ARTICLE IN PRESS

Optics Communications ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at SciVerse ScienceDirect

Optics Communications

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

Real-time tracking of multiple objects using adaptive correlation filters with complex constraints

Q1 Victor H. Diaz-Ramirez [a,*], Viridiana Contreras [a], Vitaly Kober [b], Kenia Picos [a]

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

A R T I C L E I N F O

Article history:
Received 8 May 2013
Received in revised form
9 July 2013
Accepted 18 July 2013

Keywords:
Object recognition
Object tracking
Composite correlation filters
Adaptive filters
Graphics processing unit
Opto-digital correlators

A B S T R A C T

A real-time system for classification and tracking of multiple moving objects is proposed. The system employs a bank of composite correlation filters with complex constraints implemented in parallel on a graphics processing unit. When a scene frame is captured, the system splits the frame into several fragments on the base of a modeling kinematic prediction of target's locations. The fragments are processed with a bank of adaptive filters. The filters are synthesized with the help of an iterative algorithm, which optimizes discrimination capability for each target. Using complex constraints in the filter design, multiple objects in the input frame can be detected and classified by analyzing the intensity and phase distributions on the output complex correlation plane for each fragment. The performance of the proposed system in terms of tracking accuracy, classification efficiency and time expenses is tested and discussed with synthetic and real input-scene sequences. The results are compared with those of common techniques based on correlation filtering.

1. Introduction

Research in pattern recognition has recently received a notable attention due to a growing interest in developing imaging systems for assisting people in real life. Automatic object identification, video surveillance, vehicle navigation and object tracking are examples of pattern recognition applications [1–3]. Pattern recognition consists of effective detection, localization and classification of objects of interest in an input observed scene. These problems can be addressed by different approaches, for instance, by feature-based methods or by correlation filtering. In feature-based methods [3,4], the observed scene is firstly preprocessed to extract relevant features of potential targets. Next, the features are analyzed in order to make a decision. A main drawback of feature-based methods is that its overall performance depends on some ad-hoc decisions, which often require optimization [5].

On the other hand, correlation filtering is an attractive alternative to feature-based methods, and it is a suitable option for real-time applications [6,7]. Correlation filters possess a good mathematical basis, and they can be implemented by exploiting massive parallelism either in hybrid opto-digital correlators [6,8,9] or at high speed in digital hardware such as Graphic Processing Units (GPUs) [7,10] or Field Programmable Gate Arrays (FPGAs) [11]. A correlation filter is a linear system. The coordinates of the system output maximum are estimates of the target coordinates in the observed scene [12,13]. An important feature of correlation filters is that they are able to recognize objects in highly cluttered and noisy environments; for instance, when input images are corrupted by additive and disjoint noise or when targets suffer from geometrical distortions such as in-plane rotations or scaling [14–19]. The filters can be designed by optimizing various performance criteria [20,21] and taking account models of signals and noise. For distortion invariant pattern recognition, composite filters are commonly used [22–24]. These filters are synthesized by combining several training templates, which represent a set of possible target views. Thus, a single composite filter can be used to recognize different target versions.

When a target moves across the observed scene, the target's appearance with respect to the observer varies with time. Actually, target tracking consists in the estimation of target trajectory in the observed scene while the object moves across the environment [25,26]. Hence, the problem of target tracking can be treated with composite correlation filters applied to multiple frames. The tracking problem can be solved by detecting the object in the frames and by finding the correspondence between object states across the frames. Commonly object detection is carried out by employing feature-based methods [4,27], while tracking is performed by matching the states of detected objects in consecutive frames by taking into account a state-space model [28]. There are

* Corresponding author. Tel.: +52 664 623 1344x82856.
E-mail addresses: dira.vh@gmail.com, vdiazr@ipn.mx (V.H. Diaz-Ramirez), vcontreras@citedi.mx (V. Contreras), vkober@cicese.mx (V. Kober), kpicos@citedi.mx (K. Picos).

0030-4018/$ - see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.optcom.2013.07.044

several fruitful proposals which perform object tracking with the help of feature-based methods and state-space models [25,26,29,30]. However, these techniques face some unsolved issues when there are abrupt changes in target motion; target exits and reenters to the observed scene; the scene is distorted by additive and disjoint noise; the target is geometrically deteriorated by rotation and scaling. In order to overcome these difficulties, it is necessary to explore new approaches for target tracking. The use of correlation filtering for target tracking has been investigated as well.

Bhuiyan et al. [31,32] proposed a two-step approach for object classification and tracking. The first step in the approach is to use a bank of maximum average correlation height (MACH) filters [33] to detect potential targets within the scene. This is done by looking for correlation peaks in the outputs of the filter bank and then by taking the peak coordinates as location estimates for potential targets. Next, around the location estimates fragments of interest are formed from the current frame. It is assumed that each fragment contains one potential target of unknown class. In a second step, the fragments are processed with a bank of polynomial distance classifier (PDCC) filters [34] to establish the class for each detected object. This process is repeated for all frames. A main drawback of this proposal is that each frame is considered independent with respect to others. So information from past estimates is not used. Another drawback is a high computational complexity because the system consists of two filter banks.

Recently, Manzur et al. [6] introduced a new opto-digital correlator for target detection, classification and tracking. This proposal takes advantage of inherent parallelism of optics for processing large volumes of data in real-time. The correlator employs a bank of binary phase-only (BPO) filters [35,36]. Each filter is able to recognize only one version of a target with low tolerance to geometrical distortions. The filters enter consequently and rapidly in the filter plane, and the system yields a sequence of output correlation responses. The obtained correlation planes are analyzed and processed to classify and to track the object across the frames.

In this work, we propose a real-time system for multiclass object recognition and tracking. The proposed system employs a bank of adaptive composite correlation filters with complex constraints. The filters are implemented in parallel on a GPU. First detection and classification of multiple targets are carried out by analyzing the correlation peaks at the system output for the current frame. Next, the system predicts states of the targets for the subsequent frame, and based on the prediction creates fragments of interest in the input frame and modifies the number of filters in the bank. When a new frame enters in the system, several small fragments around predicted locations for the targets are formed. The number of filters in the bank is also adjusted accordingly to predicted orientation of the targets 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 state-space motion model. The resultant system is able to perform classification and tracking of multiple moving objects in real-time by exploiting the benefits of massive parallelism of the GPU and by reducing false alarm probabilities by focusing the processing only on small fragments.

The paper is organized as follows. Section 2 presents a brief review of constrained composite filters for multiclass object recognition problems. In Section 3, we explain the proposed real-time system for multiclass object detection, classification and tracking. Section 4 presents the results obtained with the proposed system by testing its performance in synthetic and real-life input scene sequences. The obtained results in terms recognition performance, classification efficiency, tracking accuracy and speed of processing are presented. The results are compared with those of common techniques based on correlation filtering. Finally, Section 5 summarizes our conclusions.

2. Constrained composite filters for multiclass pattern recognition

Composite correlation filters use a set of training templates to represent multiple versions of a target in the observed scene. These filters are designed to produce a specific response in the presence of one of the known target versions used for training. Synthetic discriminant functions (SDFs) [22,37] yield a filter that is a linear combination of the training templates. Therefore, a SDF filter can be used for one-, two- and multiclass pattern recognition problems.

2.1. One-class problem

Let $T=\{g_i(x,y)|i=1,\ldots,N_T\}$ be a set of N_T different templates, each of them represents a different version of a given target $s(x,y)$. A SDF filter is able to recognize different versions of the target with only one correlation operation. Let $h(x,y)$ be the impulse response of a SDF filter given by

$$h(x,y)=\sum_{i=1}^{N_T} a_i g_i(x,y), \tag{1}$$

where $\{a_i|i=1,\ldots N_T\}$ are unknown weighting coefficients. The filter $h(x,y)$ is a linear combination of N_T training templates $g_i(x,y)$. The weighting coefficients are calculated subject to the inner-product constraints

$$\langle h^*(x,y), g_i(x,y)\rangle = c_i, \tag{2}$$

where "*" means complex conjugate and c_i is prespecified quantity imposed to the output cross-correlation value in the origin, between filter $h(x,y)$ and template $g_i(x,y)$. Let $\boldsymbol{\Upsilon}$ be a $N_T\times d$ matrix, where d is the number of elements in each template. The ith column in $\boldsymbol{\Upsilon}$ is given by $\mathbf{g}_i$, a $d\times 1$ vector obtained by reordering elements of $g_i(x,y)$ in lexicographical order. Let $\mathbf{a}$ and $\mathbf{c}$ represent column vectors of $\{a_i\}$ and $\{c_i\}$. Using matrix-vector notation, the filter $h(x,y)$ and constraints $\{c_i\}$ can be rewritten, respectively, as

$$\mathbf{h}=\boldsymbol{\Upsilon}\mathbf{a}, \tag{3}$$

and

$$\mathbf{c}^*=\boldsymbol{\Upsilon}^+\mathbf{h}, \tag{4}$$

where superscript "+" means conjugate transpose. By combining Eqs. (3) and (4), the solution vector (if matrix $(\boldsymbol{\Upsilon}^+\boldsymbol{\Upsilon})$ is nonsingular) is

$$\mathbf{h}=\boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}^+\boldsymbol{\Upsilon})^{-1}\mathbf{c}^*. \tag{5}$$

The SDF filter in Eq. (5) can be used for one-class pattern recognition problems, that is, to recognize different versions of a given target $s(x,y)$. This can be done by setting all elements of vector $\mathbf{c}$ equal to unity, i.e., $\mathbf{c}=[1,1,\ldots,1]^T$.

2.2. Two-class problem

Assume that we have a true-class set T and a false-class set $F=\{p_i(x,y)|i=1,\ldots,N_p\}$ consisting of N_p templates $p_i(x,y)$ (each one with d elements). The latter represents unwanted patterns to be rejected. Let $\mathbf{p}_i$ be a $d\times 1$ vector containing the elements of $p_i(x,y)$ in lexicographical order.

We can synthesize a two-class SDF filter which is able to recognize all target versions in T and to reject unwanted patterns in F, by solving Eq. (5). Here matrix $\boldsymbol{\Upsilon}$ is a $d\times(N_T+N_P)$ matrix,