2. OPTICAL CORRELATION

2.1 CORRELATION AND OPTICS

Fourier optical signal processing is a class of operations by which intentional modification of the spectrum (Fourier transform) of an image occurs. An important class of Fourier optical signal processing systems is the correlators. The usual task for a correlator is to quantify the similarity between two patterns. Of the two patterns to be correlated, one is designated as the reference image, \( r(x,y) \), and is the template for the targets which one wishes to detect in the input scene. The scene image \( s(x,y) \) is searched for targets, which may be buried in clutter, rotated, obscured, or embedded in additive noise. The objective of correlation is to identify and locate all instances of targets which match the reference in the scene image.

Typically, the similarity between two patterns \( s(x,y) \) and \( r(x,y) \) is measured by their cross-correlation function stated by the Wiener-Khinchine theorem:\[1\]

\[
O_{sr}(x,y) = s(x,y) \otimes r(x,y) = \iint s(\xi,\eta) \cdot r^*(\xi-x,\eta-y) d\xi d\eta
\] (2-1)

where \( \otimes \) denotes the correlation operator. Fig. 2-1 demonstrates an example of the two-dimensional correlation function: \( r(x,y) \), a disk function of 14 pixels base radius and height 1 is depicted in Fig. 2-1a, and \( s(x,y) \), a rectangle with a 35x35 pixel base and height 0.5 is depicted in Fig. 2-1b. Their cross-correlation function \( s(x,y) \otimes r(x,y) \) is shown in Fig. 2-1c, and the autocorrelation function of \( r(x,y) \), that is \( r(x,y) \otimes r(x,y) \), is shown in Fig. 2-1d. Note that the base dimensions of the correlation function is the sum of the corresponding dimensions of the correlated images.
Computation of such a function is numerically intensive, and thus pattern recognition is a computationally demanding problem; this is particularly true when dealing with two-dimensional images. Fourier optical correlators\(^2\),\(^3\) are well-suited for pattern recognition because of their ability to perform an analog correlation function without any numerical computation, by simply utilizing the ability of a lens to perform two-dimensional Fourier transforms\(^4\). This computational ability is executed in parallel and its speed is unrelated to the image size.

As shown in Fig. 2-2, an input image \(s(x, y)\) placed in the front focal plane of a lens of focal length \(f\) is illuminated by coherent light of wavelength \(\lambda\). The field distribution \(U(x', y')\) at the back focal plane is given by\(^5\):
Thus \( U(x', y') \) is proportional to the complex Fourier transform of the input transmissivity distribution.

\[
U(x', y') = \frac{-i}{\mathcal{F}} \int s(\xi, \zeta) \cdot \exp\left[-i \frac{2\pi}{\lambda f}(\xi x' + \zeta y')\right] d\xi d\zeta
\]  

(2-2)

The advantages of using a lens in producing Fourier transforms and, consequently, correlation, lie in parallel processing and execution at the speed of light, as opposed to the serial processing and speed limitations of even the fastest digital processors available.

In the following, the two dimensional Fourier transforms (spectra) of \( s(x, y) \) and \( r(x, y) \) are expressed as:

\[
\mathfrak{F}[s(x, y)] = S(v_x, v_y) = S(v_x, v_y) \exp \cdot \phi_S(v_x, v_y)
\]

\[
\mathfrak{F}[r(x, y)] = R(v_x, v_y) = R(v_x, v_y) \exp \cdot \phi_R(v_x, v_y)
\]

(2-3)

where \( \mathfrak{F} \) denotes the Fourier transform operator and \( v_x \) and \( v_y \) indicate the spatial frequencies given by \( v_x = \frac{x}{\lambda f} \) and \( v_y = \frac{y}{\lambda f} \).
The cross-correlation function $O_{SR}(x,y)$ can be performed as a Fourier transform of the product of the respective Fourier transforms of the patterns\[^5\]:

\[
O_{SR}(x,y) = \mathcal{F}\{s(\xi, \zeta) \cdot r^*(\xi - x, \zeta - y) d\xi d\zeta\} = s(x,y) \otimes r(x,y) = \\
\mathcal{F}^{-1}\left\{[S(v_x, v_y) \exp i\varphi_s(v_x, v_y)] \cdot [R(v_x, v_y) \exp -i\varphi_R(v_x, v_y)]\right\}
\]

(2-4)

Thus an optical correlator involving two Fourier transforming lenses can compute the correlation function of two distributions, provided that a way of realizing the product $[S(v_x, v_y) \exp i\varphi_s(v_x, v_y)] \cdot [R(v_x, v_y) \exp -i\varphi_R(v_x, v_y)]$ is possible. Figure 2-3 displays a block diagram of a correlation-theorem based optical correlator system.

![Figure 2-3: Block diagram of an optical correlator.](image-url)
2.2 CLASSICAL MATCHED FILTER

In the thin-hologram optical correlator of VanderLugt\textsuperscript{[6]}, as well as its application, the classical matched-filter correlator\textsuperscript{[7]} a Fourier-transform product is realized by coherently illuminating a filter mask (with at least one transmissivity coefficient proportional to the complex spectrum of the reference) with the Fourier transform of the scene. The difficulty lies in synthesizing such a complex filter. VanderLugt’s breakthrough was the use of interference with a second, coherent beam to convert amplitude and phase information to intensity variations that were recorded on intensity-responding photographic film.

As shown in Fig. 2-4, an object transmissivity \(r(x,y)\) is Fourier transformed by the lens \(L\), yielding at the filter plane an amplitude distribution proportional to its complex Fourier transform \(R(v_x, v_y)\exp i \varphi_R(v_x, v_y)\). The tilted wave emerging from the prism produces an amplitude distribution given by:

\[
A(v_x, v_y) = A(v_x, v_y) \cdot \exp(-i2\pi \frac{\sin \theta}{\lambda} v_y)
\]  \hspace{1cm} (2-5)

where \(\theta\) is the tilt angle and \(\lambda\) is the coherent wavelength. The intensity distribution at the filter can be expressed as:

\[
I(v_x, v_y) = |A(v_x, v_y) + R(v_x, v_y)|^2 = |A(v_x, v_y)|^2 + |R(v_x, v_y)|^2 + 2A(v_x, v_y)R(v_x, v_y) \cdot \cos\left[2\pi \frac{\sin \theta}{\lambda} v_y - \varphi_R(v_x, v_y)\right]
\]  \hspace{1cm} (2-6)

The third term in Eq. 2-6 shows that both the amplitude \(R(v_x, v_y)\) and the phase \(\varphi_R(v_x, v_y)\) of the reference Fourier transform can be recorded as intensity variations, and thus as filter transmissivity. Thus the first holographic realization of a complex filter mask was developed.
Because it is utilizing the third term in 2-6, the classical matched-filter correlator\cite{7} (Fig. 2-5) can be seen as a direct application of the VanderLugt filter. A linear space-invariant filter is called a \textbf{matched filter} for an image $r(x,y)$ if its impulse response $h(x,y)$ is given by $h(x,y) = r^*(-x,-y)$. For real-valued images, $h(x,y)$ is the point-inversion of $r(x,y)$\cite{8}.

In the 4-f matched-filter correlator shown in Fig. 2-5, a scene image is displayed at the input plane 1, which is coherently illuminated, and a lens forms the two-dimensional Fourier transform of the scene image at plane 2. A pre-fabricated matched-filter of a reference image is placed at the Fourier plane. The exiting field from the filter plane contains a term proportional to the filter transmissivity (spectrum of the reference image) times the field incident to the filter (spectrum of the scene image); thus the cross-product of the spectra (Eq. 2-4) is realized. A second lens is used to take an inverse Fourier transform. The output plane 3 displays the cross-correlation function $OSR(x,y)$ of the two images.
The difficulties with a filter-based correlator lie in the fabrication of the filter and the optical alignment. Once the filter is fabricated, it cannot be updated. In addition, precise alignment is required so that both the recorded reference spectrum and the real-time scene spectrum coincide: the filter must be positioned to within a fraction of the diffraction limit $\Delta$ of the input aperture, where $\Delta$ is given by:

$$\Delta = 2(\lambda f/X)$$

(2-7)

where $X$ is the input aperture width. For $\lambda = 0.6\mu m$, $f = 50mm$, and $X = 35mm$, then $\Delta = 20\mu m$; the required accuracy is some fraction of this. A comparable accuracy is required in the orthogonal direction.

A considerably improved variant of the classical-matched filter is the phase-only filter, which achieves high light throughput by modifying only the phase of the incident light\cite{9,10,11}. (see section 2.4II for definitions and discussion). The significant performance improvements\cite{12} introduced by the phase-only filter and the binary phase-only filter have helped maintain a strong interest in the feasibility of an operational correlator. In fact, the only field-demonstrated correlator up today\cite{13}, was based on the binary phase-only technique.
2.3 Joint Transform Correlator

An alternative approach in correlation architectures is the Joint Transform Correlator (JTC), in which the interference pattern of the Fourier transforms of two images, the scene and the reference, is recorded by an intensity-responding detector (square-law receiver), such as a photographic film, placed at the Fourier plane. The square-law reception generates the joint Fourier transform from the joint power spectrum, thus producing the cross-correlation components as transmissivity function of the “filter”. This transparency is subsequently illuminated with a coherent plane wave to obtain the correlation output.

By putting both the reference and the scene at the input plane, the JTC offers certain advantages over a filter-type correlator:

*Figure 2-6: Joint Transform Correlator.*
1. precise alignment requirement between the prefabricated filter and the scene
   spectrum is not necessary,
2. a prefabricated complex-valued filter is not required in the Fourier plane,
3. a priori knowledge of the reference image is not necessary, and
4. a JTC can be operated in real-time in a compact design.

Assuming images symmetrically positioned along the vertical axis, the input signal
\( g(x, y) \) consists of the reference, \( r(x, y-y_0) \), and the scene, \( s(x, y+y_0) \). Their joint Fourier
transform, \( G(v_x, v_y) \), is the sum of their individual complex Fourier transforms:

\[
G(v_x, v_y) = R(v_x, v_y) \exp[i(\phi_R(v_x, v_y) - 2\pi v_0 v_y)] + S(v_x, v_y) \exp[i(\phi_S(v_x, v_y) + 2\pi v_0 v_y)]
\]

(2-8)

This field is recorded by an intensity-responding detector at the filter plane. Originally this was a photographic film\[^{14}\,^{15}\]. Any linearly absorbing medium can be used; the term \textit{classical} square law receiver function:

\[
f_{CL}(v_x, v_y) = |R(v_x, v_y) + S(v_x, v_y)|^2
\]

(2-9)
is expressing such an operation regardless of the nature of the medium. The result of this
operation is the mixing the spectra so that the cross-spectral products are produced:

\[
f_{CL}(v_x, v_y) = |R(v_x, v_y) + S(v_x, v_y)|^2 + 2\text{Real}[S(v_x, v_y)R^*(v_x, v_y)(1 + \cos 2\pi v_0 v_y)]
\]

(2-10)

Thus the filter transmissivity, which is assumed proportional to the recorded
intensity, consists of the dc \( |R(v_x, v_y)|^2 \) and \( |S(v_x, v_y)|^2 \), and the cross-products \( S(v_x, v_y)R^*(v_x, v_y) \).
and \( S^* (v_x, v_y) R (v_x, v_y) \). The last two components are superimposed on the carrier frequency prescribed by the scene-reference separation \( (2y_0) \).

On illuminating by a plane wave, the output field \( g(x', y') \) is produced by an inverse Fourier transformation of the field exiting the filter, which is assumed proportional to the filter transmissivity, that is, of \( f_{cl}(v_x, v_y) \):

\[
g(x', y') = U_{SS}(x', y') + U_{RR}(x', y') + U_{RS}(x', y') + U_{SR}(x', y')
\]

where

\[
U_{SS}(x', y') = s(x', y') \otimes s(x', y') = \iint s(\xi, \zeta) s^*(\xi-x', \zeta-y') d\xi d\zeta
\]

\[
U_{RR}(x', y') = r(x', y') \otimes r(x', y') = \iint r(\xi, \zeta) r^*(\xi-x', \zeta-y') d\xi d\zeta
\]

\[
U_{RS}(x', y') = r(x', y') \otimes s(x', y') \otimes \delta(x', y' - 2y_0) = \iint r(\xi, \zeta) r^*(\xi-x', \zeta-y' + 2y_0) d\xi d\zeta
\]

\[
U_{SR}(x', y') = s(x', y') \otimes r(x', y') \otimes \delta(x', y' + 2y_0) = \iint s(\xi, \zeta) s^*(\xi-x', \zeta-y' - 2y_0) d\xi d\zeta
\]

The first two terms in Eq. 2-12 are the autocorrelation of the inputs \( s(x, y) \) and \( r(x, y) \). Both are centered at the optical axis, and are together called the dc component [although they have also non-dc content]. These terms carry no useful information: any image always auto- correlates with itself.

The latter two components in Eq. 2-12, called the correlation peaks, are symmetrically positioned along the optical axis, each separated from the optical axis by the scene-reference separation, and consist of the cross-correlation of the inputs \( s(x,y) \) and \( r(x,y) \). Existence of these correlation peaks indicates the presence and the location of the target within the scene.
Fig. 2-7 provides an example: Fig. 2-7a displays an input plane of a disk and a rectangle, and the output amplitude plane is shown in Fig. 2-7b.

Thus the JTC output is different from the matched-filter correlator output in that there are three components, one of which, the dc component, is uninteresting, and of the remaining two, one is constitute redundant because there is a pair of symmetrical correlation peaks for every matching instant of reference in the scene. For example, with a filter-based correlator, only the peak centered at pixels \((0,30)\) in Fig. 2-7b would appear. These multiple output peaks can be a significant problem when more than one instances of the reference is present in the scene, that is in the multi-object detection problem.

In addition to this, there is a limit to the input-plane scene-reference separation \(2y_0\): since, for example, the \(y\)-dimension of \(s(x, y) \odot r(x, y)\) is \(y_r + y_s\), to avoid overlap between the
correlation peaks and the dc component, the scene-reference separation must follow the limit:

\[ 2y_0 > \max(y_r, y_s) + y_r/2 + y_s/2 \]  

(2-13)

Unfortunately, increasing the separation between target and reference increases the detector resolution requirement, so the separation needs to be kept to a minimum. Dividing the input plane in thirds and zeroing the central region is a reasonable compromise.

The above discussion illustrates that a JTC requires a space-bandwidth product twice that of an otherwise similar matched-filter correlator\(^{16}\), since not only it must be shared by the input and reference signals but must be reserved as a guard band. Still, more practical issues, specifically the absence of precise alignment requirement and no need for pre-fabricated filters, have better positioned joint-transform correlators for real-time applications. These advantages made possible real-time update JTC applications\(^{17,18,19,20}\) with the use of spatial-light modulators (SLMs) and TV-cameras such as vidicon or charge-coupled devices (CCDs).

It should be emphasized, however, that the classical JTC produces at the correlation peak, to an extent of a simple proportionality, the same function as the matched-filter, that is the cross-correlation, and thus their actual performance for a given set of scene and reference is identical.
2.3 **Joint Transform Correlator**

*(i) Similar Inputs, No Noise*

If the scene \( s(x,y) \) and reference \( r(x,y) \) signals are identical, their spectra have the same amplitude \( (R(\nu_x, \nu_y) = S(\nu_x, \nu_y)) \) and, more important, phase \( (\phi_R(\nu_x, \nu_y) = \phi_S(\nu_x, \nu_y)) \). Then \( f_{\text{CL}}(\nu_x, \nu_y) \) reduces to:

\[
f_{\text{CL}}(\nu_x, \nu_y)_{R=S} = 2R^2(\nu_x, \nu_y) \cdot [1 + \cos(2\pi 2v_0 \nu_y)]
\]  

(2-14)

Fig. 2-8 provides an example with two similar disks. Fig. 2-8a displays the input plane transmissivity and fig. 2-8b the corresponding filter plane screen caption.

![Figure 2-8](image)

*Figure 2-8: Joint-transform correlation examples with two similar disks
(a) input plane transmissivity, left, and (b) filter plane, right.*

The JTC filter transmissivity, which is assumed linearly proportional to \( f_{\text{CL}}(\nu_x, \nu_y) \), consists of the reference power spectrum multiplied by a sinusoidal grating. This grating can be viewed as is the scene-reference interference; the spacing and orientation of the grating is associated with the relative separation between the target and the scene through the Fourier transform relationship.
A Fourier analysis of the JTC operation can be described as follows: The input plane (Fig. 2-7a) results from a convolution of the disk function \( r(x,y) \) with a pair of delta functions separated by \( 2y_0 \), \( \delta(y+y_0)+\delta(y-y_0) \). The field incident to the filter plane is the Fourier transform of such a convolution, which therefore is the product\(^{(1)} \) of their respective transforms, that is the reference spectrum \( R(v_x,v_y) \) multiplied by the function \( 2\cos(2\pi y_0 v_y) \).

The grating of Fig. 2-7b is the incident to the filter field squared, e.g., the reference power spectrum \( (R^2(v_x,v_y)) \) acting as an envelope function on the rectified grating \( 2[1+\cos(2\pi y_0 v_y)] \) (see Eq. 2-12). We note here the spatial frequency doubling: the grating spatial frequency becomes \( 2y_0 \). In addition, we observe in Fig. 2-7b that the JTC filter transmissivity is dominated by the on-the axis, low spatial frequency spectral components: the spatial frequencies away from the optical axis drop quickly, following the envelope \( R^2(v_x,v_y) \). As a result, only a few fringes of the interference pattern are visible.

Fig. 2-9 illustrates this operation: 2-9a shows a typical power spectrum from a reference image \( r(x,y) \), 2-9b an interference pattern similar to the \( 1+\cos(2\pi y_0 v_y) \), and 2-9c shows their product, as a result of the power spectrum acting as an envelope on the interference pattern.

![Figure 2-9: JTC frequency plane operation](image)

(a) reference power spectrum, left, (b) interference pattern, center, and (c) the resulting field, right.
It is evident that if the grating envelope $R^2(\nu_x, \nu_y)$ is removed then the grating becomes uniform; its inverse transform yields a pair of delta-function bright spots at the location of the correlation peaks. Thus grating enhancement results in more localized, sharper, and more intense correlation peaks. The need for such a grating-enhancement has been identified by a large number of researchers. A detailed discussion of similar operations is presented in the following sections.

The field at the output plane is the Fourier transform of this product, which is, again, the convolution of the respective transforms: the transform of the power spectrum $R^2(\nu_x, \nu_y)$ is the autocorrelation of the reference $r(x, y)$, while the transform of the grating $2 + 2 \cos(2\pi y_0 \nu_y)$ consists of the three-term function $2\delta(y) + [\delta(y+2y_0) + \delta(y-2y_0)]$. The first term produces the dc auto-correlation component, and the latter two the symmetrical correlation peaks. Again, this output field can be viewed as a result of diffraction from the filter-plane grating, producing the dc (0th order) and the correlation peaks as the ±1 diffraction orders. Fig 2-10 displays the output amplitude plane of such an operation.

![Figure 2-10: JTC output amplitude plane.](image-url)
2.3 Joint Transform Correlator

(ii) Inputs With Noise

The scene $s(x,y)$ signal is assumed to consist of the reference $r(x,y)$ and noise $n(x,y)$:

$$s(x,y) = r(x,y) + n(x,y),$$

(2-15)

whose spectrum is:

$$\mathfrak{F}[s(x,y)] = S(v_x, v_y) = R(v_x, v_y) \exp i \phi_R(v_x, v_y) + N(v_x, v_y) \exp i \phi_N(v_x, v_y)$$

(2-16)

Of the various classes of noise, two classes of particular interest are examined here: the first is the additive type of noise, arising for example, from in-device speckle and laser speckle. Such noise normally overlaps the target, is zero-mean white (though of the $\nu^t$ type for low spatial frequencies $\nu$) or filtered noise. The second is the background, clutter type of noise, which is not overlapping the target, and is a non-negative array. It is a common occurrence for real-time surveillance operations, where the presence of surround noise is unavoidable. Both types of noise can be present simultaneously, and typically any type of noise can be studied as a combination of these two classes.

Figure 2-11 gives an example of noise patterns which will be used extensively henceforth. Fig. 2-11a shows a 128x128 pixel zero-mean white noise array and Fig. 2-11b displays a 128x128 clutter array from a realistic desert scene. These arrays can be used as additive (Fig. 2-11a) and background noise (Fig. 2-11b).

When combined with a reference image of a 54x26 tank shown in an 128x128 array in Fig. 2-12a, the additive-noise scene image shown in Fig. 2-12b is produced by simply
adding the noise array shown in Fig. 2-11a and the tank array. To produce the clutter-noise scene image shown in Fig. 2-12c, the desert scene in Fig. 2-11b was used to fill the zeros in Fig. 2-12a.

![Noise image arrays](image)

*Figure 2-11: Noise image arrays (a) band-limited Gaussian noise, left, and (b) clutter noise, right.*

The input signal $g(x, y)$ consists of the reference image, $r(x, y-y_0)$ (Fig. 2-11a), and a scene image, $r(x, y+y_0) + n(x, y+y_0)$ (Fig. 2-12b or 2-12c). Their joint Fourier transform $G(\nu_x, \nu_y)$ is given by:

![Reference and scene images](image)

*Figure 2-12: (a) Reference image of a tank, left
(b) Scene image with additive noise, middle, and (c) Scene image with clutter noise, right.*
\[ G(v_x, v_y) = R(v_x, v_y) \cdot \exp \left[ i \varphi_R(v_x, v_y) + 2\pi y_0 v_y \right] + \]

The filter plane transmissivity \( f_{CL}(v_x, v_y) = |G(v_x, v_y)|^2 \) is therefore:

\[
2R^2(v_x, v_y) + N^2(v_x, v_y) + \\
\left[ R(v_x, v_y) \exp i\varphi_R(v_x, v_y) \right] \cdot \left[ N(v_x, v_y) \exp -i\varphi_N(v_x, v_y) \right] + \\
\left[ R(v_x, v_y) \exp -i\varphi_R(v_x, v_y) \right] \cdot \left[ N(v_x, v_y) \exp i\varphi_N(v_x, v_y) \right] + \\
\left[ R(v_x, v_y) \exp i\varphi_R(v_x, v_y) \right] \cdot \left[ R(v_x, v_y) \exp -i\varphi_R(v_x, v_y) \right] \cdot \exp(-i2\pi y_0 v_y) + \\
\left[ R(v_x, v_y) \exp -i\varphi_R(v_x, v_y) \right] \cdot \left[ N(v_x, v_y) \exp -i\varphi_N(v_x, v_y) \right] \cdot \exp(-i2\pi y_0 v_y) + \\
\left[ R(v_x, v_y) \exp -i\varphi_R(v_x, v_y) \right] \cdot \left[ R(v_x, v_y) \exp i\varphi_R(v_x, v_y) \right] \cdot \exp(i2\pi y_0 v_y) + \\
\left[ R(v_x, v_y) \exp -i\varphi_R(v_x, v_y) \right] \cdot \left[ N(v_x, v_y) \exp i\varphi_N(v_x, v_y) \right] \cdot \exp(i2\pi y_0 v_y) + \\
\]

or

\[
f_{CL}(v_x, v_y) = N^2(v_x, v_y) + 2R^2(v_x, v_y)[(1 + \cos 2\pi y_0 v_y)] + 2\text{Real}[R(v_x, v_y)N^*(v_x, v_y)(1 + \cos 2\pi y_0 v_y)]
\]

(2-18)

Comparing Eq. 2-18 to Eq. 2-10, there is a number of additional terms: the noise power spectrum,

\[
N^2(v_x, v_y), \quad (2-19)
\]

which is centered on the dc overlapping with the autocorrelation terms \( 2R^2(v_x, v_y) \); the noise-reference cross-correlation term, which is also centered on the optical axis; and the noise-reference correlation term which is of particular interest as it is produced on both the correlation spots overlapping with the correlation peaks.
The statistical properties of the noise \( n(x,y) \), and as a result, of the noise spectrum \( N(\nu_x, \nu_y) \), depend largely on the class of the noise. For example, a zero-mean white noise, (or its approximation, a band-limited Gaussian filtered speckle noise), like the array shown in Fig. 2-11a has a spectrum whose dc component is zero. On the contrary, the background noise array shown in Fig. 2-11b has a strong dc component and spectrum components comparable to the reference, particularly if similar objects are in the background. In addition, such an array, being non-negative, carries a strong dc component, which largely affect the dc component of its spectrum.

For the case of a wide-sense stationary, spatially invariant zero-mean additive white noise \( n_A(x,y) \), its spectral amplitude \( N_A(\nu_x, \nu_y) \) is also zero-mean over a wide spatial frequency bandwidth as a result of linear filtering of a zero-mean Gaussian random process\textsuperscript{[21],[22]}. Thus the ensemble average over the terms of Eqs. 2-20 and 2-21 is nearly zero, which means that the contribution of the noise to the correlation peak is very small; a classical correlation peak is similar to an auto-correlation peak resulting from a no-noise similar input\textsuperscript{[23],[24]}.

The optimality of matched-filter to optimum pulse detection in additive noise has long been established in one-dimensional electronic signals\textsuperscript{[25]} as well as in two-dimensional optical signals\textsuperscript{[26]}.

For the case of the disjoint background noise \( n_B(x,y) \), its spectral amplitude \( N_B(\nu_x, \nu_y) \) has strong non-zero dc component; the ensemble average over the terms of Eqs. 2-20 and 2-21 is comparable (or even higher than) to that of the cross-correlation. Thus the noise
contribution to the correlation peak is significant, largely degrading it, depending on the dc component (illumination level) of the background noise array\cite{27,28}.

An example of such correlation peaks is furnished in Figs. 2-13 and 2-14: the 128x128 pixel tank image of Fig. 2-12a serves as the reference, which is correlated with (a) a no-noise similar image, to produce a “clean” correlation peak, shown in Fig. 2-13a, (b) the zero-mean additive noise scene shown in 2-12b, to produce the correlation peak shown in Fig. 2-13b, and (c) the clutter noise scene shown in 2-12c to produce the correlation peak shown in Fig. 2-13c. Fig. 2-14 provides screen captures for the surface plots shown in Fig. 2-13. In all cases the center-to-center separation between scene and reference was 256 pixels, and the input plane array was embedded in a zero-filled 256x1024 array. The arrays illustrated in Figs. 2-13 and 2-14 correspond to a selected 200x200 intensity array centered around each correlation peak.

Figure 2-13 : Surface intensity plots, classical correlation peaks with tanks (a) no noise, left (b) additive noise, middle, and (c) clutter noise, right.
These results illustrate that the classical correlator performance with the Gaussian additive noise is similar to that of a no-noise similar input, since the correlation peak with additive noise (Figs. 2-13b and 2-14b) is very similar to the correlation peak with the no-noise input (Figs. 2-13a and 2-14a). On the contrary, the classical correlator performance with the clutter noise is significantly degraded: the correlator fails to identify the tank in clutter, since there is no distinct correlation peak.

Several metrics have been developed to establish quantitative comparison standards between various correlators. The simplest metric is the correlation peak intensity, which is simply the maximum value of an intensity array centered around the correlation peak. The peak sharpness, which is in many cases a more interesting quantity, and the relative value of the correlation peak compared to the noise level, that is the surrounding pixels, is expressed by two metrics, the signal-to-noise ratio[29][30] (SNR) and the peak-to-noise ratio[31] (PNR).
The SNR of a distribution \( k_i \) measures the peak height \( (k_{\text{max}}) \) divided by the root mean square of the distribution outside the full-width-at-half-maximum (FWHM) of the peak, and is defined as:

\[
\text{SNR} = \frac{k_{\text{max}}}{\left[ \frac{1}{n_T} \sum_{i=1}^{n_T} k_i^2 \left( < k_{\text{max}}^2 \right) \right]^{1/2}}
\]  

(2-22)

where \( n_T \) is the number of the points outside the FWHM of the peak.

The PNR, which was originally introduced as peak-to-correlation energy (primed) in Eq. 8 of Ref. [29], measures the peak height divided by the root mean square of the full distribution excluding the peak value \( k'_{\text{max}} \), after the average dc level has been subtracted from the distribution, and is defined as:

\[
\text{PNR} = \frac{k'_{\text{max}}}{\left( \frac{1}{n-1} \sum_{i=1}^{n-1} k_i' \right)^{1/2}}, \quad k'_i \neq k'_{\text{max}}
\]  

(2-23)

where the primed quantities \( k'_i \) indicate values after the subtraction of the average level.

The values of both SNR and PNR depend the size of the array around the correlation peak and on whether the array is one of amplitude or intensity. Amplitude arrays, 200x200 pixel size will be used henceforward. The results for the correlation peak arrays of Figs. 2-13 are in Table 2-1.

<table>
<thead>
<tr>
<th>type of scene image</th>
<th>PNR</th>
<th>SNR</th>
<th>Peak Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>clean tank, no noise</td>
<td>11.47</td>
<td>15.22</td>
<td>0.19</td>
</tr>
<tr>
<td>tank in additive noise</td>
<td>10.91</td>
<td>12.6</td>
<td>0.18</td>
</tr>
<tr>
<td>tank in clutter noise</td>
<td>1.83</td>
<td>5.7</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Table 2-1: Classical correlator performance metrics for various noise types.
2.3 JOINT TRANSFORM CORRELATOR

(III) DISSIMILAR INPUTS

Correlators are often presented with scenes containing objects similar to the reference; depending on the requirements, one would be interested the correlator to provide a negative identification (no correlation peak) or positive identification. The former is known for high discrimination ability, and the latter for low discrimination ability. The discrimination ability parameter measures the capacity of a correlator to distinguish the target among different objects. A high autocorrelation peak is desired with as low cross-correlation peak as possible. A discrimination ability metric\(^{[30]}\) is the quotient between the auto-correlation peak intensity to the cross-correlation.

\[
2\text{Real}[R(v_x, v_y)S^*(v_x, v_y)(\cos 2\pi 2v_0 v_y)] \tag{2-24}
\]

![Figure 2-15](image)

\(\text{Figure 2-15 : (a) Reference image of a tank, left, and (b) Scene image of an armored vehicle, right.}\)

In general, the correlation peak produced by a set of dissimilar inputs like those shown in Fig. 2-15, is the realization of their cross-correlation function, that is the inverse Fourier transform of the spectral product.
The Fourier transforms for similarly-sized objects contain the spectral information in reverse order of the feature size; a large feature in the input plane, for example the periphery, is decomposed as low spatial frequency, whereas a small feature, for example the turret, is decomposed as high spatial frequency component. Thus for the images shown in Fig. 2-15 their corresponding spectra are very similar for the low spatial frequencies (because of their size similarity) and the differences only appear at the high spatial frequency content (where the feature details are accounted for). It is the intense low spatial frequencies that dominate a joint power spectrum, however, as the weak high spatial frequency content quickly vanishes.

Figures 2-16 and 2-17 demonstrate this concept: Fig. 2-16a displays a slice cut through the $y$-axis of the reference-only power spectrum, Fig. 2-16b displays the joint power spectrum of two similar tanks, and Fig. 2-16c displays the joint power spectrum of the tank and the armored vehicle.

---

Figure 2-16: Power spectra plots (a) power spectrum of a tank only, left (b) joint power spectrum of two similar tanks, middle, and (c) joint power spectrum of a tank and an armored vehicle, right.
Figure 2-17 repeats this with screen captures. The similarity between these joint power spectra gratings reflects the similarity of the diffraction output, that is, the correlation peaks.

Figure 2-17: Power spectra screen captures (a) power spectrum of a tank only, left (b) joint power spectrum of two similar tanks, middle, and (c) joint power spectrum of a tank and an armored vehicle, right.

Reflecting the similarity in their spectra, the cross-correlation peak (see Figs. 2-18b and 2-19b) between the set of dissimilar images shown in Fig. 2-15 is very similar to the no-noise auto-correlation peak of the tank (see Figs. 2-18a and 2-19a).

Figure 2-18: Surface intensity plots, classical correlation peaks with tanks (a) similar images, left (b) tank versus an armored vehicle, right.
As the metrics comparison of Table 2-2 demonstrates, the classical correlator has a low discrimination capability, in the sense that for dissimilar objects the corresponding correlation peaks have intensity and sharpness metrics within the same order of magnitude. In such a correlator the false alarm rate, that is the possibility that an object will be wrongly identified as a reference instance, is high.

<table>
<thead>
<tr>
<th>type of scene image</th>
<th>PNR</th>
<th>SNR</th>
<th>Peak Intensity</th>
<th>Discrimination Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>clean tank, no noise</td>
<td>11.47</td>
<td>15.22</td>
<td>0.19</td>
<td>1.73</td>
</tr>
<tr>
<td>armored vehicle</td>
<td>6.78</td>
<td>13.98</td>
<td>0.11</td>
<td></td>
</tr>
</tbody>
</table>

*Table 2-2: Classical correlator performance metrics for dissimilar images.*
2.4 **Nonlinear JTC**

(i) **Hybrid Implementation**

The previous analysis identified some of the classical correlator performance shortcomings, in particular, its failure to identify objects in clutter array\cite{27,28} and poor discrimination ability\cite{32}. Only for the problem of additive, zero-mean Gaussian noise the classical correlator has performance similar to the clean inputs. The classical JTC, even with a no-noise scene produces a correlation peak which is, in general, of relatively low intensity and broad\cite{33}. In addition, owing to the fact that a cross-correlation peak has base dimensions typically twice that of the corresponding images, in a multi-object detection problem the space-bandwidth limitations become significant.

Overall, this is not a satisfactory performance, and there is a lot of room for improvement. An improved output should have the following features: (a) sharper correlation peak when the two input signals are identical, (b) no peak (or significantly reduced) when the two inputs are different, and (c) satisfactory performance with target embedded in clutter.

In addition to these performance concerns, implementation of the joint transform (or a filter-based) correlator in real-time dictates that the use of film as the recording medium be replaced by a real-time device like a vidicon or charge-coupled device (CCD) (Fig. 2-20); the filter function is then displayed by a spatial-light modulator (SLM)\cite{18,19,20,34,35,36} which can be electronically or optically addressed. The SLM display is subsequently illuminated by a laser beam to produce the output field. Such a correlator becomes a hybrid opto-electronic device.
The presence of these additional electronics, however, compromises the parallel processing speed and simplicity advantages of an all-optical device and imposes a series of additional limitations. The Fourier-plane dynamic range requirements may not be well served by a SLM with limited contrast ratio, notwithstanding the additional phase factors introduced by virtually all current SLMs. For example, the dynamic range of intensity at the Fourier plane for realistic imagery typically is six (or more) orders of magnitude; the contrast ratio in current SLMs is limited to a couple of orders of magnitude. In fact, the limitations of SLMs, which were and still are predominantly binary, motivated the binary joint-transform correlator\[37],[38].

On the other hand, the presence of these additional electronics introduces the possibility of subsequent nonlinear spatial frequency plane processing via the introduction of a serial processor in tandem between the CCD and the display SLM (Fig. 2-21). Thus, various forms of nonlinear spatial frequency processing have been studied and designed\[39],[40],[41],[42],[43] intending to emphasize via several contrast-improvement algorithms

\begin{figure}
\centering
\includegraphics[width=\textwidth]{Figure2-20.png}
\caption{Implementation of the hybrid opto-electronic Joint Transform Correlator.}
\end{figure}
the cross-correlation grating at the high spatial frequencies, where typically it is weak, and de-emphasize the intense self-correlation.

All these techniques have been shown to be superior to classical correlator in areas such as correlation peak intensity, peak localization, and discrimination sensitivity both theoretically\cite{44,45,46} and experimentally\cite{47}. However, the introduction of a sequential processor to implement these nonlinear algorithms further compromises the speed and simplicity of an all-optical correlator. In addition, the output of the digitally processed power spectrum has to be displayed via an SLM; this subsequently limits the nonlinearities that actually can be implemented. Nevertheless, the nonlinear JTC has been well established as a tool in pattern recognition.

Figure 2-21: Implementation of the nonlinear Joint Transform Correlator.
2.4 Nonlinear JTC

(II) Examples of Established Techniques

The objective of a nonlinear operation is to reduce the effect of the intense dc component, as it is suggested in section 2.3(i). The simplest of all nonlinear JTCs involves the use of a small opaque mask or “dc block” placed at the optical axis to prevent over-saturation by the high intensity at the center of the joint power spectrum (JPS). This technique allows for the low-intensity fringes to be detected over a broader area, but is signal-dependent and therefore not easily implemented. The dc block technique is a linear filter however; in a nonlinear operation, examples of which follow, modification of the spatial frequencies occurs.

The global median thresholding binary JTC\textsuperscript{[48],[49]} has been implemented by the necessity of the existing hardware, that is that first SLMs were binary. The processed Fourier plane output joint power spectrum (JPS) has the form of:

\[
JPS_{\text{out}} = \text{sgn}[S^2(v_x,v_y) + R^2(v_x,v_y) - \text{Global Median}].
\] (2-25)

This technique helps improve the high spectral content significantly, but backgrounds with high energy contents relative to the energy on target pose a challenge to the constant threshold since the modulation depth of the fringes diminishes with increasing energy in the lower spatial frequencies.

The localized-median binary JTC\textsuperscript{[50],[51],[52]} utilizes a binarization thresholding level based not only on the global median, but on a number of neighboring spectral values. The technique, which is also known as Fourier-plane windowing\textsuperscript{[53]}, involves masking (convolving) the Fourier plane with a transmissivity window to apply a spatially adaptive threshold only:
\[ JPS_{out} = \text{sgn}\{S^2(v_x, v_y) + R^2(v_x, v_y) - \text{Median (of local values only)}\}. \] (2-26)

Figure 2-22 offers an example of binarized joint power spectra. Fig. 2-22a displays a JPS binarized by the global thresholding level, and Fig. 2-22b displays a JPS binarized by a spatially adaptive threshold.

The intensity compensation JTC\(^{[54]}\) divides the Fourier plane output by the reference power spectrum:

\[ JPS_{out} = \frac{[S(v_x, v_y) + R(v_x, v_y)]^2}{|R^2(v_x, v_y)|^2}. \] (2-27)

Other techniques include the fringe-adjusted JTC\(^{[55],[56],[57]}\) the input plane varying illumination JTC\(^{[58]}\), the multiobject preprocessing JTC\(^{[59]}\), the error-diffusion binarization JTC\(^{[60]}\), and the time modulation JTC\(^{[61]}\), which applies subtraction and addition of partial filter plane information to rid of the dc component. The Wiener filter JTC\(^{[62]}\) divides the Fourier plane output by the reference and noise power spectrum:

\[ JPS_{out} = \frac{[S(v_x, v_y) + R(v_x, v_y)]^2}{|R^2(v_x, v_y)|^2 + |N(v_x, v_y)|^2}. \] (2-28)
Other recent developments include the Wavelet-transform based JTC\textsuperscript{[63],[64]}, which employs circular harmonics expansion to compensate for image rotational invariance. Multiple circular harmonic filters\textsuperscript{[65],[66],[67]} employ relative location as rotation-, shift-, and intensity- invariant features for pattern recognition.

In all these nonlinear algorithms aim to selectively amplify the low-intensity high frequency spectral content. The logarithmic nonlinearity and the fractional power nonlinearity\textsuperscript{[68],[69]} can be viewed as applications of compressive, nonlinear mapping of the joint power energy:

\[
JPS_{\text{out}} = \left[ |S(v_x, v_y) + R(v_x, v_y)|^2 \right]^{\mu}
\]

(2-29)
is a general tool for studying such nonlinearities. The classical correlator is achieved with \( \mu=1 \). To achieve high frequency weighting \( \mu < 1 \).

2. Optical Correlation

Examples of Established Techniques 39

Figure 2-23: Nonlinear mapping of the joint power spectrum (a) nonlinear mapping curves, top (b) power spectrum, left, bottom, and (c) fringe-enhanced joint power spectrum, right, bottom.
Depending on the value of the fractional power $\mu$, such a nonlinearity can be either an implementation of the phase-only correlation ($\mu=\frac{1}{2}$), or the phase-extraction correlation ($\mu=0$). The case $\mu=\frac{1}{2}$ produces a correlator with performance equivalent to that of the phase-only filter\textsuperscript{[60]}, defined as:

\[
f_{\text{PO}}(v_x, v_y) = \exp(-i\varphi_R(v_x, v_y)) = \frac{R^*(v_x, v_y)}{R(v_x, v_y)} \tag{2-30}
\]

The phase-only filter output is:

\[
R(v_x, v_y) \cdot \exp(i\varphi_R(v_x, v_y)) \cdot \exp(-i\varphi_R(v_x, v_y)) = R(v_x, v_y) \tag{2-31}
\]

Rewriting (Eq. 2-29) for $\mu=\frac{1}{2}$ we obtain the transfer function of the phase-only JTC as:

\[
f_{\text{PO}}(v_x, v_y) \propto |R(v_x, v_y) + S(v_x, v_y)|^1 \tag{2-32}
\]

and for the case of scene being identical to the reference:

\[
f_{\text{PO}}(v_x, v_y) \propto R(v_x, v_y). \tag{2-33}
\]

Thus the phase-only correlator has a transfer function proportional to the reference spectral amplitude; such a frequency-plane function drop rate for increasing spatial frequencies is half the rate of the classical correlator transfer function, which is proportional to the spectral energy.

The case $\mu=0$ produces a correlator with performance equivalent to that of the inverse filter, defined as\textsuperscript{[70],[71]}:

\[
f_{\text{IN}}(v_x, v_y) = \frac{R^*(v_x, v_y)}{|R(v_x, v_y)|^2} = \frac{\exp(-i\varphi_R(v_x, v_y))}{R(v_x, v_y)} \tag{2-34}
\]

The inverse filter output is:
\[ R(v_x, v_y) \cdot \exp(i\varphi_R(v_x, v_y)) \cdot \frac{\exp(-i\varphi_R(v_x, v_y))}{R(v_x, v_y)} = \text{constant} \quad (2-35) \]

Rewriting (Eq. 2-29) for \( \mu=0 \) we obtain the transfer function of the \( \mu=0 \) nonlinear JTC as:

\[ f_{\text{IN}}(v_x, v_y) = |R(v_x, v_y) + S(v_x, v_y)|^0 \quad (2-36) \]

and for the case of scene being identical to the reference:

\[ f_{\text{IN}}(v_x, v_y) = 2\delta(x,y) + [\delta(x, y+2y_0) + \delta(x, y-2y_0)]. \quad (2-37) \]

which is the interference pattern with the power spectrum envelope removed. Thus the phase-extraction correlator\(^{[12]}\) has a transfer function with constant amplitude with spatial frequencies.

Table 2-3 furnishes a comparison between standard filter-based correlators and equivalent nonlinear JTCs for clean inputs.

<table>
<thead>
<tr>
<th>joint transform correlator</th>
<th>nonlinearity</th>
<th>equivalent filter-based correlator</th>
</tr>
</thead>
<tbody>
<tr>
<td>classical JTC (</td>
<td>R(v_x, v_y) + S(v_x, v_y)</td>
<td>^2 )</td>
</tr>
<tr>
<td>phase-only JTC ( (</td>
<td>R(v_x, v_y) + S(v_x, v_y)</td>
<td>^2)^{1/2} )</td>
</tr>
<tr>
<td>phase-extraction JTC ( (</td>
<td>R(v_x, v_y) + S(v_x, v_y)</td>
<td>^2)^0 )</td>
</tr>
</tbody>
</table>

Table 2-3: Comparison of standard filter-based correlators and nonlinear JTCs.
Based on the phase-only filter, a large number of nonlinear filter algorithms has been developed\cite{27,72,73}, one of which is the Yaroslavsky filter

\[ f_{YA}(v_x, v_y) = \frac{R^*(v_x, v_y)}{\langle S(v_x, v_y) - R(v_x, v_y) \rangle^2} \quad (2-38) \]

claims optimality by treating the additional information to be rejected as noise. Here the <> symbol stands for average over realizations of sensor noise. This filter is used for comparison purposes along with the classical-matched, phase-only and inverse filters in the following chapters.
2.5 References


References 45

2. Optical Correlation


2. Optical Correlation

References 46


