Signal Processing 117 (2015) 343–354

Contents lists available at ScienceDirect

Signal Processing

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

Autonomous detection and tracking under illumination changes, occlusions and moving camera

Harish Bhaskar [a,b,*], Kartik Dwivedi [c], Debi Prosad Dogra [d], Mohammed Al-Mualla [a], Lyudmila Mihaylova [e]

[a] Visual Signal Analysis and Processing (VSAP) Research Center, Khalifa University of Science, Technology and Research (KUSTAR), P.O. Box 127788, Abu Dhabi, United Arab Emirates
[b] University of Bristol, Merchant Venturers' Building, Woodland Road, Clifton, Bristol BS8 1UB, United Kingdom
[c] Department of Electronics and Electrical Engineering, Indian Institute of Technology (IIT), Guwahati 781039, India
[d] School of Electrical Sciences, Indian Institute of Technology (IIT), Bhubaneswar 751013, India
[e] Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, United Kingdom

ARTICLE INFO

Article history:
Received 11 December 2014
Received in revised form
30 May 2015
Accepted 3 June 2015
Available online 19 June 2015

Keywords:
Target detection
Tracking
Illumination variations
Occlusion
Camera movements
Reverse analysis
Rao-Blackwellized particle filter
Likelihood model

ABSTRACT

In this paper, an autonomous multiple target detection and tracking technique for dynamic scenes that are influenced by illumination variations, occlusions and camera instability is proposed. The framework combines a novel Dynamic Reverse Analysis (DRA) approach with an Enhanced Rao-Blackwellized Particle Filter (E-RBPF) for multiple target detection and tracking, respectively. The DRA method, in addition to providing accurate target localization, presents the E-RBPF scheme with costs associated with the differences in intensity caused by illumination variations between consecutive frame pairs in any video of a dynamic scene. The E-RBPF inherently models these costs, thus allowing the framework to (a) adapt learning parameters, (b) distinguish between camera-motion and object-motion, (c) deal with sample degeneracy, (d) provide appropriate appearance compensation during likelihood measurement and (e) handle occlusion. The proposed detect-and-track method when compared against other competing baseline techniques has demonstrated superior performance both in accuracy and robustness on challenging videos from publicly available datasets.

1. Introduction and related work

Detection and tracking are challenging research problems, particularly in unconstrained surveillance scenarios. Much research efforts have been spent in developing state-of-the-art detection and tracking methodologies including detect-and-track [1], track-before-detect [2,3], Probability Hypothesis Density (PHD) filter based multiple target tracking techniques [4,5], among many others. Despite advances, detection and tracking are still challenged by the presence of illumination variations [4], occlusions [6,7], and camera movements [8]. Although many approaches have aimed to address these issues in a mutually exclusive manner, the joint problem is still far from being solved.

The concept of illumination invariance during target detection and tracking has been addressed in different ways which can be categorized into feature-based [9] and appearance-based methods [10]. A good example of the feature-based detection technique can be found in [9], where a sparse set of salient illumination invariant features are considered. Similarly, in the work of [11,12], the bi-parametrization of different combinations of colour spaces

* Corresponding author at: Visual Signal Analysis and Processing (VSAP) Research Center, Khalifa University of Science, Technology and Research (KUSTAR), P.O. Box 127788, Abu Dhabi, United Arab Emirates.
E-mail addresses: harish.bhaskar@kustar.ac.ae (H. Bhaskar), k.dwivedi@iitg.ac.in (K. Dwivedi), dpdogra@iitbbs.ac.in (D.P. Dogra), almualla@kustar.ac.ae (M. Al-Mualla), l.s.mihaylova@sheffield.ac.uk (L. Mihaylova).

http://dx.doi.org/10.1016/j.sigpro.2015.06.003
0165-1684/

have been studied for dynamic target tracking. However, such methods have failed to adequately discriminate targets against the background, during detection. A comprehensive overview of the recent efforts in background modelling based detection approaches can be found in [13,14]. From a tracking point-of-view, some initial work has been done in the joint target localization and estimation of illumination variations as in [15,16]. Similarly, a method for coping with appearance changes of targets during tracking has been proposed in [17,18]. Although such methods are proven to handle constrained gradual illumination changes, modelling of illumination changes in real-time continues to be highly complex. This can mainly be attributed to factors such as: (a) non-linearity of illumination changes in real-world scenarios, (b) the ambiguity in the interpretation of the differences in intensity variations caused by the motion of targets as against due to illumination changes, and (c) disregarding certain pertinent visual information to provide illumination invariance that can cause difficulties in handling occlusion and related challenges.

Occlusion detection during tracking is considered a hard problem in most general-purpose tracking algorithms [19]. The primary challenge in handling occlusion is to accumulate sufficient evidence from observations so that reliable data association becomes possible. Research indicates that performing occlusion handling within tracking is limited only to analysing pixel variations using multi-modal distributions in order to encompass statistical properties of occluders to distinguish it from the target(s)-of-interest [20]. However, most assumptions of feature-level similarity become invalid when considering real-world scenarios. In the study by [20], it has been shown that the contextual content which encapsulates motion information is also capable of handling occlusion. In another example, the problem of target initiation and termination has been shown to be handled using a hierarchical particle filtering framework [21]. Further, the use of spatio-temporal modelling has been proposed for human silhouette extraction from noisy and occluded data [7]. Compressive sensing [22] and sparse representation based methods have also been used for occlusion handling and illumination invariant tracking [23]. Methods such as [24,25], cast target tracking as a sparse approximation problem within typical particle filtering framework usually combining multi-features [25]. Though attempts have been made to tackle occlusion issues during tracking, the following complexities continue to remain: (a) localization of the targets when occlusion is unknown, (b) updating target descriptors during appearance changes, (c) robustness against noise and clutter, and (d) coping with disappearances and re-appearances of targets.

Motion in the background, target deformation and changes in the camera position during jitter all present similar effect on the spatio-appearance of targets during detection and tracking. In order to model spatio-temporal appearance changes of targets in the joint space, the use higher order distributions has become a popular choice [26]. Recent studies have focused on using Alpha-stable [26] and Cauchy [27] distributions to model pixel intensity variations caused by camera shake during detection. Further, the generation of spatio-temporal methods for handling camera movements within a background modelling framework has also been recently proposed in [28]. However, the robustness of such models for dynamic scenes has not been fully explored.

This brief survey of the literature has clearly highlighted that the treatment of these challenges in a mutually exclusive manner cannot facilitate robust detection and tracking in real-world scenarios. On the other hand, incorporating different adaptations into a singular model may not always help solving all problems jointly. Therefore, in this paper, a tight integration of the constituent processes into a unified framework for the detection and tracking of dynamic scenes is proposed. It is hypothesized that an integrated detect-and-track technique capable of generating sufficient statistics of illumination variations during accurate detection, when tunnelled across to an enhanced tracking framework shall provide robust localization of multiple targets within a dynamic scene.

2. Novelty and contributions

One key novelty of the proposed framework is the use of DRA within background modelling for accurately detecting (or spatially localizing) multiple moving targets under changing illumination conditions. Furthermore, such a detection procedure allows extracting sufficient statistics that are indicative of the temporal location, type and extent of the illumination variations in the dynamic scene. Another important novelty is the tight integration of the qualitative and quantitative outputs from the DRA-based target detection method together with the E-RBPF tracking algorithm for adaptation against changing illumination and camera movements. Finally, the loose integration of the E-RBPF framework with an appropriate noise model and relevant likelihood measurements has allowed the framework to compensate for local (dis)order.

The rest of the paper is organized as follows. In Section 3, a detailed description of the proposed detect-and-track framework is presented. Following this, performance evaluation and comparison of the proposed detection and tracking methodologies against the state-of-the-art is described in Section 4. Section 5 concludes.

3. Proposed methodology

The proposed framework for multiple target tracking is a modularized yet coupled approach. The method tightly integrates (a) a hybrid background modelling scheme for accurate target detection with (b) an enhanced particle filtering framework for robust target tracking. An illustration of the proposed framework is presented in Fig. 1.

The proposed hybrid background modelling method combines conventional background initialization and maintenance processes with a reverse analysis scheme for accurate target detection. The DRA technique exploits deviations in the foreground detection procedure between forward and reverse directions to extract sufficient statistics on the changes in illumination conditions to determine (a) the composition of optimal frames that produces a representative background model and (b) control of adaptation parameters for coping with dynamic changes.

In contrast to the rigorous detection process, tracking is developed as an enhancement to the Rao-Blackwellized