

Improved Phase Correlation Matching

Javed Ahmed and M. Noman Jafri

Image Processing Centre, NUST Military College of Signals, Rawalpindi, Pakistan
{javed, mnjafri}@mcs.edu.pk

Abstract. Phase correlation based template matching is an efficient tool for translation estimation which is in turn required for the image registration and the object tracking applications. When a template of an object is phase correlated with the search image, the resulting correlation surface is supposed to contain a sharp peak corresponding to the location of the object in the search image. However, the resulting surface also contains various *false* peaks which are sometimes *higher* in magnitude than the *true* peak. In order to solve the problem, we present an efficient and effective preprocessing technique that extends the images with new pixels having decaying values. The technique is compared with two recent methods on cluttered, noisy, blurred, and slightly rotated scenes. The results show that the proposed method outperforms both of them, especially when the object is away from the central region in the image.

1 Introduction

Precise estimation of the amount of translation of one image with respect to another is an important stage in an image registration application. Similarly, accurate localization of an object in a video frame is a critical step in an object tracking application. Both of these problems can be considered as a matching problem, which can be efficiently solved by cross-correlating the two images. However, the standard correlation is illumination sensitive and produces un-normalized response with wide peaks resulting in inaccurate matching [1]. On the contrary, phase correlation (PC) is significantly robust to illumination variation and offers a normalized response with a sharp peak at the best match location [1,2,3,4,5]. The PC between a search image, s , and a template, t , is computed as:

$$c = \mathfrak{F}^{-1} \left(\frac{\mathfrak{F}(s)}{|\mathfrak{F}(s)|} \cdot \frac{\mathfrak{F}(t)^*}{|\mathfrak{F}(t)|} \right) \quad (1)$$

where $\mathfrak{F}(\cdot)$ and $\mathfrak{F}^{-1}(\cdot)$ are the 2-D DFT (discrete Fourier transform) and inverse DFT functions, respectively, the $|\cdot|$ operator computes the magnitude of every complex number in its input matrix, the asterisk (*) is the complex-conjugate operation, and the division and the multiplication are performed element-by-element. The normalization of the DFTs of s and t is usually called whitening of the signals [1].

If the PC is performed directly on the *original* signals, the result is the circular correlation, since the DFT assumes that each of the two signals is repeated

after its final element. In order to obtain the true PC, the signals must be appropriately zero-padded before computing their DFTs [3,6]. However, the zero-padding operation makes the spurious frequencies due to the discontinuities at the boundaries of the signals more evident, resulting in false peaks in the correlation response [4]. The problem occurs when any one of these false peaks has higher magnitude than that of the true peak.

Stone et al. [4] proposed to modulate (i.e. multiply element-by-element) the original images with a Blackman window [7] before zero-padding operation. The Blackman window smoothly attenuates the value of the image pixel depending on its position relative to the image center. Therefore, the discontinuities at the boundaries of the image are eliminated and the PC between the resulting images becomes significantly robust [4]. However, there is a disadvantage of this technique that it distorts the actual content of the image. That is, if the object of interest (e.g. a vehicle) lies significantly away from the center in the search image as shown in Fig. 1(*left*), the object becomes almost invisible in the modulated image as illustrated in Fig. 1(*right*). Similarly, the modulated template is also distorted, as depicted in Fig. 1. As a result, the modulated template does not match with the actual object as good as it does with the clutter at the central region, and the resulting PC response contains multiple false peaks at the center having magnitude higher than that of the true peak, as shown in Fig. 2(*left*). Thus, the phase correlation between the Blackman windowed images is failed to localize the actual object [Fig. 2(*right*)].

Keller et al. [5] suggested performing post-processing instead of pre-processing. That is, they phase correlated the original images and used a projection operator to restrict the support of the resulting surface. The projection operator zeros the correlation result beyond a rectangular region at the center. This approach unreasonably assumes that the translation is not significant between the two images being registered, or that the object of interest cannot lie away from the rectangular region at the center. Figure 3(*left*) shows the surface obtained after applying the projection operator on the PC performed between the original search image and template [Fig. 1(*left*)]. Since the object and the corresponding true peak were away from the rectangular support at the center, the true peak was set to zero and there remained only the false peaks around the center in the



Fig. 1. Blackman window modulation distorts images. (*left*) Original search image and template; (*right*) Modulated search image and template.

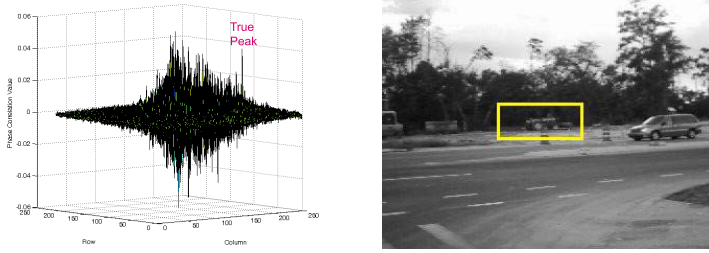


Fig. 2. Blackman window causes various false peaks in the correlation surface (if the object is away from the central region) resulting in false detection. (*left*) PC of the Blackman window modulated images indicating that the true peak is smaller than the false peaks and the highest false peak is at (105, 117) with magnitude only 0.0524; (*right*) Best-match rectangle with top-left position at (105, 117).

surface. Figure 3 (*right*) shows that the highest peak in the processed surface does not correspond to the vehicle.

In order to attenuate the false peaks and amplify the true peak, we propose an effective method that (a) slightly extends the search image and the template with new pixels having gradually decaying values (without changing the actual content of the images), (b) pads the extended images with zeros (without introducing any discontinuity), and (c) phase correlate the resulting images.

The next section describes the proposed technique of extending an image. Sect. 3 compares the proposed method with the two techniques mentioned in this section for cluttered, noisy, blurred, and slightly rotated scenes having reflection effect, and finally Sect. 4 concludes the paper.

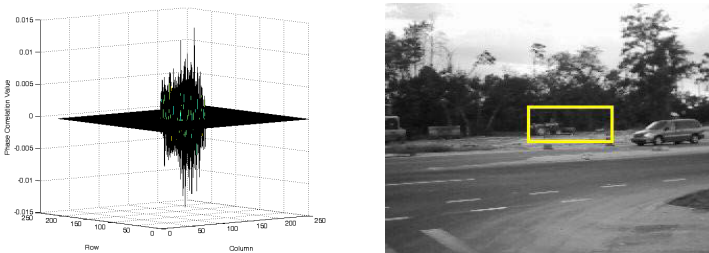


Fig. 3. The projection operator zeros the PC response [of the original images shown in Fig. 1(*left*)] beyond a rectangular region (typically 21×21) at the center, nullifying the true peak at the boundary, and resulting in a false detection. (*left*) The resulting surface shows the highest (false) peak at location (104, 141) with magnitude only 0.014; (*right*) Incorrect detection corresponding to the highest available peak.

2 Decaying Extension of an Image

We propose to extend an image using three steps as described in the following sub-sections.

2.1 Initializing the Extended Image

Let the original image be denoted by I having width, W , and height, H . We want to extend the image from every direction by δ pixels, as shown in Fig. 4(left). Thus, the size of the extended image, I_e , becomes $W_e \times H_e$, where $W_e = W + 2\delta$ and $H_e = H + 2\delta$. If $\delta < 4$, the discontinuities are not eliminated adequately, resulting in unsatisfactory attenuation of the false peaks. On the contrary, if $\delta \gg 4$, the discontinuities are eliminated very smoothly, but the computation time taken by the PC between the larger images is increased accordingly. In order to remain in the safe side, we set $\delta = 5$ in all of our experiments to have robust image matching without significantly increasing the computation time. We initialize every pixel in the extended regions with the value of its nearest boundary pixel in the original image. The process of initialization of the extended image can be described mathematically, as:

$$I_e(x, y) = \begin{cases} I(0, 0) & \text{if } 0 \leq x \leq \delta - 1 \text{ and } 0 \leq y \leq \delta - 1, \\ I(x - \delta, y - \delta) & \text{if } \delta \leq x \leq W + \delta - 1 \text{ and } \delta \leq y \leq H + \delta - 1, \\ I(0, H - 1) & \text{if } 0 \leq x \leq \delta - 1 \text{ and } H \leq y \leq H + \delta - 1, \\ I(W - 1, 0) & \text{if } W \leq x \leq W + \delta - 1 \text{ and } 0 \leq y \leq \delta - 1, \\ I(W - 1, H - 1) & \text{if } W + \delta \leq x \leq W_e - 1 \text{ and } H + \delta \leq y \leq H_e - 1. \end{cases} \quad (2)$$

2.2 Generating the Decaying Weights

Once the extended image is initialized, we proceed to generate the decaying weights for the new pixels. We obtain the weights using a Gaussian function because of its smooth and symmetric behavior.

Consider a \mathbb{R}^K column vector, where $K = 2\delta + 1$, containing the Gaussian weights: $g(k) = \exp[-(1/2)\{(k - \delta)^2/\sigma^2\}]$, where $k = 0, 1, \dots, K - 1$, and the standard deviation, σ , controls the spread of the Gaussian function. If the value of σ is too large, the function will go down too slow, and the boundary elements will be too high from zero, resulting in a sharp discontinuity during the zero-padding operation. On the contrary, if its value is too low, the function will go down abruptly to near-zero value, resulting in a discontinuity even before the zero-padding operation. We want to have the function, that goes down smoothly and becomes near-zero right at the boundary elements of the vector. We achieve this objective by computing an appropriate value of σ automatically according to the size of the vector, using: $\sigma = 0.3[(K/2) - 1] + 0.8$, as in [8]. Then, we split the vector g into its top and bottom halves (i.e. \mathbb{R}^δ vectors), as: $g_t(k) = g(k)$ and $g_b(k) = g(k + \delta + 1)$, $\forall k = 0, 1, \dots, \delta - 1$.

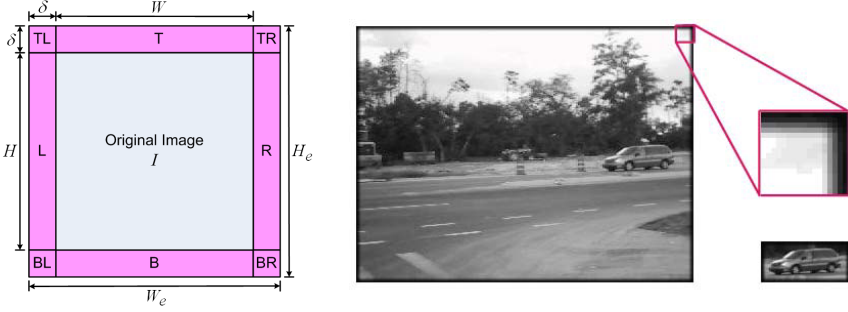


Fig. 4. Decaying extension of an image. (*left*) Structure of an extended image, I_e , where T = Top, L = Left, R = Right, and B = Bottom. (*right*) Search image and template from Fig. 1(*left*) extended using $\delta = 5$ with new pixels having decaying magnitude down to near-zero, as can be observed easily in the zoomed-in corner region.

Now, consider a $\mathbb{R}^{K \times K}$ matrix containing the weights obtained by a 2D Gaussian function, as:

$$g(k_x, k_y) = \exp \left[-\frac{1}{2} \left\{ \frac{(k_x - \delta)^2 + (k_y - \delta)^2}{\sigma^2} \right\} \right], \quad \forall k_x, k_y = 0, 1, \dots, \delta - 1. \quad (3)$$

The value of σ is the same as computed previously. Then, we split the 2-D Gaussian function into its four quarters (i.e. $\mathbb{R}^{\delta \times \delta}$ matrices). The top-left, top-right, bottom-left, and bottom-right quarters are obtained as: $G_{tl}(k_x, k_y) = G(k_x, k_y)$, $G_{tr}(k_x, k_y) = G(k_x + \delta + 1, k_y + \delta + 1)$, $G_{bl}(k_x, k_y) = G(k_x, k_y + \delta + 1)$, and $G_{br}(k_x, k_y) = G(k_x + \delta + 1, k_y)$, $\forall k_x, k_y = 0, 1, \dots, \delta - 1$.

2.3 Applying the Decaying Weights

We multiply the g_t vector element-by-element with every column in the \mathbf{T} patch in the initialized extended image [see Fig. 4(*left*)], g_b vector with every column in the \mathbf{B} patch, g_t^T row-vector with every row in the \mathbf{L} patch, g_b^T row-vector with every row in the \mathbf{R} patch, G_{tl} matrix with the \mathbf{TL} patch, G_{tr} matrix with the \mathbf{TR} patch, G_{bl} matrix with the \mathbf{BL} patch, and G_{br} matrix with the \mathbf{BR} patch. Figure 4(*right*) illustrates the extended versions of the search image and template shown in Fig. 1(*left*). It may be noted that all the content of the original images is intact in the extended images, and that only the pixels at the external regions are gradually getting more and more black as depicted by the zoomed-in top-right corner of the extended search image.

3 Experimental Results

When we zero-padded the extended versions of the search image and the template [Fig. 4(*right*)] and phase correlated the resulting images with each other,

we obtained a single dominant peak in the correlation surface [Fig. 5(*left*)] corresponding to the true location of the vehicle in the search image [Fig. 5(*right*)]. We can see that the proposed technique has drastically attenuated all the false peaks in the PC response. Moreover, since the absolute of the sum of all the peaks in the PC response is always 1 [4], the energy released by the attenuating false peaks combines with that of the true peak, resulting in significant amplification of the true peak. This is a significant improvement introduced by the proposed approach in the PC response as compared to the response obtained using Blackman window (Fig. 2) or projection operator (Fig. 3).

Furthermore, we compared the proposed approach with the two techniques also for the template matching based object tracking application on numerous real videos. The size of the rectangular support of the projection operator [5] was enlarged from 21×21 to one-half of the frame size in order to give it some more advantage. We show two image sequences due to space constraint. Figure 6 illustrates a very noisy, shaky, and blurred video of a ground vehicle recorded from a UAV (unmanned aerial vehicle). The ground vehicle is being continuously translated and slightly rotated, and there is a noticeable reflection effect at the end of the video. The template – shown at the bottom-left of every frame – is kept constant to create a more challenging scenario. We can observe that only the proposed approach is successful in tracking the ground vehicle persistently in this video even when the vehicle is far from the central region of the frame. Figure 7 depicts the results of the three techniques on a cluttered and noisy video (OneShopOneWait2cor.avi) from CAVIAR database¹. Since the PC is only for estimating the translation between two images, it is not supposed to handle the significant variation in the appearance of the walking woman to be tracked in this video. Therefore, we smoothly update the template in each iteration of the techniques being compared, using α -tracker template updating method [9] (with $\alpha = 0.25$).

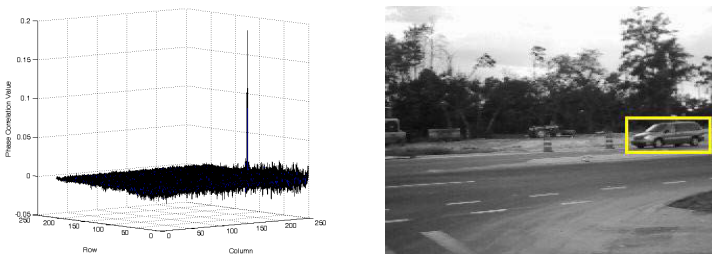


Fig. 5. Effect of the decaying extension of the images. (*left*) Phase correlation of the extended versions of the search image and the template that shows a single dominant peak at the true location of the best match, (111, 237), having magnitude amplified to 0.1836, and all the false peaks attenuated to a very small magnitude; (*right*) Correct detection of the vehicle at (111, 237) position.

¹ Available at <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>

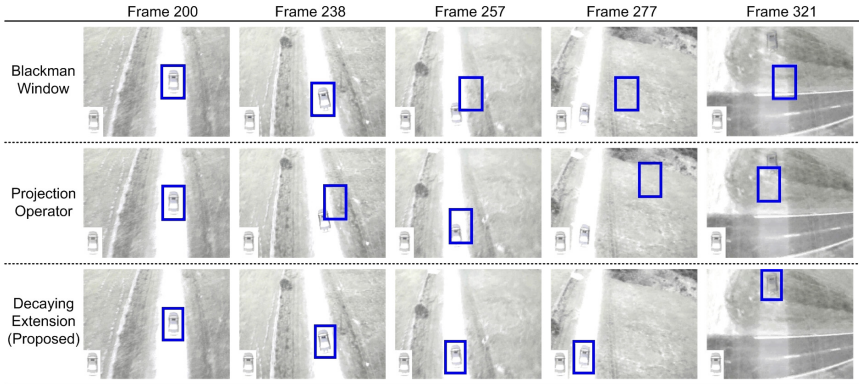


Fig. 6. Tracking a ground vehicle in a noisy, shaky, and blurred UAV (unmanned aerial vehicle) video containing slight rotations and significant reflection effect, when the template (shown at the bottom left of every frame) is not updated. Only the proposed method tracks the ground vehicle persistently.

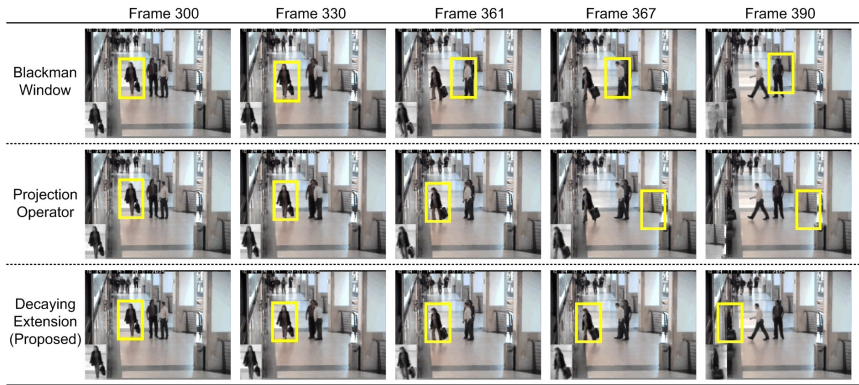


Fig. 7. Tracking a woman in a noisy and cluttered video from CAVIAR database, when the template (shown at the bottom-left of every frame) is iteratively updated. The proposed method tracks the woman robustly, while the other methods do not.

The adaptive template is shown at the bottom-left of every frame. It can be again observed that the proposed method outperforms the other two techniques in tracking the walking woman robustly even when she goes away from the central region of the frame.

We also compared the computation time of each technique by using a fixed 512×512 search image and varying the template size from 25×25 to 250×250 . In all cases, the proposed method took about as much time as the standard PC or the projection operator based PC, and only one-half of the time taken by the Blackman window based PC.

4 Conclusion

Phase correlation of the original images sometimes results in false peaks having magnitude higher than that of the true peak. This is mainly due to the discontinuities resulting from the mismatch between the elements at the opposite boundaries of each image and the zero-padding operation (necessary for obtaining the true correlation). We propose an efficient and effective technique that extends the original images with new pixels having decaying values, so that the discontinuities are eliminated without distorting the actual scene. Our method outperforms two previous techniques, even if the object is slightly rotated and away from the central region of the image suffering from significant reflection, noise, blur, and clutter. The accurate and robust pixel-level registration result from the proposed approach can definitely increase the true positive rate and the precision of the sub-pixel registration technique (described in [10]), that increase the number of samples in the phase correlation response and fill the in-between samples using some interpolation method.

References

1. Foroosh, H., Zerubia, J., Berthod, M.: Extension of phase correlation to sub-pixel registration. *IEEE Trans. on Image Processing* 11(3), 188–200 (2002)
2. Kuglin, D., Hines, D.: The phase correlation image alignment method. In: *Proc. IEEE Conference on Cybernetics and Society*, pp. 163–165 (1975)
3. Manduchi, R., Mian, G.: Accuracy analysis for correlation-based image registration algorithms. In: *Proc. ISCAS 1993: Int'l Symposium on Circuits and Systems*, vol. 1, pp. 834–837 (1993)
4. Stone, H., Tao, B., McGuire, M.: Analysis of image registration noise due to rotationally dependent aliasing. *Journal of Visual Communication and Image Representation* 14(3), 114–135 (2003)
5. Keller, Y., Averbuch, A., Miller, O.: Robust phase correlation. In: *Proc. 17th IEEE International Conference on Pattern Recognition* (2004)
6. Ahmed, J., Jafri, M., Shah, M., Akbar, M.: Real-time edge-enhanced dynamic correlation and predictive open-loop car-following control for robust tracking. *Machine Vis. and App.* 19(1), 1–25 (2008)
7. Oppenheim, A., Schaffer, R., Buck, J.: *Discrete-Time Signal Proc.*, 2nd edn. Prentice Hall, Englewood Cliffs (1999)
8. Bradski, G., Kaehler, A., Pisarevsky, V.: Learning-based computer vision with open source computer vision library. *Intel Tech. J.* 9(3), 118–131 (2005)
9. Blackman, S., Popoli, R.: *Design and Analysis of Modern Tracking Systems*, 2nd edn. Artech House, Boston (1999)
10. Tian, Q., Huhns, M.N.: Algorithms for subpixel registration. *Comp. Vision, Graph., and Image Proc.* 35, 220–223 (1986)