Optics Communications 323 (2014) 32–43

Contents lists available at ScienceDirect

Optics Communications

journal homepage: www.elsevier.com/locate/optcom

Target tracking in nonuniform illumination conditions using locally adaptive correlation filters

Victor H. Diaz-Ramirez [a,*], Kenia Picos [a], Vitaly Kober [b]

[a] Instituto Politécnico Nacional – CITEDI, Ave. del Parque 1310, Mesa de Otay, Tijuana B.C. 22510, Mexico
[b] Department of Computer Science, Division of Applied Physics, CICESE, Carretera Ensenada-Tijuana 3918, Zona Playitas, Ensenada B.C. 22860, Mexico

ARTICLE INFO

Article history:
Received 4 January 2014
Received in revised form
19 February 2014
Accepted 26 February 2014
Available online 11 March 2014

Keywords:
Target tracking
Locally adaptive correlation filters
Nonuniform illumination
Graphics processing unit (GPU)
Opto-digital correlators

ABSTRACT

An accurate method for tracking the position and orientation of a moving target in nonuniformly illuminated and noisy scenes is proposed. The approach employs a filter bank of space-variant correlation filters which adapt their parameters accordingly with local statistics of the observed scene in each frame. When a scene frame is captured, a fragment of interest is extracted from the frame around predicted coordinates of the target location. The fragment is firstly preprocessed to correct the illumination. Afterwards, the location and orientation of the target are estimated from the corrected fragment with the help of the filter bank. The performance of the proposed system in terms of tracking accuracy is tested in nonuniformly illuminated and noisy scene sequences. The obtained results are discussed and compared with those of similar state-of-the-art techniques for target tracking in terms of objective metrics.

1. Introduction

Nowadays, object recognition and tracking receive a great interest due to their high impact on real-life activities such as surveillance, robotics, vehicle navigation, biometrics, among others. Target tracking consists in estimation of the state trajectory of a target while its moves across the environment. The state of the target in a scene can be specified as a vector of attributes such as position, orientation, velocity, among others. The tracking problem can be solved by detecting the target in successive frames and by finding the correspondence between target states across scene frames. The state-of-the-art of target tracking considers various approaches [1,2]. The approaches differ in representation of the target in the scene and in chosen features for description of the target. A classic approach in target tracking is given by Kalman filtering [3]. This method is a recursive Bayesian estimator of the target position assuming a linear motion. The Kalman filter yields good results only when the state and observation variables are random and jointly Gaussian [1]. Moreover, the detection problem must be usually solved before the estimation problem (i.e., tracking of the objects) can begin. The Kalman filter requires an initial state for each object, and that initial state estimate must be obtained by first detecting the target. An extension of the Kalman filter for target tracking is given by the particle filter [4–6]. This approach utilizes Markov chain and Monte Carlo methods to approximate the solution of sequential target's state estimation. The particle filtering is useful for nonlinear motion. However, it is computationally intensive and it also faces convergence problems [7]. Another successful tracking method is given by the mean-shift algorithm [8–10]. This algorithm utilizes a basic shape (ellipse) to represent the target and uses the histogram as a main target descriptor. The mean-shift algorithm is computationally efficient and effective for tracking the position and orientation of a target by exploiting perspective transforms [1]. A main drawback of the mean-shift algorithm is its high sensitivity to illumination changes and occlusions. Recently, a tracking-learning-detection (TLD) algorithm for real-time target tracking was introduced [11]. This method learns past detection errors and self-adjusts to avoid the errors in future. The TLD algorithm is able to estimate the target position in every frame with tolerance to illumination changes, presence of noise and background clutter, and partial occlusion of a target. Other known tracking methods use key-point features to improve their tolerance to occlusions [12,13]. A detailed review of existing tracking algorithms can be found in Refs.[1,2]. One can note that majority of these methods still face difficulties in the presence of additive sensor's noise and cluttering background, exiting and re-entering of the target to the observed scene, illumination changes, and real-time requirements. In this work, we deal with these problems by proposing alternatives for target tracking.

* Corresponding author. Tel.: +52 664 623 1344x82856.
E-mail addresses: dira.vh@gmail.com, vdiazr@ipn.mx (V.H. Diaz-Ramirez).

http://dx.doi.org/10.1016/j.optcom.2014.02.063

An attractive alternative to conventional tracking methods is given by correlation filtering. Correlation filters have a good formal basis, and they can be implemented for real-time applications either in hybrid opto-digital correlators [14,15] or in high-performance hardware such as graphics processing units (GPUs) [16] exploiting massive parallelism. Recently, several successful proposals of effective implementation of correlation filters for pattern recognition in digital hardware were proposed [17]. A correlation filter is a linear system where the coordinates of the system output maximum are estimates of the target location in the observed scene [18]. An important characteristic of correlation filters is that they can be designed to recognize targets in high cluttering and noisy environments [19–21]. Furthermore, many attempts have been made to design correlation filters with tolerance to illumination changes [22–25]. Tara et al. [22] suggested the use of a nonlinear correlation filtering which can recognize a target with invariance to intensity changes and with tolerance to spatially disjoint noise. This filtering improves the peak-to-sidelobe ratio (PSR) [26]. However, it requires three correlation operations which results in a considerable increase of computational complexity. In order to recognize different views of a target a bank of correlation filters is commonly used [27,28]. There exits a trade-off among robustness to different views of the target and computational complexity of the overall system. In practice, correlation filters are able to provide a precise estimation of the target location in the observed scene [29], even in the presence of scene noise. We believe that the desired properties of correlation filters can be extended to the problem of target tracking by applying correlation filters to multiple frames. Various proposals for target tracking based on correlation filtering were suggested [15,30,31]. Manzurv et al. [15] introduced a new opto-digital correlator for real-time target tracking. This system employs a bank of binary-phase-only (BPO) filters [32]. The BPO filters are placed consequently and rapidly in the filter (Fourier) plane of the correlator to produce a sequence of output correlation planes. The obtained planes are subsequently analyzed and processed to track the target across scene frames. An important limitation of this architecture is that BPO filters are time-invariant. Thus, they are unable to correctly deal with eventual nonstationary noisy environment. Another disadvantage of this proposal is that all input frames are considered as independent signals. Thus, important information which can be obtained from past state estimates is discarded.

In this work, we propose a correlation-based approach for accurate recognition and tracking of a moving target in nonuniformly illuminated and noisy scenes. The proposed approach employs a filter bank of locally adaptive correlation filters. The frequency response of the filters is varied in each frame accordingly with local statistics of the observed signal. When a scene frame enters the system, a pointwise locally adaptive illumination correction is firstly applied to the frame. Afterwards, the target is detected in the restored frame by analyzing the correlation intensity planes obtained at the output of the filter bank. Next, the system predicts the state of the target (position and orientation) for next frame, and based on this prediction creates a fragment of interest in the frame and modifies the number of filters in the bank. The state prediction is computed by analyzing current and past estimates of the target and by taking into account a two-dimensional kinematic model [33]. The proposed approach is able to track with high accuracy a moving target in nonuniformly illuminated and noisy scenes. Additionally, in order to achieve a real-time performance (up to 60 frames-per-second) the bank of correlation filters is implemented in parallel on a GPU. The paper is organized as follows. Section 2 presents the proposed approach for target detection in nonuniformly illuminated and noisy scenes. Section 3 discusses the proposed system for accurate target tracking. Section 4 shows the results obtained with the proposed system by testing its performance in nonuniformly illuminated and noisy scenes. Finally, Section 5 summarizes our conclusions.

2. Target detection in nonuniformly illuminated and noisy scenes

In this section we describe the proposed approach for target detection in nonuniformly illuminated and noisy scenes. First, we briefly review illumination models. Next, we define the scene model and describe the suggested preprocessing for illumination correction. Finally, we explain the suggested method for target detection in nonuniform illumination conditions.

2.1. Illumination models

An illumination model describes relationships between the illuminant direction and a surface shape. These models can be broadly categorized in terms of the reflectance properties of illuminated surfaces. The Lambertian model is used to represent opaque surfaces, whereas specular models are used to describe mirror-like surfaces. Consider the setup shown in Fig. 1, where a target moves in horizontal direction of two-dimensional plane. The surface in Fig. 1 is being illuminated by a point-light-source with location parameters $\mathbf{S}=[\rho,\phi,\psi]^T$. Here, ρ is the distance between a point in the surface and the light source, and ϕ and ψ are the tilt and slang angles, respectively, between the surface normal and the observation point. In the Lambertian model, the surface reflects incident light in all directions with equal amplitudes [34]. Let Ω be the incident angle of light, i.e., the angle between the surface normal vector $\mathbf{N}$ and the illuminant direction vector $\mathbf{S}$. The light reflected by a Lambertian surface is given by $I_L=\cos(\Omega)$. In the specular model, incident light is reflected by the surface as [20] $I_s=\delta(\Omega_s-\Omega)$, where $\delta(\Omega)$ is the Dirac delta function, $\Omega_s=(\Omega_N-\Omega_v)$, Ω_N is the angle between $\mathbf{N}$ and the z-axis, and Ω_v is the angle between $\mathbf{N}$ and the observation point. Hybrid reflectance models are also used. They are given as a combination of Lambertian and specular components as follows: $I_H=k_1I_L+k_2I_S$, where $k_1+k_2=1$. In this work we focus on the Lambertian model. According to the setup given in Fig. 1, the light reflected by the surface for a known position of the light source and assuming the z-axis as the observation point is given by[23]

$$d(x,y)=\cos\left\{\frac{\pi}{2}-\arctan\left[\frac{\rho}{\cos(\phi)}[(\rho\tan(\phi)\cos(\psi)-x)^2+(\rho\tan(\phi)\sin(\psi)-y)^2]^{-1/2}\right]\right\}. \tag{1}$$

Note that $d(x,y)$ in Eq. (1) is a multiplicative function which depends on the parameters ρ, ϕ, and ψ. In practice the location parameters of light source may be unknown. Thus, a processing algorithm to compensate inhomogeneous effects introduced by the illumination function $d(x,y)$ is required.

2.2. Scene model and illumination correction preprocessing

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 by a zero-mean stationary white Gaussian noise $n(x,y)$. In this case, the observed scene can be expressed by the nonoverlapping signal model [19] as follows:

$$f(x,y)=[t(x-\alpha,y-\beta)+\overline{w}(x-\alpha,y-\beta)b(x,y)]\,d(x,y)+n(x,y), \tag{2}$$

where $\overline{w}(x,y)$ is the inverse region of support of the target given by unity outside the target area and zero elsewhere. According to the