# Asymmetric Correlation: A Noise Robust Similarity Measure for Template Matching

Elhanan Elboher and Michael Werman

***Abstract*— We present an efficient and noise robust template matching method based on *asymmetric correlation* (ASC). The ASC similarity function is invariant to affine illumination changes and robust to extreme noise. It correlates the given non-normalized template with a normalized version of each image window in the frequency domain. We show that this asymmetric normalization is more robust to noise than other cross correlation variants, such as the correlation coefficient. Direct computation of ASC is very slow, as a DFT needs to be calculated for each image window independently. To make the template matching efficient, we develop a much faster algorithm, which carries out a prediction step in linear time and then computes DFTs for only a few promising candidate windows. We extend the proposed template matching scheme to deal with partial occlusion and spatially varying light change. Experimental results demonstrate the robustness of the proposed ASC similarity measure compared to state-of-the-art template matching methods.**

***Index Terms*— Asymmetric correlation, cross correlation, noise robust similarity, phase correlation, template matching.**

## I. Introduction

Template matching is an important tool in image processing and computer vision. A *template*, a relatively small image patch, is sought within a larger reference image. Finding such correspondences is useful for many computer vision tasks such as object tracking [1], 3D reconstruction [2], image editing [3] and medical imaging [4]. For real world applications, template matching methods need to be both efficient and robust to various visual distortions.

In this paper we address the problem of template matching in the presence of extreme noise where common template matching methods fail. We propose an efficient method based on a new similarity function, *asymmetric correlation* (ASC). The proposed method is also useful when the noise distortion is combined with poor illumination and other photometric distortions.

The proposed ASC similarity is a robust cross correlation variant. As described below (Section III-B), most cross correlation variants use one of two common normalizations. In *symmetric normalization* both the compared vectors are normalized the same way, either in the spatial domain (in normalized cross correlation) or in the frequency domain (in phase correlation, see Section III-B). In *filter design* only the template is modified and then is correlated with the reference image.

The ASC similarity measure takes a different approach. Here, the image windows are normalized independently in the frequency domain and then correlated with the unmodified template $q$.

In Section VI we analyze the effect of this asymmetric normalization, and show that it makes ASC highly robust to noise. For the additive Gaussian noise model, ASC is proven to be more robust than the matched filter [5] and the popular normalized cross correlation, NCC, which is also known as the correlation coefficient or zero-mean NCC. The robustness of ASC is demonstrated experimentally in Section V for various noise models and additional photometric distortions.

### A. Efficient Computation

As shown by Lewis [6], most cross correlation variants can be computed efficiently using the fast Fourier transform algorithm (FFT) and a few additional operations. However, this precludes fast computation of ASC since the Fourier spectrum of each image window is different. By direct computation, an FFT should be calculated for each image window independently. Given a $N \times N$ iamge and a $M \times M$ template, the time complexity of this process is $O(N^2M^2 log M)$.

In Section IV we propose a much faster algorithm that operates in two steps. The first step computes partial ASC scores and predicts the most promising candidate windows. To do so, we developed a linear time projection scheme on a single Fourier basis vector using the integral image technique [7]. This scheme is as efficient as previous sliding DFT methods [8] but is simpler, more exact and numerically stable (see Section IV-A).

In the second step an exact ASC similarity is computed using FFT but only for a few promising candidates. The proposed algorithm, implemented in C++, is more than 700 times faster than the exact ASC computation with an almost identical success rate (see Section V-C).

### B. Extensions

The ASC similarity is robust to noise and invariant to global affine illumination change. In Section IV-E we extend the ASC based template matching scheme to also deal with partial occlusion and spatially varying illumination. An experimental evaluation is presented in Section V-B.

The rest of this paper is organized as follows. In Section II we review related work. Section III defines the proposed asymmetric correlation (ASC) and compares it with common cross

Manuscript received September 2, 2012; revised January 31, 2013; accepted March 22, 2013. Date of publication April 12, 2013; date of current version May 24, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hitoshi Kiya.

The authors are with the School of Computer Science, The Hebrew University of Jerusalem, Jerusalem 91904, Israel (e-mail: elhanan.elboher@mail.huji.ac.il; werman@cs.huji.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2257811

correlation variants. Section IV presents a fast algorithm for ASC template matching and an extension for partial occlusion. Section V presents our experimental results. In Section VI we analyze the noise robustness of ASC mathematically using an additive white Gaussian noise model. Section VII summarizes our work.

## II. Related Work

We briefly review the main similarity and distance functions used for template matching, with a focus on correlation related functions. More complete reviews can be found in [5], [9], [10].

### A. SSD and SAD

The sum of squared differences (SSD) is probably the most popular distance measure for many applications, including template matching, due its nice mathematical properties and very efficient computational schemes (see review and performance evaluation in [11]). However, SSD is very sensitive to noise and illumination change. Another popular distance is the sum of absolute differences (SAD) which is robust to outlier ('salt-pepper') noise and partial occlusion. However, SAD is also sensitive to additive noise and illumination change.

### B. NCC and Other Cross Correlation Variants

One very useful similarity is the normalized cross correlation (NCC). NCC is invariant to affine illumination change, and more robust to noise than SSD and SAD. Several efficient methods have been proposed for template matching using NCC [12]–[15]. Most of these methods compute a partial similarity score for each candidate image window and use the lower and upper bounds to eliminate or skip irrelevant candidates.

Although NCC is more robust to noise than SAD and SSD, it is sensitive to outliers and partial occlusion. Some NCC related methods overcome such distortions using gradient information, either by correlating the gradient magnitudes (e.g. [16]) or the partial derivatives [17]. This method, called the *matching function* (MF), was shown in [18] to outperform many other methods in the presence of spatially varying illumination and real noise. An experimental comparison to the MF method is presented in Sections V-A and V-B.

Other cross correlation variants include the phase correlation (PC) similarity [19] and correlation filters such as the phase only filter (POF) [20], the synthetic discriminant function (SDF) [21], the minimum average correlation energy filter (MACE) [22], the maximum average correlation height filter (MACH) [23] and optimal tradeoff filters [5]. The relation between ASC and these cross correlation variants is discussed in detail in Section III-B.

### C. Robust Metrics and Additional Methods

Another important family is composed of robust metrics, which are less affected by outlier noise than cross correlation related methods. In addition to SAD, these include Hamming based distances [24], [25] and M-estimators such as Huber's estimator, Tukey's estimator or trimmed SAD. Fast template matching using M-estimators was suggested in [26] using a hierarchical approach and in [27] by decomposition to a sum of cross correlations computed by multiple FFTs.

Other robust similarity functions were proposed in [28] that used voting scheme to overcome outliers and partial occlusions. Another method described in [29] uses higher order statistics to overcome strong Gaussian noise. This method, called $M4_c$, is computed efficiently using a sum of cross correlations, similar to [27]. A comparison with $M4_c$ is presented in Section V-A.

A recent method that has attracted attention is matching by tone mapping (MTM) [30]. This method is invariant to an arbitrary intensity transformation ('tone mapping') between the query template and the reference image. While ASC is only invariant to affine intensity transformation, it is shown experimentally in Section V to be more robust to noise than MTM.

## III. Asymmetric Correlation (ASC) and Related Correlation Functions

### A. ASC Definition

Let $q$ be a 2D template (query) of size $M \times M$, and $w$ a candidate image window of the same size. For simplicity we consider an odd size $M = 2m + 1$, and determine the centers of the template and the image window as the origins $q_{00}, w_{00}$. In the origin, the cross correlation between $q$ and $w$ is the dot product

$$\begin{aligned} q^T w &= \sum_{i=-m}^{m} \sum_{j=-m}^{m} q_{ij} w_{ij} \\ &= \frac{1}{M^2} \sum_{u=-m}^{m} \sum_{v=-m}^{m} Q_{uv} W_{uv}^* \end{aligned} \tag{1}$$

where $Q_{uv}$ and $W_{uv}$ are the Fourier coefficients of $q$ and $w$ (the second equality follows from Parseval's theorem).

The asymmetric correlation (ASC) is defined by modifying the dot product as follows:

$$ASC(q, w) = c_q \sum_{(u,v)\neq(0,0)} Q_{uv} \frac{W_{uv}^*}{|W_{uv}|} \tag{2}$$

The zero frequency is ignored, and the amplitudes of the Fourier coefficients of all the other windows are normalized.

The total sum is normalized by $c_q = \left(\sum_{(u,v)\neq(0,0)} |Q_{uv}|\right)^{-1}$. This bounds ASC to be between $-1$ and 1 independent of the template size.

An equivalent definition of ASC can be expressed by the amplitude $|\cdot|$ and the phase $\theta(\cdot)$ of the Fourier coefficients,

$$ASC(q, w) = c_q \sum_{(u,v)\neq(0,0)} |Q_{uv}| \cos\Big(\theta(Q_{uv}) - \theta(W_{uv})\Big) \tag{3}$$

This polar form is useful for the analysis of ASC noise robustness (see Section VI). Due to the conjugate symmetry of the DFT of real vectors, Equation 3 does not include imaginary sine elements that cancel each other.

Due to the normalization by $c_q$, ASC is invariant to affine illumination changes of the template $q$. In addition, from

Equations 2 and 3 it follows that ASC is invariant to affine illumination changes of the reference image; i.e.,

$$ASC(q, aw + b) = ASC(q, w), \quad a > 0 \tag{4}$$

Adding a constant $b$ to $w$ changes only $W_{0,0}$ which is ignored by ASC. By Fourier linearity, the other coefficients $W_{uv}$ are multiplied by $a$. This does not affect ASC, which considers the normalized window coefficients (Equation 2). In terms of Equation 3, the phase $\theta(W_{uv})$ is unchanged.

### B. Comparison With Other Correlation Variants

In the following we review common cross correlation variants and highlight the basic differences between them and the proposed ASC. These help to account for the differences in performance between ASC and the other correlation related methods while looking for a small template in a large noisy image.

*1) Symmetric Normalization Variants:* Using the notations of $q$ and $w$, the normalized cross correlation (NCC) is defined as

$$NCC(q, w) = \left(\frac{q - \mu_q}{\sigma_q}\right)^T \left(\frac{w - \mu_w}{\sigma_w}\right) \tag{5}$$

where $\mu_q, \sigma_q$ and $\mu_w, \sigma_w$ are the means and standard deviations of the template $(q)$ and the window $(w)$. We refer to this as symmetric normalization, since the template and each image window are normalized the same way.

The phase correlation (PC) [19] is also based on a symmetric normalization:

$$PC(t, f) = \sum_{(u,v)\neq(0,0)} \frac{T_{uv} F^*_{uv}}{|T_{uv}||F_{uv}|} \tag{6}$$

where $T_{uv}$ and $F_{uv}$ are the Fourier coefficients of $t$ and $f$ (the notation change from $q, w$ to $t, f$ is explained below).

Unlike NCC and PC, the ASC similarity is based on asymmetric normalization. In Equation 2, the normalized window coefficients $W_{uv}/|W_{uv}|$ are correlated with the unnormalized template coefficients $Q_{uv}$. The rationale for this asymmetry is that normalizing $Q_{uv}$ removes important information about the sought object. On the other hand, when the reference image is noisy, normalizing $W_{uv}$ reduces the noise effect significantly. This non-intuitive observation is proved in Section VI based on the statistics of natural images.

There is a crucial difference between PC and ASC as regards template matching. The PC similarity has a sharp correlation peak, which makes it useful for an accurate global image registration [31], [32]. PC has also been used for block matching; i.e., alignment of two image patches that are known to be shifted by a few pixels, using a coarse to fine approach [2]. However, implementing PC for full search template matching is more complicated. Computing $PC(q, w)$ between the template $q$ and each image window $w$ is computationally expensive since the DFT has to be computed for each window $w$ independently. Given two images of different sizes, the most common solution is to expand the smaller image to the size of the larger one by zero padding. Thus in Equation 6 the notation $F$ for the DFT of the complete $N \times N$ reference image $f$, and $T$ for the DFT of $t$, the $N \times N$ expansion of the template $q$ by zero padding.

Since the template $q$ is typically much smaller than the reference image, the DFT $F$ is mostly determined by irrelevant data that do not come from window $w$ which best matches $q$. This reduces the detection rate of PC significantly (see Section V-A). An improvement is possible by dividing the image into smaller regions and computing a PC for each of them. This has an overhead of computing multiple DFTs for overlapping regions. Instead, the algorithm that we propose for ASC in Section IV can be modified for fast computation of $PC(q, w)$ without an additional overhead. The proposed algorithm for $PC(q, w)$ performs better than $PC(t, f)$ both on clean and noisy images and results in a very sharp and accurate peak. However, it is not as robust to noise as ASC (see Section V).

*2) Correlation Filters:* Given a single template $q$ (or multiple templates), a *correlation filter* $h_q$ can be designed and correlated with the reference image. The cross correlation $(h_q)^T w$ is then computed instead of $q^T w$ (see Equation 1). This may be useful where a specific criterion needs to be optimized; e.g. peak sharpness. It is also valuable in capturing the consensus of multiple training examples (templates) that represent the variation of an object class [5].

In this paper we focus on template matching based on a single example. One common filter that is based a single template is the *matched filter*, which is the template $q$ itself, or $Q$ in the frequency domain.[1] The matched filter is the linear correlation filter that is most robust to noise [5]. However, it has no sharp correlation peak and is therefore not as discriminating.

Another filter based on a single template is the phase only filter (POF) [20]. POF is defined by the normalized template's coefficients $Q_{uv}/|Q_{uv}|$. Due to this normalization POF has a sharp correlation peak. However, the normalization also makes POF sensitive to noise since high template frequencies are amplified. These frequencies contain unreliable information that varies even between slightly different instances of the same object. The proposed ASC similarity does exactly the opposite of POF: the noisy window coefficients are normalized, whereas the reliable template coefficients remain unchanged (Equation 2). For this reason ASC is both more robust to noise and more discriminating than POF.

Given multiple example templates, more advanced correlation filters can be designed such as SDF, MACE and MACH. An optimal tradeoff filter can also be defined that maximizes a linear combination of the criteria optimized by these filters. In the case of a single example, SDF and MACH reduce to the matched filter. The MACE filter reduces to an inverse filter defined by $Q_{uv}/|Q_{uv}|^2$. The optimal tradeoff between SDF and MACE filter reduces to $Q_{uv}/(\alpha+\beta|Q_{uv}|^2)$ where $\alpha$ and $\beta$ are the tradeoff parameters.

The SDF-MACE optimal tradeoff filter is more robust to noise than POF. However, as shown in Section V, optimal

[1] It is common to define the matched filter by the complex conjugate $Q^*$. This is accepted while inverting the order of the scalar product (Equation 1) from $q^T w$ to $w^T q$, which is equivalent.

tradeoff filters based on a single example also fail to find templates that are matched correctly by common similarities as NCC (see Section V-A).

## IV. Algorithm

A naive implementation of ASC based template matching is very slow since an FFT computation is required for each image window. Given a template $q$ of size $M \times M$ and a reference image $f$ of size $N \times N$, the time complexity is $O(N^2 M^2 \log M)$.

We present a much faster two step algorithm for finding the window $w$ which maximizes $ASC(q, w)$. The first step consists of an efficient prediction of promising candidate windows. In the second step ASC is computed exactly but only for these few candidates. In the following we describe the main steps of our algorithm. Then we extend the basic computation scheme to deal with partial occlusion and spatially varying illumination of the sought template.

### A. Fast Fourier Windows Projection (FFWP)

The core of the proposed algorithm is an efficient scheme that computes the Fourier coefficient (amplitude and phase) of a single frequency $W_{uv}$ for all image windows $w$. The proposed FFWP scheme has linear time complexity in the image size, $O(N^2)$, regardless of the template (or window) size.

Let $f(x, y)$ be the grayscale image values and $w$ an $M \times M$ window centered at $(x_0, y_0)$ (i.e., $w_{0,0} = f(x_0, y_0)$). For simplicity we use an odd size $M = 2m + 1$ (as in Equation 1 above). For each frequency $(u, v)$, $W_{uv}$ is the projection of $w$ on the $(u, v)$ basis vector:

$$\begin{aligned} W_{uv} &= \sum_{x=x_0-m}^{x_0+m} \sum_{y=y_0-m}^{y_0+m} f(x,y) e^{-\frac{2\pi i}{M}\left(u(x-x_0)+v(y-y_0)\right)} \\ &= e^{\frac{2\pi i}{M}(ux_0+vy_0)} \sum_{x=x_0-m}^{x_0+m} \sum_{y=y_0-m}^{y_0+m} f(x,y) e^{-\frac{2\pi i}{M}(ux+vy)} \end{aligned} \tag{7}$$

Each addend in the resulting sum is computed from a single pixel by its coordinates $x, y$ and its grayscale value $f(x, y)$. We use the integral image technique [7] to compute the sum of the addends in each window $w$. This is done for all the image windows in $O(N^2)$ operations regardless of window size. Then for each location $(x_0, y_0)$ we multiply the resulting sum by $\exp\left((2\pi i/M)(ux_0 + vy_0)\right)$.

The proposed FFWP scheme extends our previous work in the cosine transform domain [33]. It provides a simple and exact alternative to previous linear time techniques such as the sliding DFT (the common methods are reviewed in [8]). These techniques compute the Fourier coefficients of local windows using a recursion formula derived from the Fourier circular shift property [34]. Therefore they perform repeated multiplications of complex numbers, which is numerically unstable [8]. In addition, these techniques use windowing functions which affect the precision of the Fourier coefficients. On the other hand, FFWP computes the exact windows' Fourier coefficients without involving windowing functions. Our scheme is not recursive, and therefore it is numerically stable. An additional advantage of FFWP, which is not discussed in this paper, is that efficient computation can be performed in sampled locations. Given the integral image, one can compute the Fourier coefficients of any window independently, which is complicated when using recursive methods (see [8]).

### B. Prediction Using Partial ASC Similarity

The above FFWP scheme is used to compute partial ASC similarity scores on the largest magnitude template frequencies. Tipcally these frquencies are low, thus the prediction step is equivalent to full search in a coarse resolution reference image.

For each window $w$, the partial ASC score is

$$ASC^{(k)}(q, w) = c_q \sum_{(u,v)\in\mathcal{F}} Q_{uv} \frac{W^*_{uv}}{|W_{uv}|} \tag{8}$$

The set $\mathcal{F}$ contains the $k$ frequencies of the template with the largest amplitudes,[2] which makes the complexity of the prediction step $O(kN^2)$. $k$ is typically a small number (see below).

Fast template matching algorithms such as [12], [13], [15], [35] use partial scores to bound the full score and eliminate irrelevant candidates. However, for the ASC similarity, tight bounds require too many frequencies (large $k$) and the computation becomes inefficient. Therefore we do not use the partial scores for elimination of irrelevant candidates, but to *predict* the best candidates.

As demonstrated in Figure 1(c), the ASC similarity has a very sharp peak in the correct match. Since most of the similarity score is determined by the leading frequencies of the template, the partial scores $ASC^{(k)}$ of true matches are much higher than the partial scores of non matches even when $k$ is small. Good candidates for the correct match can therefore predicted when the peak of the partial score $ASC^{(k)}$ is sharp enough.

To avoid parameter tuning, we add the contribution of the next frequency (i.e. increase $k$ by 1) until only a small fraction $p$ (e.g. $p = 0.001$) remains of *promising candidates*. An image window is defined as a promising candidate if its partial score is at least $a = 0.8$ times the maximal partial value $\mathrm{ASC}^{(k)}_{max}$.[3] Usually 6–30 frequencies are computed. This mainly depends on the template structure. We limit $k$ to below 50 for the exceptional cases where we are left with a proportion larger than $p$ of promising candidates.

### C. Finding the Best Match

In the final step we compute the exact ASC similarity (Equation 2) for all the remaining promising candidates using

[2] $\mathcal{F}$ does not contain the zero frequency. In addition, due to Fourier conjugate symmetry, we skip one frequency from each pair $(u, v), (-u, -v)$.

[3] A straightforward implementation requires two passes over the image: one for finding $\mathrm{ASC}^{(k)}_{max}$ and another to compare it with the partial score of each window. For efficiency we make a single pass and compare the partial scores with the current maximum, which could be a slight under estimate of $\mathrm{ASC}^{(k)}_{max}$.

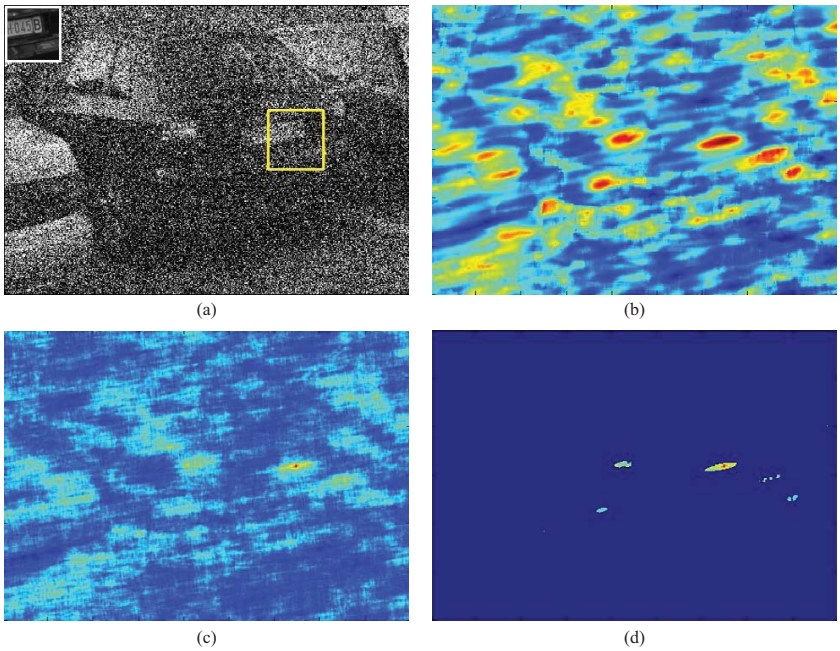



Fig. 1. Template matching by the proposed fast ASC scheme. **This figure is best viewed in color.** (a) A template taken from the 'Leuven' image is sought in the same image after adding strong Gaussian noise ($\sigma = 0.3$, or 76.5 out of 255). The original template is shown in the top left corner (white square). The correct match, which was found by ASC, is delineated by the lower right yellow square. Figures (b,c,d) show similarity maps scaled to [0, 1] where blue indicates low similarity and red indicates high similarity. (b) Partial ASC score provided by the prediction step of our fast algorithm (Section IV-B). We compute this score by an efficient projection of all the image windows on the frequency domain (Section IV-A). (c) Exact ASC similarity; note the unique and very sharp peak. Computing the complete similarity map is time consuming. Due to the prediction step, our proposed algorithm focuses only on a few promising candidate windows (d). This dramatically reduces the computation time (see Section V-C).

FFTs. The complexity of this step is $O(M^2 \log M)$ per candidate window. Usually the number of promising candidates is very small. Therefore in practice this step requires less time than the prediction step.

The candidate with the highest score is returned as the best match. Although theoretically the true best match might be missed during the prediction step, in practice it is very rare that many irrelevant candidates have higher partial scores even with the first few leading frequencies.

### D. Modification for Phase Correlation

The above algorithm can be modified to solve the problem of fast computation of the phase correlation with the template, $PC(q, w)$ (see the discussion in Section III-B.1 above). As for ASC, the partial PC score can be computed by changing Equation 8 as follows:

$$PC^{(k)}(q, w) = \sum_{(u,v)\in\mathcal{F}} \frac{Q_{uv} W^*_{uv}}{|Q_{uv}||W_{uv}|} \tag{9}$$

which is equivalent to computing $ASC^{(k)}(\tilde{q}, w)$, where $\tilde{q}$ is the result of normalizing each template's Fourier coefficient $Q_{uv}$ by its amplitude.

As in Equation 8, the set $\mathcal{F}$ contains the template's frequencies $Q_{uv}$ with the largest amplitudes. Although the magnitudes are ignored in $PC^{(k)}$, using these leading frequencies is important since they are more robust to noise than others (see below in Section VI).

The rest of the algorithm remains the same: the windows' frequencies are computed by the FFWP scheme (section IV-A), and a few promising candidates are predicted. In the final step $PC(q, w)$ is computed exactly for these promising candidates. Experimental evaluation of the proposed algorithm for $PC(q, w)$, compared with the common computation of $PC(t, f)$ (Section III-B.1, Equation 6) is presented in Section V-A.

### E. Extension Using a Robust Sum

The ASC similarity as defined in Section III-A is invariant to affine illumination change, but not to spatially varying illumination change or to partial occlusion. We deal with these distortions by matching the image with small sub-templates and combining their similarity scores using a robust sum:

$$ASC_{sub} = \sum_i \max\left(0, ASC(q^{(i)}, w^{(i)})\right) \tag{10}$$

where $q^{(i)}$ are sub-templates and $w^{(i)}$ are the corresponding sub-windows.[4]

$\text{ASC}_{sub}$ is robust to partial occlusion which typically changes only some of the sub-windows. Since each sub-window $w^{(i)}$ is normalized independently, $\text{ASC}_{sub}$ is also robust to non uniform illumination change.

To predict the best candidates, the windows' Fourier coefficients are computed by the FFWP scheme (Section IV-A) but this time for a smaller window size corresponding to the sub-templates. In our experiments we used five overlapping sub-templates of the same size: top left, top right, bottom left, bottom right and center.

[4]We found that alternatives such as maximum, do not perform as well as the robust sum.

Since each sub-template can have a different order of frequency amplitudes, we use a predefined order so that low frequencies $(u, v)$ are computed first. We store the partial ASC scores of each sub-template. After computing a new coefficient $W_{uv}$ and normalizing it, each entry is updated by adding the multiplication to the coefficient of the appropriate sub-template. This eliminates the need for repeating the whole computation many times and does not significantly increase the execution time compared to template matching using the original template.

An experimental evaluation of $ASC_{sub}$ is presented in Section V-B compared to the original ASC and other similarities.

## V. Results

We evaluated the performance of the proposed ASC template matching scheme on several image data sets. In all our experiments the templates were taken from high quality images (i.e. clean, bright and sharp images) and were sought in poor quality images containing noise and other distortions such as poor illumination, light changes, blur, etc. For each distortion ASC was compared against other relevant similarity functions as listed below.

### A. Noise, Blur, Poor Illumination

*1) Experimental Setting:* In this experiment we tested ASC and other template matching methods with noise produced by three models: additive Gaussian noise, multiplicative Gaussian noise and outlier ("salt-pepper") noise. In the additive Gaussian model, the value of each pixel was changed by adding noise sampled i.i.d from a Gaussian distribution $N(0, \sigma^2)$. In the multiplicative Gaussian model, the value of each pixel was multiplied by $N(1, \sigma^2)$. In the salt-pepper noise model, the image values in random locations $(x, y)$ were replaced by 0 or 1 randomly. The parameter of this noise model was the proportion of the corrupted pixels.

For each noise model (additive, multiplicative, outliers) we examined the performance in three conditions: (1) without additional distortions (noise only), (2) when the reference image had poor illumination, (3) when the reference image was blurred by camera defocus. We produced test images for all the possible combinations by corrupting the source with increasing amounts of noise.

For the noise only experiments we used a test set of 204 templates, each of size $61 \times 61$. The templates were selected from three large grayscale images: two images used in [36] (available in [37]), 'leuven-1' (80 templates) and 'bikes-1' (54 templates), and an additional test image 'boats' (70 templates). The templates were selected by sampling the images uniformly and eliminating flat windows with standard deviations smaller than 0.1.

For the experiments involving noise and poor illumination we chose 83 templates from the bright image 'leuven-1' and sought them in the dark image 'leuven-6', which was captured from a slightly different viewpoint with poor illumination. For the experiments involving blur and noise, 62 templates from 'bikes-1' were sought in 'bikes-6' which was captured with a defocused camera.[5]

A mismatch was defined as returning a center location with an Euclidean distance greater than 5 pixels from the ground truth center location (more than 85% overlap between the result window and the ground truth window). Because of the extreme noise that was added to the images, setting a threshold smaller than 5 pixels reduced the success rate of all the methods almost to zero. On the other hand, setting a larger threshold (e.g. 10, 15) had almost no effect on the results. Most of the mismatches were caused by finding a completely random location.

*2) Compared Methods:* ASC was compared to the following functions: the SSD and SAD metrics, three correlation variants: NCC, phase correlation (PC) and the matching function (MF) [17], two correlation filters: the phase only filter (POF) and an SDF-MACE optimal tradeoff filter (OTF, see below), the moment based method $M4_c$ [29] and the pattern-to-window variant of the MTM method [30] (using 20 bins of gray levels).

For the phase correlation we tested two variants. One was the common $PC(t, f)$ between the expanded template and the complete reference image (see Section III-B.1, Equation 6). This is denoted as 'PC-global'. The second was $PC(q, w)$ between the template $q$ and each local window $w$, which was computed using our efficient algorithm (Section IV-D). This is denoted as 'PC-local'.

The optimal tradeoff filter (OTF) was constructed by normalizing each of the template's Fourier coefficients as follows: $Q_{uv}/(|Q_{uv}|^2 + M^2)$.[6] This is the optimal tradeoff filter between SDF and MACE based on a single training example.

*3) Results:* The results of our experiment are shown in Figure 2. ASC performed best in all cases when the reference image was distorted by an additive or multiplicative Gaussian noise, with or without additional distortions (blur or poor illumination). The results of the additive Gaussian noise experiments support the thoretical analysis presented in Section VI.

In two of the salt and pepper noise experiments (noise only, noise + blur) ASC ranked second best with a success rate lower than SAD but higher than all the other methods. SAD, however, is sensitive to affine illumination change. Thus it had low success rate in the experiment involving salt and pepper noise and poor illumination. In this experiment ASC again outperformed all the other methods.

Another important point is the success rate of PC, POF and OTF. These methods had success rates of 100% when seeking a template $q$ in the image from which it was chosen, as done in the noise only experiments with a zero noise level. However, when seeking a template in another noiseless image of the same scene, the success rate of these methods was much lower (see the zero noise level in the experiments involving poor illumination and image blur; note that NCC and other methods had a 100% success on the same templates). This can be accounted for by the normalization of the template

[5]Due to the shifts between image pairs, the set of used templates were sampled differently than in the noise only experiment.

[6]Several different weights of $|Q_{uv}|^2$ and $M^2$ were examined, but did not impact the success rate significantly.

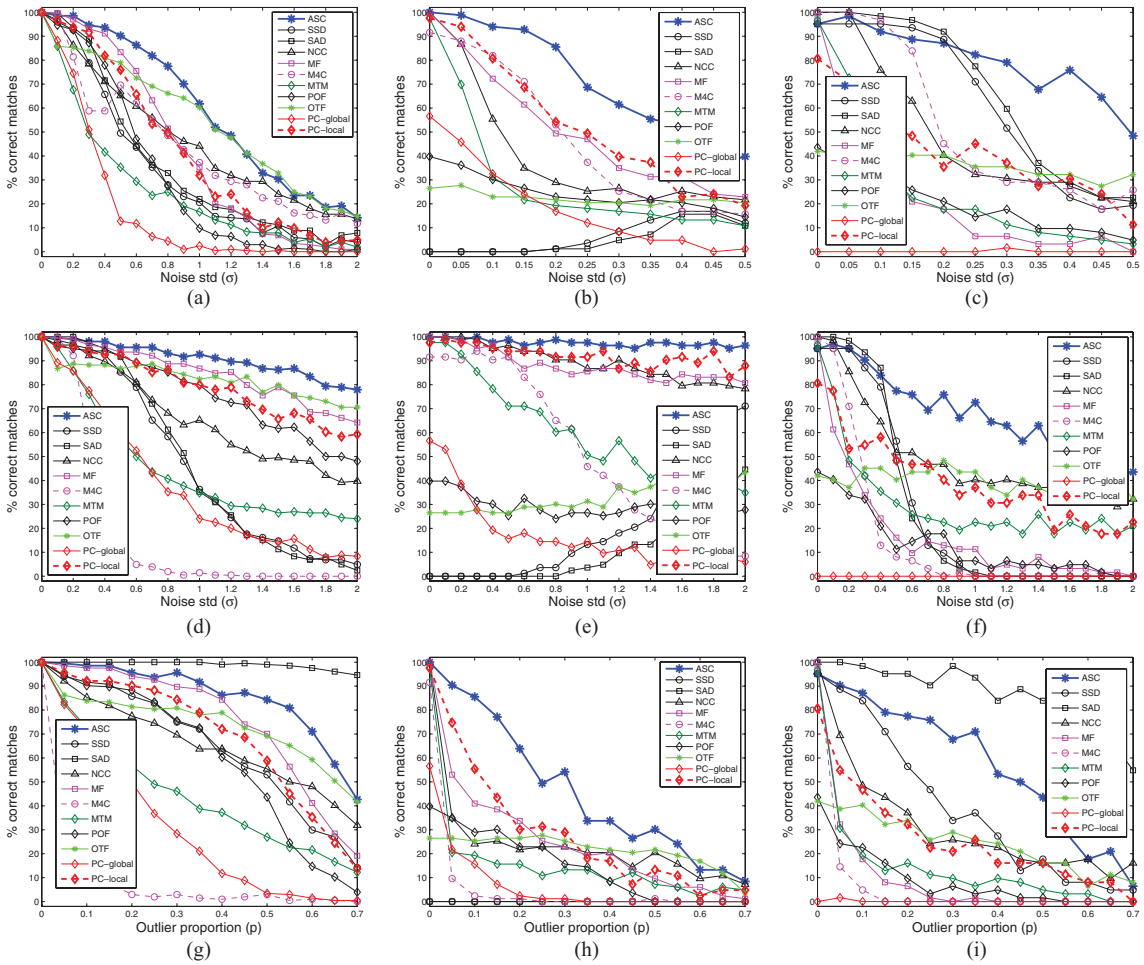



Fig. 2. Results of the synthetic noise experiments (Section V-A). First row: additive Gaussian noise without additional distortions (a), with poor illumination (b), and with image blur (c). Second row: multiplicative noise without additional distortions (d), with poor illumination (e), and with image blur (f). Third row: outlier (salt-pepper) noise without additional distortions (g), with poor illumination (h), and with image blur (i). The proposed ASC method is marked by the thick blue line.

Fourier spectrum which amplifies high frequencies. These frequencies contain unreliable information that varies even between slightly different instances of the same object. On the other hand, it can be seen that normalizing the Fourier spectrum of each image window, as done by ASC and PC-local, makes the detection more robust.

Note also that computing PC-local using our proposed algorithm (Section IV-D) was significantly better than computing PC-global using the standard method (Section III-B.1). This is because PC-local considers the DFT of local image windows, which is not the case for PC-global.

*4) Aerial Dataset:* We tested ASC on another dataset published by [14] which consists of a large aerial image ($1453 \times 1548$ pixels) and 59 templates of size $99 \times 99$. The dataset is available in [38]; example templates are shown in Figure 3(a). ASC was compared to SSD, NCC and PC in the presence of additive Gaussian noise. As ground truth we used the locations where SSD and NCC agreed on the noiseless reference image. The mismatch threshold was 5 pixels. As in the previous experiment, setting a larger threshold had almost no effect on the results.

As shown in Figure 3(b), our proposed ASC outperformed the other methods. The advantage of using PC-local instead of PC-global is also demonstrated.

### *B. Partial Occlusion and Spatially Varying Light*

The $ASC_{sub}$ variant was tested on two datasets, Leuven and Guitar. The Leuven dataset contains two images used in the noise experiment (Section V-A), one clean and one noisy ($\sigma = 0.1$) and 83 bright clean templates. The reference images in this dataset have poor illumination, for which standard methods used to overcome partial occlusion (e.g. SAD) are not appropriate. ASC, on the other hand, is invariant to illumination change. We compared ASC to NCC, POF and to two variants of the MF method [17] with 2-neighbors ($MF_2$) and 1,2-neighbors ($MF_{1,2}$) which were shown in [18] to be very robust in similar conditions. All the methods were examined both in their original form; i.e., matching a template to a single image window, and using the sub-templates extension described in Section IV-E.

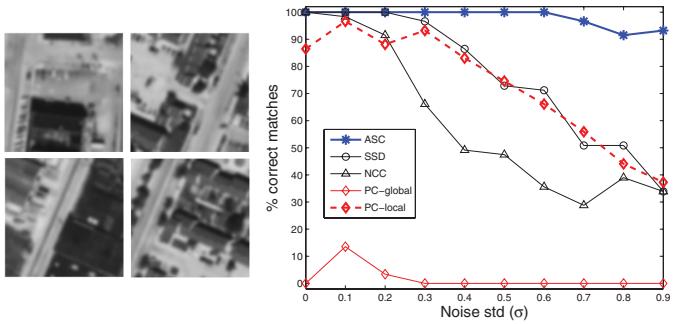
In each of the 83 tests on the Leuven dataset, the left half of the target template was occluded by a black rectangle.

The results are shown in Table I. The success rate of ASC was higher than NCC and POF but lower than the MF variants. Using the sub-template extension improved the results significantly: $ASC_{sub}$ was very close to MF when the reference image was clean, and had an equal success rate to the best MF variant ($MF_2$ with single template) when the reference image was noisy.

TABLE I

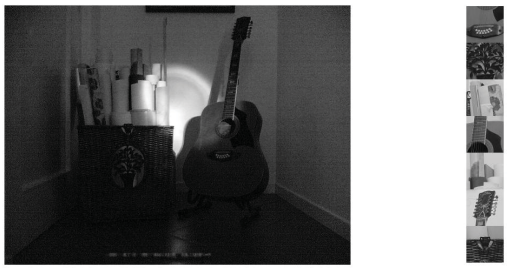
PARTIAL OCCLUSION AND SPATIALLY VARYING LIGHT (SECTION V-B). THE TABLE SHOWS THE NUMBER OF CORRECT MATCHES OUT OF THE TOTAL NUMBER OF TESTS. FOR EACH METHOD THE TABLE INDICATES THE RESULTS OF MATCHING A SINGLE TEMPLATE (LEFT SIDE) AND USING A ROBUST SUM ON SUB-TEMPLATES (RIGHT SIDE, SEE SECTION IV-E)

| Dataset | Single Template | | | | | Sub-Templates | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $MF_2$ | $MF_{1,2}$ | POF | NCC | ASC | $MF_2$ | $MF_{1,2}$ | POF | NCC | ASC |
| Leuven (occlusion, no noise) | 72/83 | 72/83 | 32/83 | 29/83 | 54/83 | 73/83 | 73/83 | 19/83 | 30/83 | 69/83 |
| Leuven (occlusion, noise std = 0.1) | 55/83 | 53/83 | 24/83 | 26/83 | 47/83 | 43/83 | 42/83 | 14/83 | 21/83 | 55/83 |
| Guitar [18] (varying light) | 64/70 | 64/70 | 50/70 | 61/70 | 59/70 | 66/70 | 66/70 | 58/70 | 64/70 | 67/70 |




(a) (b)

Fig. 3. Experimental results on the Aerial image dataset [14] (Section V-A). (a) shows several template examples. (b) shows the success rate in the presence of additive Gaussian noise.

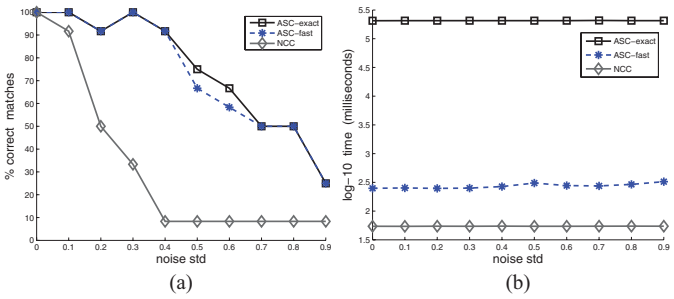


(a) (b)

Fig. 4. Guitar dataset [18], [39]. (a) sample image (number 5); (b) the 7 sought templates. The proposed $ASC_{sub}$ variant was the best performing method on this dataset, slightly better than MF [17] (see Table I).

NCC had a low success rate since it is less discriminating than ASC; i.e., its peak is not as sharp. Since the left half of the template was occluded, only two sub-windows were not distorted (top right and bottom right). Thus, the sum of two smooth peaks was not sufficient for NCC to find the correct match. POF was also unsuccessful on this task. Although POF has a sharp peak, it is sensitive even to small distortions (see Section I and Section V-A). On the other hand, ASC exhibited good performance since it is both a discriminating and robust similarity measure.

The Guitar dataset (available online in [39]) contains 7 clean templates and 10 images with various lighting conditions. An example of this dataset is shown in Figure 4. In [18], the best performing method on this dataset was the MF method [17]. The original, single template ASC similarity performed better on this dataset than POF, slightly less well than NCC but much less well than the examined MF variants. Using the sub-templates extension improved the performance of all similarities but not to the same extent. The ASC similarity showed the greatest improvement, making $ASC_{sub}$ the best performing method with 67 correct matches out of 70 tests.




(a) (b)

Fig. 5. Comparison of the C++ implementations of NCC, exact (naive) ASC, and the fast ASC template matching scheme proposed in Section IV. The proposed algorithm had an almost identical success rate to the full computation (a), but was 744 times faster on average (b).

### C. Execution Time

In order to evaluate the computational aspects of the ASC template matching scheme we implemented the proposed fast algorithm (Section IV) and an exact computation of ASC for each image window in C++. FFT computations in both these implementations used the FFTW3 library [40]. We compared the performance of these programs to the openCV implementation of the NCC (the 'cvMatchTemplate' function). The results are shown for 12 61 $\times$ 61 templates in the presence of illumination changes and 10 noise levels.

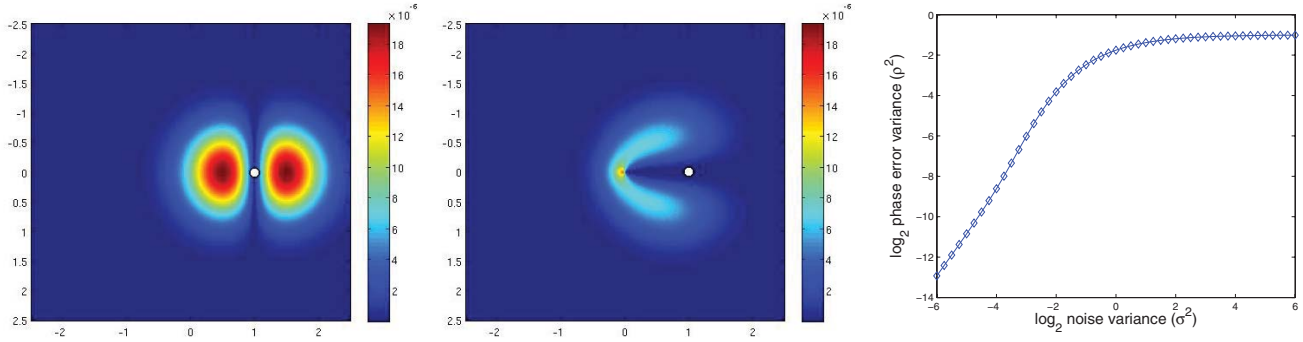
Figure 5(a) shows that the success rate of the proposed fast algorithm is almost identical to the exact full computation, and significantly higher than NCC. On the other hand, the average computation time for the fast algorithm was 274 milliseconds. This was 5.9 times slower than NCC (54 milliseconds) and on 744 times faster on average than the full ASC computation (206.3 seconds = 3.4 minutes). A $\log_{10}$ plot of the execution time is shown in Figure 5(b). It can be seen that the execution time of the proposed fast scheme is independent of the noise level.

## VI. NOISE ROBUSTNESS ANALYSIS

We analyzed the robustness of the proposed ASC similarity to additive white Gaussian noise. We assume w.l.o.g that the given template $q$ has zero mean and a variance $\text{var}(q) = q^T q = 1$.

One common measure of noise robustness of a similarity function is the *signal to noise ratio* (SNR). Given a template $q$, the SNR of a similarity function 'sim' is

$$SNR(\text{sim}) = \frac{\text{sim}(q,q)^2}{\text{var}(\text{sim}(q, q+\eta))} \tag{11}$$




(a) (b) (c)

Fig. 6. Phase error variance (Section VI-A). In (a,b) two error measures are illustrated: cross correlation error (a) and phase error (b). In both figures the complex number $z = 1 + 0i$ (marked by a white point) is compared with a noisy number $z + \eta$, where $\eta$ is complex Gaussian noise with zero mean and variance $\sigma^2 = 0.5$. (a) measures the difference of the cross correlation from 1, while (b) measures the difference of the phase from 0. The value plotted at each point $x + iy$ is the error at that point multiplied by the probability $Pr(x + iy)$. The variance of the cross correlation error equals the noise variance $\sigma^2$. The variance of the phase error, $\rho^2(\sigma^2)$ is much smaller. In (c) the phase error variance $\rho^2$ is plotted as a function of the noise variance $\sigma^2$ in $\log_2$:$\log_2$ scale. $\rho^2$ is smaller than $\min(\sigma^2, 1/2)$.

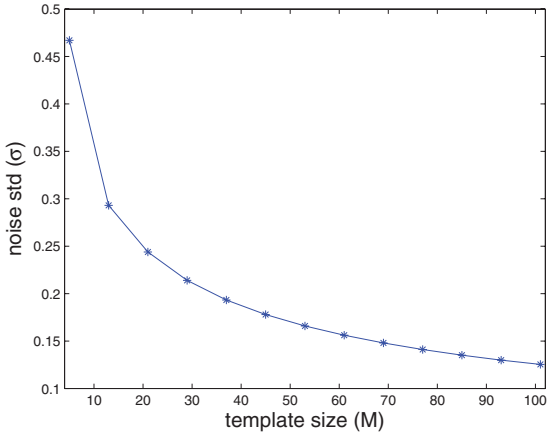



Fig. 7. Comparison between the noise robustness of ASC and NCC (Section VI-C). For each template size $M \times M$ we plot the noise standard deviation $\sigma$ from which $SNR(ASC) \geq SNR(ASC)$ (Equation 27). We proved that for noise with standard deviation greater than $\sigma$, ASC is more robust than NCC. For smaller standard deviation there is no theoretical guarantee, but the practical performance of ASC is better (see Section V).

where $\eta$ is white Gaussian noise with zero mean and variance $\sigma^2$. The larger the SNR the more robust the similarity is to noise. By Equation 11, the SNR of the cross correlation is

$$\frac{\left(q^T q\right)^2}{\text{var}\left(q^T(q+\eta)\right)} = \frac{1}{\sigma^2} \tag{12}$$

The cross correlation (a.k.a. matched filter) is the linear filter with maximal SNR [5], [10]. However, NCC and ASC are not linear correlation filters; i.e., they are non linear functions of $w$. As shown below, the SNR of these similarities is higher than the cross correlation.

### A. SNR of ASC

From the definition of ASC (Equation 2) it follows that the SNR of ASC is

$$\frac{\left(ASC(q,q)\right)^2}{\text{var}(ASC(q,q+\eta))}$$

$$= \frac{\left(\sum_{(u,v)} |Q_{uv}|\right)^2}{\sum_{(u,v)} |Q_{uv}|^2 \text{var}\Big[\cos\Big(\theta(Q_{uv}) - \theta(Q_{uv} + \eta_{uv})\Big)\Big]} \tag{13}$$

where $\theta(\cdot)$ denotes the phase of a complex number and $\eta_{uv}$ is the noise added to each frequency; i.e., a complex Gaussian noise $\eta_{uv} \sim N(0, \sigma^2)$. The decomposition of the denominator in Equation 13 is based on the independence of white noise across different frequencies $(u, v)$.

On the right side of Equation 13 the noise $\eta_{uv}$ appears only inside cosine terms. Since the cosine is in $[-1, 1]$, the denominator of the right side in Equation 13 is smaller than

$$\sum_{(u,v)} |Q_{uv}|^2 \;=\; q^T q \;=\; 1 \tag{14}$$

Therefore the SNR of ASC is greater than $\Big(\sum_{u,v} |Q_{uv}|\Big)^2$ which is greater than 1. In the following we provide a tighter lower bound on the SNR of ASC, and show that it is larger than the SNR of the cross correlation and NCC.

*Definition 1: Phase Error Variance:* Let $\eta$ be a complex Gaussian noise, $\eta \sim N(0, \sigma^2)$. The phase error variance is defined as

$$\rho^2(\sigma^2) := \text{var}\Big[\cos\Big(\theta(1+\eta)\Big)\Big] \tag{15}$$

where $\theta(\cdot)$ denotes the phase of a complex number.

The $\rho^2$ function is analyzed in the Appendix, in which we prove the following two claims:

*Claim 1:* For any $\sigma$ it holds that $\rho^2(\sigma^2) \leq 1/2$.

*Claim 2:* Let $z$ be a complex number with amplitude $|z|$ and phase $\theta(z)$, $\eta$ complex Gaussian noise, $\eta \sim N(0, \sigma^2)$. Then

$$\text{var}\Big[\cos\Big(\theta(z) - \theta(z+\eta)\Big)\Big] = \rho^2\left(\frac{\sigma^2}{|z|^2}\right)$$

In the Appendix we evaluated $\rho^2(\sigma^2)$ for sampled $\sigma^2$ values. This is done by developing the first and second moments of $\cos\Big(\theta(1+\eta)\Big)$ and integrating them numerically.

Figure 6 illustrates the difference between the cross correlation error (6(a)) and the phase error (6(b)). In both graphs a complex number $z = 1 + 0i$ is compared to a noisy number $z+\eta$. The cross correlation error is measured by $|1-z(z+\eta)^*|$, and its variance equals to the noise variance $\sigma^2$. The phase error is measured by $\cos\Big(\theta(z)-\theta(z+\eta)\Big)$ and by Definition 1 its variance is $\rho^2(\sigma^2)$. Note that $\rho^2(\sigma^2) \ll \sigma^2$. This is because that part of the noise is ignored when considering the phase rather than the actual noisy number.

Figure 6(c) plots $\rho^2(\sigma^2)$ vs. $\sigma^2$ in $\log_2$:$\log_2$ scale. It can be seen that $\rho^2$ increases monotonically in $\sigma^2$ and that it approaches $1/2$ when $\sigma^2$ becomes large. For small noise levels $\sigma^2$ we consistently have $\rho^2 \ll \sigma^2$.

Using Claims 1 and 2, the SNR of ASC (Equation 13) can be rewritten and bounded from below:

$$SNR(ASC) = \frac{\Big(\sum_{uv} |Q_{uv}|\Big)^2}{\sum_{(u,v)} |Q_{uv}|^2 \rho^2\left(\frac{\sigma^2}{|Q_{uv}|^2}\right)} \geq \frac{2\Big(\sum_{uv} |Q_{uv}|\Big)^2}{\sum_{(u,v)} |Q_{uv}|^2} \tag{16}$$

In the following we use this lower bound to compare the SNR of ASC and NCC.

### B. SNR of NCC

Interestingly, the SNR of NCC is also related to the phase error variance $\rho^2$. Recall that $q$ has, w.l.o.g, zero mean and that $NCC(q, q) = 1$. Hence,

$$SNR(NCC) = \frac{1}{\mathrm{var}\Big(\frac{q^T(q+\eta)}{\|q\|\|q+\eta\|}\Big)} = \frac{1}{\mathrm{var}\Big(\cos(\theta(q, q+\eta))\Big)} \tag{17}$$

where $\theta(q, q+\eta)$ is the angle between the two vectors $q$ and $q+\eta$. This angle is well defined since any two non zero vectors $q \neq (q+\eta)$ span a unique 2D plane that crosses the origin.

Denote by $\mathcal{P}$ the pencil of all the 2D planes in $R^M$ containing the origin and the point $q$. Since the noise is white and Gaussian, the distribution of $q+\eta$ on $\mathcal{P}$ is uniform. In other words, the probability $Pr(q+\eta \in P)$ is equal for all the planes $P \in \mathcal{P}$. Moreover, the conditional distribution of $q+\eta$ on each plane $P$ is Gaussian and can be characterized as follows.

Denote by $x_q, y_q$ the projection of $q$ on the 2D plane $P$, and assume that $q+\eta \in P$. Let $p(x, y)$ be the distribution of $q+\eta$ on $P$. Then $p(x, y)$ is a 2D Gaussian distribution with mean $(x_q, y_q)$ and covariance $\sigma^2 I$. Equivalently, $p(x, y)$ represents a Gaussian distribution on 1D complex numbers $x+iy$ with mean $x_p + iy_p$ and variance $\sigma^2$.

Since the vectors $q$ and $q+\eta$ are in the plane $P$, the angle between their projections $(x_p, y_p)$ and $(x, y)$ equals to $\theta(q, q+\eta)$. Considering the 2D coordinates $x, y$ as 1D complex numbers it follows that

$$\theta(q, q+\eta) = \Big|\theta(x+iy) - \theta(x_p + iy_p)\Big| \tag{18}$$

Hence, by Definition 1 and Claim 2,

$$\begin{aligned}&\mathrm{var}_{(q+\eta)\in P}\Big(\cos(\theta(q, q+\eta))\Big) \\ &= \mathrm{var}_{(x,y)}\Big(\cos\big(\theta(x+iy) - \theta(x_p+iy_p)\big)\Big) \\ &= \rho^2(\sigma^2)\end{aligned} \tag{19}$$

Integrating over planes $P \in \mathcal{P}$ uniformly gives

$$SNR(NCC) = \frac{1}{\rho^2(\sigma^2)} \tag{20}$$

Since $\rho^2(\sigma^2) < \sigma^2$, it always holds that SNR(NCC) > SNR(cross correlation).

### C. Comparison of the SNR Values

In the following we compare the SNR of the cross correlation, NCC and ASC. We show that in the typical situation where the template $q$ is a natural image, ASC has a significantly higher SNR than the cross correlation or NCC. For this purpose we consider Field's model for the power spectrum of natural images [41].

Let $F$ be the Fourier spectrum of a real world image. According to Field's model, the amplitudes of $F$ coefficients are inversely proportional to the norm of the frequency, so that

$$\forall(u, v) \qquad |F_{uv}| = \frac{c_F}{\|(u, v)\|} \tag{21}$$

where $c_F$ is a constant.

For simplicity we first consider the 1D case. Let $q$ be a zero-mean 1D template of size $M$ which satisfies Field's model, and $Q$ is its Fourier spectrum. Then

$$\frac{\Big(\sum_u |Q_u|\Big)^2}{\sum_u |Q_u|^2} = \frac{\Big(c_F \sum_{u>0} \frac{1}{|u|}\Big)^2}{(c_F)^2 \sum_{u>0} \frac{1}{u^2}} \geq \frac{\ln^2(M)}{\frac{\pi^2}{6}} > 0.6 \ln^2(M) \tag{22}$$

Dealing with the 2D case is better accomplished using a continuous representation and polar coordinates. Let $q$ be a zero mean 2D template with radius $m$.[7] Then

$$\begin{aligned}\int_u \int_v |Q_{uv}|\, dv\, du &= \int_{r=1}^{m} \int_{\theta=0}^{2\pi} |Q_{r,\theta}|\, d\theta\, dr \\ &= \int_{r=1}^{m} 2\pi r \cdot \frac{c_F}{r}\, dr \\ &= 2\pi c_F (m-1)\end{aligned} \tag{23}$$

Similarly,

$$\begin{aligned}\int_u \int_v |Q_{uv}|^2\, dv\, du &= \int_{r=1}^{m} \int_{\theta=0}^{2\pi} |Q_{r,\theta}|^2\, d\theta\, dr \\ &= \int_{r=1}^{m} 2\pi r \cdot \frac{(c_F)^2}{r^2}\, dr \\ &= 2\pi (c_F)^2 \ln(m)\end{aligned} \tag{24}$$

Combining Equations 23 and 24 it follows that

$$\frac{\Big(\int_{uv} |Q_{uv}|\, dv\, du\Big)^2}{\int_{uv} |Q_{uv}|^2\, dv\, du} = \frac{\Big(2\pi c_F (m-1)\Big)^2}{2\pi (c_F)^2 \ln(m)} = \frac{2\pi (m-1)^2}{\ln(m)} \tag{25}$$

[7] We discuss circular templates for simplicity alone. An analogous analysis can be done with square templates but is more complicated.

Based on Equation 25 we derive a lower bound for the ratio between the SNR of ASC (Equation 16) and the SNR of NCC (Equation 20):

$$\frac{SNR(ASC)}{SNR(\text{NCC})} \geq \frac{2\Big(\int_u\int_v\ |Q_{uv}|\ dv\ du\Big)^2\rho^2(\sigma^2)}{\int_u\int_v\ |Q_{uv}|^2\ dv\ du} \geq \rho^2(\sigma^2)\ \frac{4\pi(m-1)^2}{\ln(m)} \tag{26}$$

Hence, ASC is more robust than NCC where

$$\rho^2(\sigma^2) \geq \frac{\ln(m)}{4\pi(m-1)^2} \tag{27}$$

Figure 7 plots for each templates size $M \times M$ ($M = 2m+1$) the noise standard deviation $\sigma$ from which we proved that $SNR(ASC) \geq SNR(NCC)$. For example, given a $61 \times 61$ template (radius $m = 30$) as in our major experiment in Section V-A, ASC is more robust for noise standard deviation $\sigma \geq 0.1561$. In practice, as shown in Section V, ASC outperforms NCC as well for smaller $\sigma$ values, e.g. $\sigma = 0.1$. This is because the above inequalities are based on the bound $\rho^2(\sigma^2) \leq 1/2$ (used in Equation 16 to bound $SNR(ASC)$) which is not tight. Thus performance in practice is better than in theory.

## VII. Summary

A robust and efficient template matching method using *asymmetric correlation* (ASC) was presented. The proposed ASC template matching is invariant to affine illumination changes and highly robust to noise. This was shown both theoretically and experimentally. An extension using robust sums was presented which makes ASC template matching robust to partial occlusions and spatially variant illumination changes. Future work includes extending ASC to deal with geometric transformations.

## Appendix

## VIII. Phase Error Variance

In Section VI-A we defined the phase error variance $\rho^2$ as follows. Let $\eta$ be complex Gaussian noise, $\eta \sim N(0, \sigma^2)$. Then the phase error variance is

$$\rho^2(\sigma^2) := \text{var}\Big[\cos\Big(\theta(1+\eta)\Big)\Big] \tag{28}$$

where $\theta(\cdot)$ denotes the phase of a complex number.

*Claim 1:* $\rho^2(\sigma^2) \leq 1/2$

*Proof:* The noisy complex number $1+\eta$ has a Gaussian distribution $N(1, \sigma^2)$. Its phase $\theta(1+\eta)$, $\theta$, has a unimodular symmetric distribution on $[-\pi, \pi]$ with a maximum at zero. The maximum variance of $\cos(\theta)$ is achieved when $\theta$ distributes uniformly on $[-\pi, \pi]$ ($\sigma = \infty$). In this case the expectation is

$$\text{E}[\cos(\theta)] = \int_{\theta=-\pi}^{\pi} \frac{1}{2\pi}\cos(\theta)d\theta = 0 \tag{29}$$

and the variance equals the second moment,

$$\text{var}\Big(\cos(\theta)\Big) = \int_{\theta=-\pi}^{\pi} \frac{1}{2\pi}\cos^2(\theta)d\theta = \frac{1}{2} \tag{30}$$

Thus $\rho^2(\sigma^2) < 1/2$ independent of $\sigma^2$. ■

### A. $\rho^2$ Calculation.

The value of $\rho^2(\sigma^2)$ can be calculated by the Cartesian representation of the noisy complex number, $1+\eta = x_\eta + i \cdot y_\eta$. The expected value of the phase noise error is

$$\begin{aligned}\mu_1 &= \text{E}\Big[\cos\big(\theta(1+\eta)\big)\Big] \\ &= \text{E}\Big[\frac{x_\eta}{\sqrt{x_\eta^2+y_\eta^2}}\Big] \\ &= \int_x\int_y \frac{1}{2\pi\sigma^2}\exp\Big(-\frac{(x_\eta-1)^2+y_\eta^2}{2\sigma^2}\Big)\frac{x_\eta}{\sqrt{x_\eta^2+y_\eta^2}}dydx\end{aligned} \tag{31}$$

The second moment is

$$\begin{aligned}\mu_2 &= \text{E}\Big[\cos^2\big(\theta(I+\eta)\big)\Big] \\ &= \text{E}\Big[\frac{x_\eta^2}{x_\eta^2+y_\eta^2}\Big] \\ &= \int_x\int_y \frac{1}{2\pi\sigma^2}\exp\Big(-\frac{(x_\eta-1)^2+y_\eta^2}{2\sigma^2}\Big)\frac{x_\eta}{x_\eta^2+y_\eta^2}dydx\end{aligned} \tag{32}$$

and the variance is

$$\rho^2(\sigma^2) = \mu_2 - (\mu_1)^2 \tag{33}$$

We calculated the values of $\mu_1$ and $\mu_2$ numerically using Matlab. A log:log plot is shown in Figure 6(c). It can be seen that $\rho^2$ increases monotonically in $\sigma^2$ and that it approaches $1/2$ when $\sigma^2$ becomes large. For small noise levels $\sigma^2$, we have $\rho^2 \ll \sigma^2$. This is because part of the noise is eliminated due the Fourier coefficient normalization. Figure 6(a,b) illustrates the difference between the noise error of the cross correlation for a single frequency $(u, v)$ where $Q_{uv} = 1 + 0i$ and the phase noise error.

*Claim 2:* Let $z$ be a complex number with amplitude $|z|$ and phase $\theta(z)$, $\eta$ a complex Gaussian noise, $\eta \sim N(0, \sigma^2)$. Then

$$\text{var}\Big[\cos\Big(\theta(z)-\theta(z+\eta)\Big)\Big] = \rho^2\left(\frac{\sigma^2}{|z|^2}\right) \tag{34}$$

*Proof:* Let $z_0$ be the complex number $|z_0| = |z|$, $\theta(z_0) = 0$. Then

$$\begin{aligned}\theta(z)-\theta(z+\eta) &= \Big(\theta(z)-\theta(z)\Big) - \Big(\theta(z+\eta)-\theta(z)\Big) \\ &= -\theta(z_0+\eta)\end{aligned} \tag{35}$$

Hence,

$$\begin{aligned}\text{var}\Big(\cos\big(\theta(z)-\theta(z+\eta)\big)\Big) &= \text{var}\Big(\cos\big(-\theta(z_0+\eta)\big)\Big) \\ &= \text{var}\Big(\cos\big(\theta(z_0+\eta)\big)\Big) = \text{var}\Big(\cos\big(\theta(\frac{z_0+\eta}{|z|})\big)\Big) \\ &= \text{var}\Big(\cos\big(\theta(1+\frac{\eta}{|z|})\big)\Big) = \rho^2\left(\frac{\sigma^2}{|z|^2}\right)\end{aligned} \tag{36}$$

■

## REFERENCES


[1] F. Jurie and M. Dhome, "A simple and efficient template matching algorithm," in *Proc. 8th IEEE Int. Conf. Comput. Vision*, vol. 2. Jul. 2001, pp. 544–549.

[2] K. Takita, M. A. Muquit, and T. Higuchi, "A sub-pixel correspondence search technique for computer vision applications," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. 87, no. 8, pp. 1913–1923, 2004.

[3] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–21, Aug. 2009.

[4] M. S. Sussman and G. A. Wright, "Factors affecting the correlation coefficient template matching algorithm with application to real-time 2-D coronary artery MR imaging," *Med. Imag.*, vol. 22, no. 2, pp. 206–216, Feb. 2003.

[5] B. V. K. V. Kumar, A. Mahalanobis, and R. D. Juday, *Correlation Pattern Recognition*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[6] J. P. Lewis, "Fast normalized cross-correlation," *Vis. Inter.*, vol. 10, no. 1, pp. 120–123, 1995.

[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jan. 2001, pp. 1–511.

[8] E. Jacobsen and R. Lyons, "The sliding DFT," *IEEE Signal Process. Mag.*, vol. 20, no. 2, pp. 74–80, Mar. 2003.

[9] G. S. Cox, "Template matching and measures of match in image processing," Ph.D. dissertation, Dept. Comput. Sci., Cape Town Univ., Cape Town, South Africa, 1995.

[10] R. Brunelli, *Template Matching Techniques in Computer Vision*. New York, NY, USA: Wiley, 2009.

[11] W. Ouyang, F. Tombari, S. Mattoccia, L. Di Stefano, and W.-K. Cham, "Performance evaluation of full search equivalent pattern matching algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 127–143, Jan. 2012.

[12] W. H. Pan, S. D. Wei, and S. H. Lai, "Efficient ncc-based image matching in walsh-hadamard domain,"in *Proc. 10th Eur. Conf. Comput. Vis.*, 2008, pp. 468–480.

[13] H. Schweitzer, R. F. Anderson, and R. Deng, "A near optimal acceptance-rejection algorithm for exact cross-correlation search," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Oct. 2009, pp. 1089–1094.

[14] A. Mahmood and S. Khan, "Exploiting transitivity of correlation for fast template matching," *IEEE Trans. Image Process.*, vol. 19, no. 8, pp. 2190–2200, Aug. 2010.

[15] L. Di Stefano, S. Mattoccia, and F. Tombari, "Zncc-based template matching using bounded partial correlation," *Pattern Recognit. Lett.*, vol. 26, no. 14, pp. 2129–2134, Oct. 2005.

[16] S.-D. Wei and S.-H. Lai, "Robust and efficient image alignment based on relative gradient matching," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2936–2943, Oct. 2006.

[17] F. Tombari, L. Di Stefano, and S. Mattoccia, "A robust measure for visual correspondence," in *Proc. 14th Int. Conf. Image Anal. Process.*, Sep. 2007, pp. 376–381.

[18] F. Tombari, L. Di Stefano, S. Mattoccia, and A. Galanti, "Performance evaluation of robust matching measures," in *Proc. VISAPP*, 2008, pp. 473–478.

[19] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *Proc. Int. Conf. Cybern. Soc.*, 1975, pp. 299–324.

[20] J. L. Horner and P. D. Gianino, "Phase-only matched filtering," *Appl. Opt.*, vol. 23, no. 6, pp. 812–816, Mar. 1984.

[21] C. F. Hester and D. Casasent, "Multivariate technique for multiclass pattern recognition," *Appl. Opt.*, vol. 19, no. 11, pp. 1758–1761, 1980.

[22] A. Mahalanobis, B. V. K. Kumar, and D. Casasent, "Minimum average correlation energy filters," *Appl. Opt.*, vol. 26, no. 17, pp. 3633–3640, Sep. 1987.

[23] A. Mahalanobis, B. V. K. Vijaya Kumar, S. Song, S. R. F. Sims, and J. F. Epperson, "Unconstrained correlation filters," *Appl. Opt.*, vol. 33, no. 17, pp. 3751–3759, 1994.

[24] B. G. Shin, S.-Y. Park, and J. J. Lee, "Fast and robust template matching algorithm in noisy image," in *Proc. Int. Conf. Control Autom. Syst.*, Oct. 2007, pp. 6–9.

[25] O. Pele and M. Werman, "Robust real-time pattern matching using bayesian sequential hypothesis testing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 8, pp. 1427–1443, Aug. 2008.

[26] J.-H. Chen, C.-S. Chen, and Y.-S. Chen, "Fast algorithm for robust template matching with m-estimators," *Signal Process.*, vol. 51, no. 1, pp. 230–243, 2003.

[27] A. J. Fitch, A. Kadyrov, W. J. Christmas, and J. Kittler, "Fast robust correlation," *Image Process.*, vol. 14, no. 8, pp. 1063–1073, 2005.

[28] A. Sibiryakov, "Fast and high-performance template matching method," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1417–1424.

[29] F. Essannouni and D. Aboutajdine, "Fast frequency template matching using higher order statistics," *Image Process.*, vol. 19, no. 3, pp. 826–830, 2010.

[30] Y. Hel-Or, H. Hel-Or, and E. David, "Fast template matching in non-linear tone-mapped images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1355–1362.

[31] Q. Tian and M. N. Huhns, "Algorithms for subpixel registration," *Comput. Vis., Graph., Image Process.*, vol. 35, no. 2, pp. 220–233, 1986.

[32] J. Ren, J. Jiang, and T. Vlachos, "High-accuracy sub-pixel motion estimation from noisy images in fourier domain," *Image Process.*, vol. 19, no. 5, pp. 1379–1384, 2010.

[33] E. Elboher and M. Werman, "Cosine integral images for fast spatial and range filtering," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 89–92.

[34] J. Stuller, "Generalized running discrete transforms," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 30, no. 1, pp. 60–68, Feb. 1982.

[35] G. Ben-Artzi, H. Hel-Or, and Y. Hel-Or, "The gray-code filter kernels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 382–393, Mar. 2007.

[36] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[37] K. Mikolajczyk and C. Schmid. (2007, Jul.). *Affine Covariant Features—Test Data* [Online]. Available: http://www.robots.ox.ac.uk/∼vgg/sresearch/affine/

[38] *Aerial Dataset*. (2010, Aug.) [Online]. Available: http://cvlab.lums.edu.pk/tea/

[39] *Guitar Dataset*. (2013) [Online]. Available: http://vision.deis.unibo.it/pm-eval.aspx

[40] M. Frigo and S. G. Johnson, "The design and implementation of FFTW3," *Proc. IEEE*, vol. 93, no. 2, pp. 216–231, Feb. 2005.

[41] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer.*, vol. 4, no. 12, pp. 2379–2394, Dec. 1987.


**Elhanan Elboher** received the B.Sc. degree in computer science and life science and the M.Sc. degree in computer science from the Hebrew University of Jerusalem, in 2007 and 2009, respectively. He is currently pursuing the Ph.D. degree in computer science with the Hebrew University of Jerusalem. His current research interests include image processing and computer vision.

**Michael Werman** received the Ph.D. degree from The Hebrew University, in 1986. He is currently a Professor of computer science with The Hebrew University. His current research interests include designing computer algorithms and mathematical tools for analyzing, understanding, and synthesizing pictures.