

Design of Correlation Filters for Pattern Recognition Using a Noisy Training Image

Pablo M. Aguilar-González and Vitaly Kober

Department of Computer Science, Centro de Investigación Científica y de Educación Superior de Ensenada,
Carretera Ensenada-Tijuana No. 3918, Zona Playitas, C.P. 22860, Ensenada, B.C.,
México
{paguilar, vkober}@cicese.mx
<http://www.cicese.edu.mx/>

Abstract. Correlation filters for object detection and location estimation are commonly designed assuming the shape and graylevel structure of the object of interest are explicitly available. In this work we propose the design of correlation filters when the appearance of the target is given in a single training image. The target is assumed to be embedded in a cluttered background and the image is assumed to be corrupted by additive sensor noise. The designed filters are used to detect the target in an input scene modeled by the nonoverlapping signal model. An optimal correlation filter, with respect to the peak-to-output energy ratio criterion, is proposed for object detection and location estimation. We also present estimation techniques for the required parameters. Computer simulation results obtained with the proposed filters are presented and compared with those of common correlation filters.

Keywords: correlation filters, pattern recognition.

1 Introduction

Since the introduction of the matched filter [1], correlation filters have been extensively used for pattern recognition [2-15]. Two tasks of interest in pattern recognition are detection of a target and the estimation of its location in an observed scene. With the help of correlation filters such tasks can be solved in two steps: detection is carried out by locating the highest peak in the filter output; then, the coordinates of the peak are taken as estimations of the position of the target in the observed scene [2].

The performance of correlation filters can be evaluated by quantitative performance criteria such as signal-to-noise ratio (SNR), peak sharpness, discrimination capability (DC), and probability of false alarms [3, 4]. Location accuracy can be described in terms of the variance of location errors [5, 6]. Correlation filters are designed by means of analytical optimization of one or more of these criteria. In order to perform such optimization, a mathematical model of the scene is chosen. The additive signal model is used when an input scene contains

a target distorted by additive noise. Optimizing the SNR criterion for the this model leads to the matched filter (MF) [1], while minimizing the probability of false alarms yields the optimal filter (OF) [4]. The nonoverlapping signal model is used when an opaque target is placed over a background that is spatially disjoint. Several filters have been derived for this scene model [6,7,8,9]. Maximizing the ratio of the square of the expected value of the correlation peak to the average output variance leads to the generalized matched filter [7]. Maximizing the peak-to-output energy ratio (POE) yields the generalized optimum filter [7] (GOF).

Because correlation filters are designed using the expected appearance of the target, their performance degrades rapidly if the target appears distorted in the scene. Distortions can be caused by changes of scale, rotation or perspective; blurring or defocusing, or incomplete information about the appearance of the target. Several correlation filters were proposed that take into account linear degradations of the input scene and the target [10]. Composite filters have been used to consider geometric distortions [11, 12, 13]. However, the design of these filters is done assuming that a knowledge of the target shape is explicitly known. In practical situations, the target may be given in a noisy reference image with a cluttered background. Recently [14, 15], a signal model was introduced that accounts for additive noise in the image used for filter design. In this paper, we extend that work to account for the presence of a nonoverlapping background in a training image that is corrupted by additive noise. We derive a correlation filter optimized with respect to the POE criterion. The performance of this filter is compared to that of classical correlation filters for the nonoverlapping signal model.

2 Design of Filters

The nonoverlapping signal model is used for the reference image and the input scene. We use one-dimensional notation for simplicity. Integrals are taken between infinite limits. Throughout this section we use the same notation for a random process and its realization.

Formally, the input scene and the reference image are given, respectively, by

$$s(x) = t(x - x_s) + b_s(x) \bar{w}(x - x_s) + n_s(x) , \quad (1)$$

$$r(x) = t(x - x_r) + b_r(x) \bar{w}(x - x_r) + n_r(x) , \quad (2)$$

where $t(x)$ is the target, located at unknown coordinates x_s and x_r in the input scene $s(x)$ and in the reference image $r(x)$, respectively; $b_s(x)$ and $b_r(x)$ are the disjoint backgrounds, and $n_r(x)$ and $n_s(x)$ are the additive noise signals due to sensor noise. $\bar{w}(x)$ is the inverse support region for the target, that is, it takes a value of unity outside the target area and a value of zero inside. We make the following assumptions:

- The nonoverlapping backgrounds, $b_s(x)$ and $b_r(x)$, are treated as the realization of stationary random processes that have mean values μ_s and μ_r , respectively, and have power spectral densities $B_s^0(\omega)$ and $B_r^0(\omega)$, respectively.
- The additive noise processes $n_s(x)$ and $n_r(x)$ are assumed to be stationary random processes with mean zero and spectral densities $N_s(\omega)$ and $N_r(\omega)$, respectively.
- All random processes and random variables are treated as statistically independent.
- $s(x)$ and $r(x)$ are real-valued images with Fourier transforms $S(\omega)$ and $R(\omega)$, respectively.

The goal of the filter design process is to obtain a filter frequency response $H(\omega)$ of the form

$$H(\omega) = A(\omega) R^*(\omega) , \tag{3}$$

where $A(\omega)$ is a deterministic function and $*$ denotes complex conjugate. Since the obtained filter frequency response contains non deterministic components, the filter expression represents a bank of transfer functions. A specific realization of the filter is fixed by the realization of the noises processes $b_r(x)$ and $n_r(x)$.

In (2), the location of the target in the reference image x_r , is unknown and not necessarily located at the origin. Therefore, the correlation peak is expected to be present at the coordinate $x_0 = x_s - x_r$. If x_r is close to 0, the location estimation of the target in the input scene will be in the close vicinity of its true location in the input scene. Even if the exact location of the target cannot be precisely determined, the relative position is useful for applications such as tracking [16], where the goal is to determine the relative movement of the target.

We derive the modified generalized optimum filter for the *nonoverlapping-nonoverlapping* model (GOF_{NN}) by maximizing the POE criterion, formally defined as

$$\text{POE} = |\text{E}\{y(x_0)\}|^2 / \text{E}\{\overline{|y(x)|^2}\} , \tag{4}$$

where $\text{E}\{\cdot\}$ denotes statistical averaging and the over-bar denotes spatial averaging, i.e. $\overline{|y(x)|^2} = (1/L) \int |y(x)|^2 dx$, L is the spatial extent of the signal $y(x)$. The expected value of the correlation peak in the filter output plane is

$$\text{E}\{y(x_0)\} = (2\pi L)^{-1} \int A(\omega) \text{E}\left\{(R(\omega) e^{i\omega x_r})^* S(\omega) e^{i\omega x_s}\right\} d\omega . \tag{5}$$

The denominator of the POE represents the average energy in the correlation plane. It can be calculated as

$$\text{E}\{\overline{|y(x)|^2}\} = (2\pi L)^{-1} \int |A(\omega)|^2 \text{E}\{|R^*(\omega) S(\omega)|^2\} dx . \tag{6}$$

Using (5) and (6) in (4) we get

$$\text{POE} = \frac{L \left| \int A(\omega) \text{E}\left\{(R(\omega) e^{i\omega x_r})^* S(\omega) e^{i\omega x_s}\right\} d\omega \right|^2}{2\pi \int |A(\omega)|^2 \text{E}\{|R^*(\omega) S(\omega)|^2\} dx} . \tag{7}$$

Applying the Cauchy-Schwartz inequality to (7) and substituting the optimum value for $A(\omega)$ into (3), we obtain the following frequency response for the GOF_{NN} :

$$\text{GOF}_{\text{NN}}^*(\omega) = \frac{\text{E} \{ R(\omega) e^{i\omega x_r} \}^* \text{E} \{ S(\omega) e^{i\omega x_s} \} R(\omega)}{\text{E} \{ |R(\omega)|^2 \} \text{E} \{ |S(\omega)|^2 \}}. \quad (8)$$

The expected value of the power spectra of the input scene and reference image can be calculated as follows:

$$\text{E} \{ |S(\omega)|^2 \} = |T(\omega) + \mu_s \bar{W}(\omega)|^2 + \frac{1}{2\pi} B_s^0(\omega) \bullet |\bar{W}(\omega)|^2 + N_s(\omega), \quad (9)$$

$$\text{E} \{ |R(\omega)|^2 \} = |T(\omega) + \mu_r \bar{W}(\omega)|^2 + \frac{1}{2\pi} B_r^0(\omega) \bullet |\bar{W}(\omega)|^2 + N_r(\omega), \quad (10)$$

where \bullet denotes the convolution operation and $\bar{W}(\omega)$ is the Fourier transform of the inverse support function. It can be seen that the obtained filter requires knowledge of the Fourier transform of the target and its support function. However, in the problem model we assume that this information is not available. Therefore, estimations need to be designed from the available information. A smoothing Wiener filter [17] can be used to partially suppress the background and attenuate the effects of noise in the reference image. After filtering, we can apply a threshold to the resulting image and obtain an approximate support function as follows:

$$\hat{w}(x) = \begin{cases} 1 & \mu_r > \mu_t \text{ and } \tilde{r}(x) \geq \tau(t, b_r, n_r) \\ 1 & \mu_r < \mu_t \text{ and } \tilde{r}(x) \leq \tau(t, b_r, n_r) \\ 0 & \text{otherwise} \end{cases}, \quad (11)$$

where $\tilde{r}(x)$ denotes the reference image after Wiener filtering, $\hat{w}(x)$ denotes the estimation of the inverse support function and $\tau(t, b_r, n_r)$ is the optimum threshold for separating the distributions of $t(x) + n_r(x)$ and $b_r(x) + n_r(x)$ after filtering. We make the simplifying assumption that the noise processes are approximately normal after filtering. Therefore, we can estimate their statistics, using the spectral densities we assumed known, and use them calculate the optimum threshold. When the statistics of the target are known, they can be used to improve the threshold calculation. If the statistics of the target are unknown, the optimum threshold can be determined with respect to the statistics of the background and additive noise.

Once an estimation of the inverse support function is obtained, we can estimate the expected value of the input scene as

$$T(\omega) + \mu_s \bar{W}(\omega) \approx \tilde{R}(\omega) + (\mu_s - \mu_r) \hat{W}(\omega), \quad (12)$$

where $\tilde{R}(\omega)$ is the Fourier transform of the reference image after Wiener filtering. This estimation is then used to design the frequency response of the GOF_{NN} .

It is worth noting that when there is no noise present in the reference image, the GOF_{NN} is equal to the GOF.

3 Computer Simulations

In this section we present computer simulation results. The performance of the proposed filters is evaluated in terms of discrimination capability (DC) and location errors (LE). The DC is formally defined [4] as the ability of a filter to distinguish a target from other objects in the scene. The DC is formally defined as

$$DC = 1 - \frac{|C_B|^2}{|C_T|^2}, \quad (13)$$

where C_B is the maximum value in the correlation plane over the background area, and C_T is the maximum value in the correlation plane over the target area in the scene. The background area and the target area are complementary. Ideally, values of the DC should be close to unity, indicating a good capacity to discriminate the target against unwanted objects in the background. Negative values of the DC indicate a failure to detect the target. The location accuracy can be characterized by means of the location errors defined as

$$LE = \sqrt{(x_T - \hat{x}_T)^2 + (y_T - \hat{y}_T)^2}, \quad (14)$$

where (x_T, y_T) are the exact coordinates where the correlation peak is expected to occur and (\hat{x}_T, \hat{y}_T) are the coordinates where it is actually located after filtering.

The size of the images used in the experiments is 256×256 pixels with intensity values in the range $[0-255]$. We use the image of the toy car shown in Fig. 1(a) as the target. Two types of backgrounds are used: deterministic and stochastic. The stochastic backgrounds are realizations of colored random processes with correlation coefficients of 0.70 and 0.95 for the reference image and input scene, respectively. The size of the target is 62×50 pixels. The target mean and Std. Dev. have values 105 and 45, respectively. To guarantee statistically correct results, 30 statistical trials of each experiment for different positions of a target and realization of random processes were carried out. All scenes are corrupted by additive white Gaussian noise.

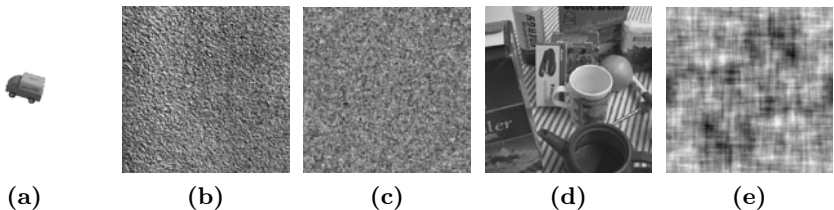


Fig. 1. (a) The target used in the experiments, (b) deterministic reference image background, (c) sample stochastic reference image background, (d) deterministic scene background, and (e) sample stochastic scene background

For comparison purposes, we include the ideal GOF, designed with all known parameters, to establish an upper bound on performance; the proposed GOF_{NN} filter when using $\hat{w}(x)$ as the estimation of the inverse support function; and a switching version of the GOF_{NN} , labeled sGOF_{NN} , which is designed by using only part of the GOF_{NN} . The GOF_{NN} can be regarded as the sum of two filters, corresponding to the part of the target and the part of the expected value of the background; that is, we can write the frequency response of the GOF_{NN} as

$$\text{GOF}_{\text{NN}}(\omega) = G(\omega)T^*(\omega) + G(\omega)\mu_s\bar{W}(\omega), \quad (15)$$

for a suitably defined $G(\omega)$. This is effectively an approximation of the ideal GOF when the reference image is corrupted by noise. However, when the estimation of the support function is degraded, it proves better to design a switching filter that uses either or both terms of (15) depending on the statistics of the target and the noise processes in a given problem instance.

We first perform experiments using synthetically generated backgrounds while varying the location of the target in the scene. We need to determine how robust the filters are with respect to the noise in the reference image. The simulation results are shown in Fig. 2 when the mean and Std. Dev. of the background in the reference image are 100 and 40, respectively. Since the statistics of the background and of the target are so similar, the estimations of the support function are severely degraded. The performance of the GOF is constant because this filter is designed with all parameters known. It can be seen that detection of the target is possible in the presence of noise with a Std. Dev. of up to 20. Location errors are small values of the Std. Dev. up to 25. In this case, the performance of the switching filter is not significantly better since there is very little of the support function available for the design of the filters.

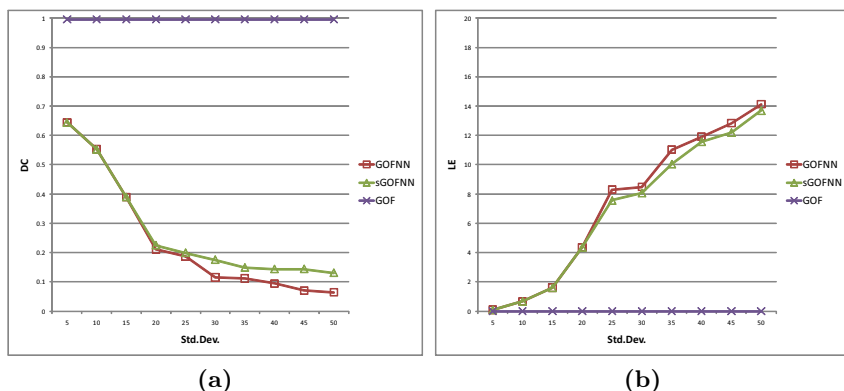


Fig. 2. Performance of filters in terms of (a) DC and (b) LE while varying the Std. Dev. of the reference image noise. The input scene is corrupted by additive noise with Std. Dev. of 10.

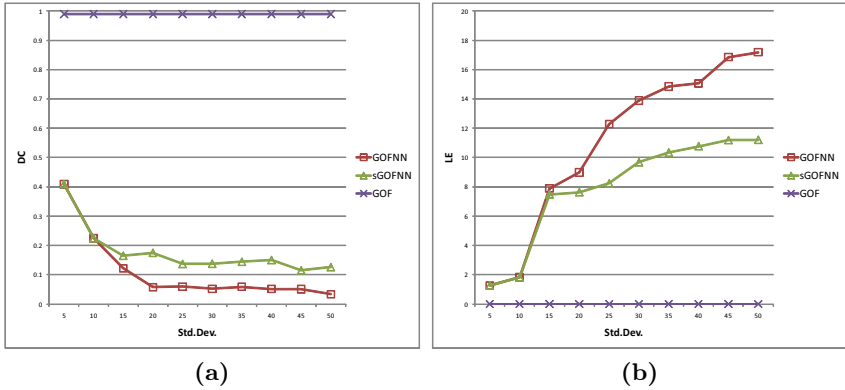


Fig. 3. Performance of filters in terms of (a) DC and (b) LE while varying the Std. Dev. of the reference image noise using deterministic backgrounds. The input scene is corrupted by additive noise with Std. Dev. of 10.

While stochastic backgrounds closely match the signal model, it is interesting to investigate the performance of the proposed filters when natural images are used as backgrounds. The simulation results are shown in Fig. 3. Because of the increased complexity of the backgrounds, it becomes harder to detect and locate the target. Thus, the performance for the GOF_{NN} is lower than when using stochastic backgrounds. However, in this case the switching filter consistently outperforms the GOF_{NN}. We can say that the target can be detected when the noise has a Std. Dev. of up to 10, while location errors are small up to a Std. Dev. of 20. When the noise levels increase, we can no longer be consider the detection results to be reliable.

4 Conclusions

In this paper were proposed a novel filter for detecting and locating a target in nonoverlapping background noise by using a noisy image. The filter expression is derived from a new signal model that accounts for the presence of a cluttered background in the training image. Filter instances are designed using only information available in the reference image and statistical information of the noise processes in the model. Estimations were given for the parameters assumed unknown in the model. A switching filter was also proposed that, under certain conditions, performs better than the approximation of the ideal filter. With the help of computer simulations, we showed that the filters along with the proposed estimations, yield good detection results in the presence of moderate levels of noise.

References

1. VanderLugt, A.: Signal detection by complex spatial filtering. *IEEE Transactions on Information Theory* 10(2), 139–145 (1964)
2. Kumar, B.V.K.V., Mahalanobis, A., Juday, R.: *Correlation pattern recognition*. Cambridge University Press, Cambridge (2005)
3. Kumar, B.V.K.V., Hassebrook, L.: Performance measures for correlation filters. *Applied Optics* 29(20), 2997–3006 (1990)
4. Yaroslavsky, L.P.: The theory of optimal methods for localization of objects in pictures. In: Wolf, E. (ed.) *Progress in Optics*, pp. 145–201. Elsevier, Amsterdam (1993)
5. Kumar, B.V.K.V., Dickey, F.M., DeLaurentis, J.M.: Correlation filters minimizing peak location errors. *Journal of the Optical Society of America A* 9(5), 678–682 (1992)
6. Kober, V., Campos, J.: Accuracy of location measurement of a noisy target in a nonoverlapping background. *Journal of the Optical Society of America A* 13(8), 1653–1666 (1996)
7. Javidi, B., Wang, J.: Design of filters to detect a noisy target in nonoverlapping background noise. *Journal of the Optical Society of America A* 11(10), 2604–2612 (1994)
8. Javidi, B., Zhang, G., Parchekani, F.: Minimum-mean-square-error filters for detecting a noisy target in background noise. *Applied Optics* 35, 6964–6975 (1996)
9. Javidi, B.: *Real-Time Optical Information Processing*. Academic Press, London (1994)
10. Ramos-Michel, E.M., Kober, V.: Design of correlation filters for recognition of linearly distorted objects in linearly degraded scenes. *Journal of the Optical Society of America A* 24(11), 3403–3417 (2007)
11. Mahalanobis, A., VijayaKumar, B.V.K., Song, S., Sims, S.R.F., Epperson, J.F.: Unconstrained correlation filters. *Applied Optics* 33(17), 3751–3759 (1994)
12. González-Fraga, J., Kober, V., Álvarez-Borrego, J.: Adaptive synthetic discriminant function filters for pattern recognition. *Optical Engineering* 45, 057005 (2006)
13. Ramos-Michel, E.M., Kober, V.: Adaptive composite filters for pattern recognition in linearly degraded and noisy scenes. *Optical Engineering* 47, 047204 (2008)
14. Aguilar-González, P.M., Kober, V.: Correlation filters for pattern recognition using a noisy reference. In: Ruiz-Shulcloper, J., Kropatsch, W.G. (eds.) *CIARP 2008*. LNCS, vol. 5197, pp. 38–45. Springer, Heidelberg (2008)
15. Aguilar-González, P.M., Kober, V.: Correlation pattern recognition in nonoverlapping scene using a noisy reference. In: Bayro-Corrochano, E., Eklundh, J.-O. (eds.) *CIARP 2009*. LNCS, vol. 5856, pp. 555–562. Springer, Heidelberg (2009)
16. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. *ACM Computing Surveys* 38(4) (2006)
17. Pratt, W.K.: *Digital Image Processing*. John Wiley & Sons, Chichester (2007)