

Object Tracking in Nonuniform Illumination Using Space-Variant Correlation Filters

Víctor Hugo Díaz-Ramírez¹, Kenia Picos¹, and Vitaly Kober²

¹ Instituto Politécnico Nacional - CITEDI, Ave. del Parque 1310, Mesa de Otay,
Tijuana B.C. 22510, Mexico

² Department of Computer Science, Division of Applied Physics, CICESE, Carretera
Ensenada-Tijuana 3918, Zona Playitas, Ensenada B.C. 22860, Mexico
vdiazr@ipn.mx, kpicos@citedi.mx, vkober@cicese.mx

Abstract. A reliable system for recognition and tracking of a moving target in nonuniformly illuminated scenes is presented. The system employs a filter bank of space-variant correlation filters adapted to local statistical parameters of the observed scene in each frame. When a scene frame is captured, a fragment of interest is constructed in the frame around predicted location of the target based on a kinematic model. The fragment is firstly pointwise processed to correct the illumination. Afterwards, the state of the target is estimated from the restored fragment by employing a bank of space-variant correlation filters. The performance of the proposed system in terms of object recognition and tracking is tested with nonuniformly illuminated and noisy scenes. The results are compared with those of common techniques based on correlation filtering.

1 Introduction

Nowadays, object recognition attracts research interests due to the need of developing imaging systems to improve activities such as video surveillance, vehicle navigation, object tracking, among others. Object recognition consists in identification of the target within a observed scene and in estimation of the target's exact coordinates. When a target moves across an environment, the appearance of the target with respect to the observer varies with time. Actually, target tracking consists in estimation of the target trajectory in the observed scene while the object moves. Target tracking can be solved by detecting the object in successive frames and by finding the correspondence between object states across scene frames. Commonly target tracking is performed by employing feature-based methods and state-space models. This approach yields good results when the target suffers from occlusions and geometrical modifications such as rotation and scaling. However, when the target exits and reenters to the observed scene and when the scene is degraded by additive and high cluttering background noise feature-based methods face some difficulties. A detailed review of tracking algorithms can be found in [1]. An attractive option for target recognition is given by correlation filtering. Correlation filters possess a good

mathematical background, and they can be implemented by exploiting massive parallelism either in hybrid opto-digital correlators [2,3] or in digital hardware such as graphic processing units (GPU) [4]. A correlation filter is a linear system where the coordinates of the system output maximum are estimates of the target coordinates in the observed scene [5,6]. Correlation filters can recognize objects in cluttered and noisy environments, even when targets suffer from geometrical distortions [7]. In this sense, the problem of target tracking can be addressed by applying correlation filters to multiple frames. Recently, several proposals have been suggested to perform target tracking with the help of correlation filters [3]. In this work, we propose a reliable system for recognition and tracking of a moving target in nonuniformly illuminated scenes using a filter bank of space-variant correlation filters. The frequency response of the filters are varied accordingly to local statistical parameters of the input signal in each frame. First, the proposed system performs a pointwise illumination correction to the input frame. Next, the target is detected from the restored frame by analyzing the correlation peaks obtained at the outputs of the filter bank. Then, the system predicts the state of the target for subsequent frame, and based on the prediction creates a fragment of interest in the input frame and modifies the number of filters in the bank using predicted orientation of the target in the current frame. Both location and orientation predictions are calculated by analyzing current and past state estimates and by taking into account a two-dimensional motion model. The resultant system is able to track a moving target in nonuniform illumination with reduced false alarms probabilities by focusing the processing only on a small fragment. The paper is organized as follows. Section 2 presents the approach used for target recognition in nonuniformly illuminated and noisy scenes. Section 3, explains the system proposed for object recognition and tracking. Section 4 presents the results obtained with the proposed system by testing its performance in nonuniformly illuminated scenes. Finally, section 5 summarizes our conclusions.

2 Recognition of a Target in Nonuniformly Illuminated and Noisy Scenes

Let $f(x, y)$ be an input scene composed by a target $t(x, y)$ located at unknown coordinates (α, β) and embedded into a disjoint background $b(x, y)$. The scene is assumed to be corrupted by a nonuniform illumination function $d(x, y)$ and with zero-mean additive noise $n(x, y)$. The input scene is expressed by

$$f(x, y) = [t(x - \alpha, y - \beta) + \bar{w}(x - \alpha, y - \beta)b(x, y)]d(x, y) + n(x, y), \quad (1)$$

where $\bar{w}(x)$ is the inverse region of support of the target given by unity outside the target area and zero elsewhere. We assume that $d(x, y)$ is a slow varying function, which is approximately constant in a small region (for instance, the region of support of the target $w(x, y)$). Note that this is the case of Lambertian surfaces. In order to correct the illumination of the input frame we perform the following pointwise processing

$$\hat{f}(x, y) = r_{x,y}f(x, y) + s_{x,y}, \quad (2)$$

where $r_{x,y}$ and $s_{x,y}$ represent unknown restoration coefficients. The mean squared error (MSE) between the restored frame $\hat{f}(x, y)$ and the reference image of the target $t(x, y)$, is given by

$$MSE_{\alpha,\beta} = \sum_{x,y \in w} \sum (r_{\alpha,\beta} f(x + \alpha, y + \beta) + s_{\alpha,\beta} - t(x, y))^2. \tag{3}$$

By minimization of the $MSE_{\alpha,\beta}$, we get

$$s_{\alpha,\beta} = \mu_t - r_{\alpha,\beta} \mu_f(\alpha, \beta), \tag{4}$$

and

$$r_{\alpha,\beta} = \frac{\frac{1}{N_w} \sum_{x,y \in w} t(x, y) f(x + \alpha, y + \beta) - \mu_t \mu_f(\alpha, \beta)}{\mu_{f^2}(\alpha, \beta) - \mu_f^2(\alpha, \beta)}, \tag{5}$$

where N_w is the number of signal elements inside $w(x, y)$, $\mu_t = \frac{1}{N_w} \sum_{x,y \in w} t(x, y)$, $\mu_f(\alpha, \beta) = \frac{1}{N_w} \sum_{x,y \in w} f(x + \alpha, y + \beta)$ and $\mu_{f^2}(\alpha, \beta) = \frac{1}{N_w} \sum_{x,y \in w} f^2(x + \alpha, y + \beta)$. Note that $\hat{f}(x, y)$ in Eq. (2) represents the input frame with approximately uniform illumination, i.e.,

$$\hat{f}(x, y) \approx t(x - \alpha, y - \beta) + \bar{w}(x - \alpha, y - \beta) b(x, y) + \tilde{n}(x, y), \tag{6}$$

where $\tilde{n}(x, y)$ is a nonstationary noise process.

Recognition of a Target in Additive and Nonoverlapping Noise

Here the goal is to recognize and to precisely estimate the location of the target from the nonoverlapping signal model of Eq. (6). In this case, the optimum filter with respect to the signal to noise ratio (SNR) and to the minimum variance of target's location error (LE) is the Generalized Matched Filter (GMF), whose frequency response is given by [8,6]

$$H^*(\mu, \nu) = \frac{T(\mu, \nu) + \mu_b \bar{W}(\mu, \nu)}{S_{b0}(\mu, \nu) \otimes |\bar{W}(\mu, \nu)|^2 + S_n(\mu, \nu)}. \tag{7}$$

In Eq. (7), μ_b and $S_{b0}(\mu, \nu)$ represent the mean value of the background $b(x, y)$ and the power spectral density of $b_0(x, y) = b(x, y) - \mu_b$, respectively. The terms $T(\mu, \nu)$, $\bar{W}(\mu, \nu)$ and $S_n(\mu, \nu)$ are the Fourier transform of the target, the Fourier transform of $\bar{w}(x, y)$ and the power spectral density of $n(x, y)$, respectively. It is important to realize that for real applications the terms $T(\mu, \nu)$ and $\bar{W}(\mu, \nu)$ are a-priori known. Nevertheless, the terms μ_b , $S_{b0}(\mu, \nu)$ and $S_n(\mu, \nu)$ are generally unknown and must be estimated.

Estimation of Nonoverlapping Noise Parameters

Assume that the target $t(x, y)$ is located inside a small fragment $\hat{f}_r(x, y)$ of the input frame and is embedded into the background $b_r(x, y)$. The mean value of the scene fragment can be computed as $\mu_{f_r} = \frac{1}{N_{\hat{f}_r}} \sum_{x,y \in \hat{f}_r} \hat{f}_r(x, y)$, where $N_{\hat{f}_r}$ is the

number of pixels in the fragment. Since the target is known and is contained in the fragment, the mean value of the background $b_r(x, y)$ can be estimated as

$$\hat{\mu}_{b_r} = \frac{\mu_{f_r} N_{f_r} - \mu_t N_w}{N_{f_r} - N_w}. \quad (8)$$

The sample variance of $b_r(x, y)$ can be computed by $\sigma_{b_r}^2 = \frac{1}{N_{b_r}} \sum_{x,y \in b_r} b_r^2(x, y) - \hat{\mu}_{b_r}^2$.

By noticing that for the disjoint model $\sum_{x,y \in b_r} b_r^2(x, y) = \sum_{x,y \in f_r} \hat{f}_r^2(x, y) - \sum_{x,y \in w} t^2(x, y)$,

and with the help of Eq. (8) the local variance of the background is estimated by

$$\hat{\sigma}_{b_r}^2 = \frac{1}{N_{f_r} - N_w} \left(\sum_{x,y \in f_r} \hat{f}_r^2(x, y) - \sum_{x,y \in w} t^2(x, y) \right) - \left(\frac{\mu_{f_r} N_{f_r} - \mu_t N_w}{N_{f_r} - N_w} \right)^2. \quad (9)$$

The parameters $\hat{\mu}_{b_r}$ and $\hat{\sigma}_{b_r}^2$ are used to estimate the power spectral density $S_{b_0}(\mu, \nu)$ required in Eq. (7). This is done by a separable exponential model of the covariance function, as follows:

$$\hat{S}_{b_0}(\mu, \nu) = \sum_{x,y \in f_r} \hat{\sigma}_{b_r}^2 \rho_x^{|x|} \rho_y^{|y|} \exp \left[-i2\pi (\mu x + \nu y) / N_{f_r} \right], \quad (10)$$

where ρ_x and ρ_y are the correlation coefficients of the background function in x and y directions.

Estimation of Additive Noise Parameters

Consider that the fragment $f_r(x, y)$ is corrupted by zero-mean additive white Gaussian noise $n_r(x, y)$. Assume that $f_r(x, y)$ and $n_r(x, y)$ are independent. The autocorrelation function of the noisy fragment can be approximated by $R(x, y) \approx R_{f_r}(x, y) + R_n(x, y)$, where $R_n(x, y) = \sigma_n^2 \delta(x, y)$ is the autocorrelation function of the noise, and $R_{f_r}(x, y)$ is the autocorrelation function of the fragment. Note that the noise variance σ_n^2 can be estimated by $\hat{\sigma}_n^2 = R(0, 0) - R_{f_r}(0, 0)$, where $R_{f_r}(0, 0)$ is unknown. We see that $R_n(x, y) = 0, \forall (x, y) \neq 0$. Thus, the values $\{R(x, y) | (x, y) \neq 0\}$ can be used as estimates of $R_{f_r}(x, y)$. To estimate the noise variance σ_n^2 we can calculate $R_{f_r}(0, 0)$ by means of a extrapolation of the values of $R(x, y)$ close to the origin. In this manner, the power spectral density of the noise can be approximated by $S_n(\mu, \nu) = \hat{\sigma}_n^2$.

3 Proposed System for Object Recognition and Tracking in Nonuniform Illumination

Now we describe the proposed system for object recognition and tracking in nonuniform illumination. Let us consider the optical setup shown in Fig. 1 in where a target moves in horizontal direction of the two-dimensional plane, and it is under the influence of a illumination source. At time t_k , the camera captures

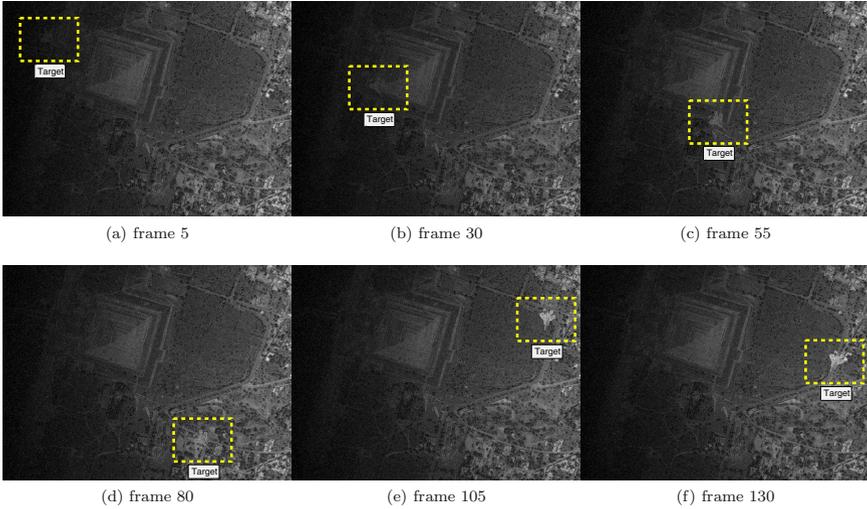


Fig. 2. Examples of nonuniformly illuminated scene frames corrupted with 20 dB SNR additive noise

$$\begin{aligned}
 \alpha_{k+1} &= \alpha_k + \frac{\sin(\omega_k \Delta_t)}{\omega_k} \dot{\alpha}_k - \frac{1 - \cos(\omega_k \Delta_t)}{\omega_k} \dot{\beta}_k + a_{\alpha,k} \frac{\Delta_t^2}{2}, \\
 \beta_{k+1} &= \beta_k + \frac{1 - \cos(\omega_k \Delta_t)}{\omega_k} \dot{\alpha}_k + \frac{\sin(\omega_k \Delta_t)}{\omega_k} \dot{\beta}_k + a_{\beta,k} \frac{\Delta_t^2}{2}, \\
 \dot{\alpha}_{k+1} &= \cos(\omega_k \Delta_t) \dot{\alpha}_k - \sin(\omega_k \Delta_t) \dot{\beta}_k + a_{\alpha,k} \Delta_t, \\
 \dot{\beta}_{k+1} &= \sin(\omega_k \Delta_t) \dot{\alpha}_k + \cos(\omega_k \Delta_t) \dot{\beta}_k + a_{\beta,k} \Delta_t, \\
 \omega_{k+1} &= \omega_k + a_{\omega,k}.
 \end{aligned} \tag{12}$$

The variables α_k and β_k denote the position of the target in Cartesian coordinates, $\dot{\alpha}_k$ and $\dot{\beta}_k$ are velocity components in α and β directions, and ω_k is the target's angular rate. Furthermore, $a_{\alpha,k}$ and $a_{\beta,k}$ are random variables representing acceleration components (due to turbulence) in α and β directions and $a_{\omega,k}$ is the angular acceleration. The position of the target in a subsequent time, is predicted by substitution of the estimated position $(\hat{\alpha}_k, \hat{\beta}_k)$, the estimated velocity components $(\hat{\dot{\alpha}}_k, \hat{\dot{\beta}}_k)$, and the estimated turn rate $\hat{\omega}_k$ (calculated from current and past frames) into the state space model in Eq. (12), and then by taking the expected value.

- Step 7: Read a new frame $f_k(x, y)$ from the scene and create a fragment of interest according to state prediction. Go to STEP 2.

4 Results

Here, results obtained with the proposed system for target tracking in nonuniformly illuminated scenes are presented. The results are given in terms of

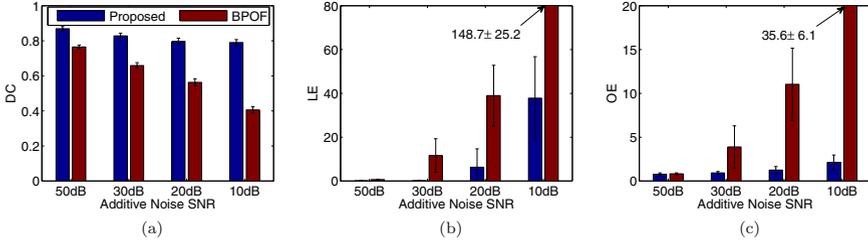


Fig. 3. Performance of tracking systems with 95% confidence in terms of: (a) DC, (b) LE, and (c) OE

recognition performance and tracking accuracy. The obtained results are compared with those obtained with a system based on binary phase-only (BPO) filters [3]. The recognition performance is measured in terms of the DC, whereas tracking accuracy is characterized by the precision of estimates carried out for the target state across scene frames. The accuracy in location estimation is characterized by the LE, which is given by [6]

$$LE = \left[(x^q - \hat{x}^q)^2 + (y^q - \hat{y}^q)^2 \right]^{1/2}, \quad (13)$$

were (x^q, y^q) and (\hat{x}^q, \hat{y}^q) are the exact and estimated coordinates of location of the target, respectively. The accuracy of estimation of the orientation angle is characterized by the orientation error (OE), defined by

$$OE = \left| \phi^q - \hat{\phi}^q \right|, \quad (14)$$

where ϕ^q and $\hat{\phi}^q$ are the true and estimated orientation angles, respectively. The units for LE and OE metrics are pixels and degrees, respectively. In our experiments we use a sequence of 200 nonuniformly illuminated scene frames with 800×600 pixels. Figure 2 shows examples of various scene frames in environment of 20 dB SNR additive noise. The target is an airplane which can move and rotate freely in the horizontal plane. With 95% confidence, the results for 200 scene frames obtained with proposed and BPO systems are presented in Fig. 3. We can see that the proposed system yields the best results in all the cases. Observe from Fig. 3(a) that the proposed system yields DC values close to unity even in highly noisy conditions of 10 dB SNR. Furthermore, we see that the

Table 1. Detection performance of tracking systems in 200 scene frames

Decision	Additive Noise SNR							
	50dB	30dB	20dB	10dB	50dB	30dB	20dB	10dB
detected	200	200	198	187	200	180	136	104
not detected	0	0	2	13	0	20	64	96
% of error	0%	0%	1%	6.5%	0%	10%	32%	48%
	Proposed system				BPO system			

proposed system can estimate with a good precision the location of the target in noisy conditions of 50 dB, 30 dB, and 20 dB SNR (see Fig. 3(b)). Additionally, from Fig. 3(c) we see that the proposed system estimates with a good accuracy the orientation angle of the target even in highly noisy conditions. The BPO system, yields good results in terms of the DC for 50 dB, 30 dB, and 20 dB SNR. Furthermore, this system yields high LE and OE values for 20 dB and 10 dB SNR. Table 1 shows the recognition performance of the tracking systems in 200 scene frames. The proposed system yields no detection errors for 50 dB and 30 dB SNR, and yields only two false detections for 20 dB SNR, and yields thirteen false detections for highly noisy conditions of 10 dB SNR. The BPO system yields good results for 50 dB SNR, however when the SNR decreases the number of false detections increases.

5 Conclusions

A tracking system for nonuniformly illuminated scenes was presented. The system employs a filter bank of time variant correlation filters to estimate the state trajectory of a moving target in a sequence of images. By incorporation of a prediction stage the system creates a fragment of interest in the observed frame and modify the number of required filters in the bank to estimate the state of the target in current frame. By means of computer simulations we showed that the proposed system yields a superior performance in terms of tracking accuracy comparing with recent state of the art tracking systems based on correlation filtering.

References

1. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. *ACM Comput. Surv.* 38(4) (2006)
2. Diaz-Ramirez, V.H., Kober, V.: Adaptive phase-input joint transform correlator. *Appl. Opt.* 46(26), 6543–6551 (2007)
3. Manzurv, T., Zeller, J., Serati, S.: Optical correlator based target detection, recognition, classification, and tracking. *Appl. Opt.* 51, 4976–4983 (2012)
4. Ouerhani, Y., Jridi, M., Alfalou, A., Brosseau, C.: Optimized pre-processing input plane GPU implementation of an optical face recognition technique using a segmented phase only composite filter. *Opt. Comm.* 289, 33–44 (2013)
5. Yaroslavsky, L.P.: The theory of optical method for localization of objects in pictures. In: Wolf, E. (ed.) *Progress in Optics*, vol. XXXII, pp. 145–201. Elsevier North-Holland (1993)
6. Kober, V., Campos, J.: Accuracy of location measurement of a noisy target in a nonoverlapping background. *J. Opt. Soc. Am. A* 13(8), 1653–1666 (1996)
7. Aguilar-Gonzalez, P.M., Kober, V.: Design of correlation filters for pattern recognition using a noisy reference. *Opt. Comm.* 285(5), 574–583 (2012)
8. Javidi, B., Wang, J.: Design of filters to detect a noisy target in nonoverlapping background noise. *J. Opt. Soc. Am. A* 11, 2604–2612 (1994)
9. Hu, W., Tan, T., Wang, L., Maybank, S.: A survey on visual surveillance of object motion and behaviors. *IEEE Trans. Syst. Man Cybern. C Appl. Rev.* 34(3), 334–352 (2004)