTs, which is equivalent to applying a DC block. For the cases shown in Fig. 5(A), the FTs were by design constructed with null values at the center, thus obviating the need for applying a DC block during the PMT process. In Fig. 5(B), we have used actual images. Since the FTs of these images have non-zero DC values, we have employed the requisite DC block with the same radius for each image.

5. Results and discussion

A 2 mm thick PQ:PMMA HMD with 1,320 images was created as described in [21]. The disc was installed in the reference arm of the HOC, serving as the projection method for the reference image. The HMD readout was set to the PMT in Fig. 5(A.1) and kept constant throughout the experiments presented here. Thus, all of the data share the same reference image. Additionally,

Optics EXPRESS

the PZTs and PID loop of the PS arm were disabled, and the optical bench was not actively isolated. Under this condition, the absolute and relative phases of the APWs vary randomly, as we had determined before. The results reported here demonstrate that the HOC yields signals that do not depend on these phase variations, as a result of using the off-axis technique.

Figure 6 shows the optical output of the HOC for three sample experiments. The FPA_{out} captures displayed in Figs. 6(A.2), 6(B.2), and 6(C.2) show various peaks that correspond to the convolution signal, the correlation signal, and additional noise from the optical FT. The region corresponding to the correlation was identified via the equations in Table 1, considering the physical shift of the SLM and the HMD. This location remains unchanged for any experiment carried out with the same shifts. Here we identify two figures of merit for the correlation results: the total power and the peak value, where the former is defined as the sum of the pixel values in the correlation region. It is useful to normalize these two coefficients separately, as they will have entirely different scales. To do this, we divide each value in an experiment by that obtained in a reference autocorrelation. The experiment shown in Fig. 6(A) was used as such an

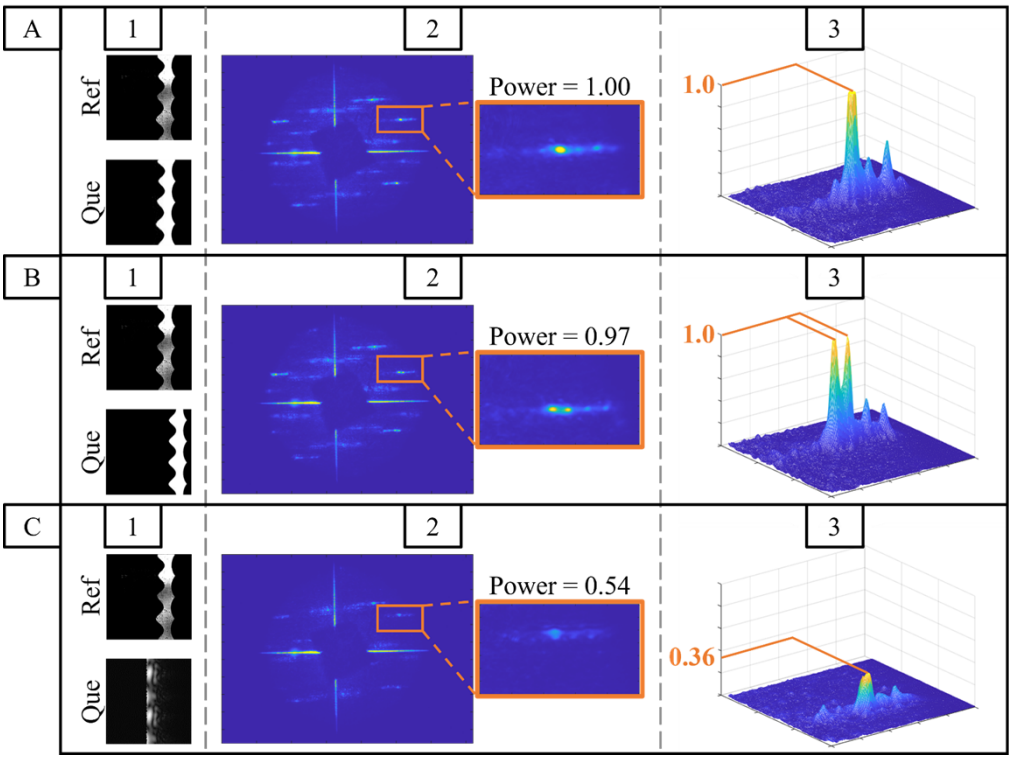

Fig. 6. Three sample experiments on the HOC. In all cases the reference image corresponds to the PMT in Fig. 5(A.1), as read out from an HMD. (A): Autocorrelation. This data was used to normalize the total and peak powers shown in all other experiments. (B): Correlation between the reference and the PMT in Fig. 5(A.3), wherein there is a relative shift and scale between the two '*original*' images. (C): Correlation between the reference and the PMT in Fig. 5(B.1), wherein no match is expected. (1): The reference and query images. (2): The optical signal measured on FPA_{out} with a DC-blocking mask. All of the insets show the same region and correspond to the isolated correlation signal. The 'Power' is taken to be the sum of the pixel values in this region divided by the result from the autocorrelation. (3): A perspective view of the isolated correlation signal, where an orange line highlights the peak. The data is here scaled by the peak value of the autocorrelation.

Optics EXPRESS

autocorrelation and serves as a reference with which to normalize subsequent results; thus, the total power and peak values shown in Figs. 6(A.2) and 6(A.3) are both equal to unity by construct.

Figure 6(B) shows the case where the query image is rotated and scaled with respect to the reference image. In this case, the PMT has a shift to the right due to the scaling, and an upward shift due to the rotation, where any section of the figure that has exited the top is reinserted at the bottom. The latter behavior is due to the cyclical nature of the PMT, and effectively splits the correlation into two sections [14,20]. This splitting manifests itself as a separation of the correlation peak into two smaller peaks in the horizontal axis of the output. The precise separation of the peaks is directly proportional to the rotation of the original image, modulo π, and so can be used to extract this information. Similarly, the horizontal shift in the PMT is proportional to the scale of the image, and results in a vertical shift of the output. Both of these effects can be observed in Figs. 6(B.2) and 6(B.3). The total correlation power was measured to be 0.97, indicating a match between the query and reference images. Additionally, the peak also had a value of unity, but this will not always be the case for images with rotation, and so the total power is the preferred figure of merit for SSRI image recognition. The peak value may be used as a secondary coefficient, whereby a large value indicates an increased confidence in the match. Figure 6(C) presents an experiment where the reference and query images were entirely different. To determine whether we have a match or not, we chose to use 0.8 as the threshold value of the (normalized) total power. The choice of this value for the threshold is justified from comparisons of the results for all the query images, as illustrated in the next paragraph. In this case, the total power shown in Fig. 6(C.2) was measured to be 0.54, thus indicating that no match was found.

A total of 85 experiments like those shown in Fig. 6 were performed, a number not previously achievable in the same experimental session due to the phase stability requirement. Figure 7 shows a violin scatter plot of the results, where the vertical axis presents the normalized total correlation power. The curves are interpolated splines of the histograms of the data and are used for visualization purposes only. The threshold value of 0.8 mentioned above was used to separate the data into 'matches' (blue) and 'mismatches' (red), achieving a detection rate of 100%. In total, the 45 results determined to be matches present a median value of 1.024, with a standard deviation of 0.062. These results show successful operation of the HOC for SSRI target recognition but are limited in that they only utilize an artificial PMT as the reference image. Further experiments are required with a real-world reference image in order to better estimate the performance of this device.

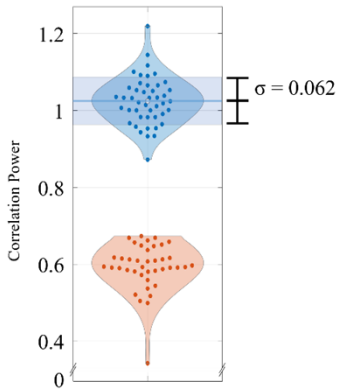

Fig. 7. Distribution of the total correlation power from 85 experiments on the HOC using the off-axis technique. The curves are interpolated splines of the histograms of the data and are used to improve visualization of the results. The values are normalized according to a reference autocorrelation. Blue: Data above a threshold of 0.8, presenting a median value of 1.024 with a standard deviation of 0.062. Red: Data below a threshold of 0.8.

Optics EXPRESS

6. Conclusion

Previous iterations of the HOC required a complex and unstable PS segment in order to extract a usable correlation signal from the output. The off-axis technique presented here allows this architecture to perform image correlations without the need for active control of the optical phases of the APWs. Both theoretical and practical aspects of this approach have been investigated, yielding results that indicate a successful decoupling of the results from optical phase instabilities. Images with variations in shift, scale, and rotation were tested without phase scanning or active isolation of the optical bench. The system was able to categorize the images correctly and consistently as 'match' or 'no match,' while also providing information about the relative rotation and scale between the two inputs. By incorporating the off-axis technique into the HOC, the phase stability requirement is essentially removed, thus opening the doors to real-world implementations of this image correlation architecture.

Funding. Air Force Office of Scientific Research (FA9550-18-01-0359).

Disclosures. The authors declare no conflict of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. J. Li, Z. Su, J. Geng, and Y. Yin, "Real-time Detection of Steel Strip Surface Defects Based on Improved YOLO Detection Network," in *IFAC-PapersOnLine* (2018), Vol. 51.
2. N. U. Dinc, J. Lim, E. Kakkava, D. Psaltis, and C. Moser, "Computer generated optical volume elements by additive manufacturing," Nanophotonics **9**(13), 4173–4181 (2020).
3. R. Xu, P. Lv, F. Xu, and Y. Shi, "A survey of approaches for implementing optical neural networks," Opt. Laser Technol. **136**, 106787 (2021).
4. X. Sui, Q. Wu, J. Liu, Q. Chen, and G. Gu, "A review of optical neural networks," IEEE Access **8**, 70773–70783 (2020).
5. D. Mengu, Y. Rivenson, and A. Ozcan, "Scale-, Shift-, and Rotation-Invariant Diffractive Optical Networks," ACS Photonics **8**(1), 324–334 (2021).
6. A. Alfalou and C. Brosseau, "Recent Advances in Optical Image Processing," in *Progress in Optics* (2015), Vol. 60, Chapter 2, pp. 119–262.
7. B. Javidi and C.-J. Kuo, "Joint transform image correlation using a binary spatial light modulator at the Fourier plane," Appl. Opt. **27**(4), 663–665 (1988).
8. A. Heifetz, J. T. Shen, J.-K. Lee, and M. S. Shahriar, "Translation-invariant object recognition system using an optical correlator and a super-parallel holographic random access memory," Opt. Eng. **45**(2), 025201 (2006).
9. D. Casasent and D. Psaltis, "Scale invariant optical correlation using Mellin transforms," Opt. Commun. **17**(1), 59–63 (1976).
10. D. Casasent and D. Psaltis, "Position, rotation, and scale invariant optical correlation," Appl. Opt. **15**(7), 1795–1799 (1976).
11. B. Javidi, J. Li, and Q. Tang, "Optical implementation of neural networks for face recognition by the use of nonlinear joint transform correlators," Appl. Opt. **34**(20), 3950–3962 (1995).
12. M. S. Monjur, M. F. Fouda, and S. M. Shahriar, "All optical three dimensional spatio-temporal correlator for automatic event recognition using a multiphoton atomic system," Opt. Commun. **381**, 418–432 (2016).
13. M. S. Monjur, S. Tseng, R. Tripathi, J. J. Donoghue, and M. S. Shahriar, "Hybrid optoelectronic correlator architecture for shift-invariant target recognition," J. Opt. Soc. Am. A **31**(1), 41–47 (2014).
14. J. Gamboa, M. Fouda, and S. M. Shahriar, "Demonstration of shift, scale, and rotation invariant target recognition using the hybrid opto-electronic correlator," Opt. Express **27**(12), 16507 (2019).
15. J. Gamboa, T. Hamidfar, and S. Shahriar, "Integration of a PQ:PMMA holographic memory device into the hybrid opto-electronic correlator for shift, scale, and rotation invariant target recognition," Opt. Express **29**(24), 40194–40204 (2021).
16. F. T. S. Yu and X. J. Lu, "A real-time programmable joint transform correlator," Opt. Commun. **52**(1), 10–16 (1984).
17. A. vander Lugt, "Signal detection by complex spatial filtering," IEEE Trans. Inform. Theory **10**(2), 139–145 (1964).
18. D. A. Gregory, J. A. Loudin, and H.-K. Liu, "Joint transform correlator limitations," Proc. SPIE **1053**, 198–207 (1989).
19. M. S. Monjur, S. Tseng, M. F. Fouda, and S. M. Shahriar, "Experimental demonstration of the hybrid opto-electronic correlator for target recognition," Appl. Opt. **56**(10), 2754–2759 (2017).
20. M. S. Monjur, S. Tseng, R. Tripathi, and M. S. Shahriar, "Incorporation of polar Mellin transform in a hybrid optoelectronic correlator for scale and rotation invariant target recognition," J. Opt. Soc. Am. A **31**(6), 1259–1272 (2014).

Optics EXPRESS

21. J. Gamboa, X. Shen, T. Hamidfar, and S. M. Shahriar, "Ultrafast image retrieval from a holographic memory disc for high-speed operation of a shift, scale, and rotation invariant target recognition system," in *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS)* (2022), pp. 505–518.
22. Q. Wang, D. Xiong, A. Alfalou, and C. Brosseau, "Optical image encryption method based on incoherent imaging and polarized light encoding," Opt. Commun. **415**, 56–63 (2018).