Optik 125 (2014) 2243–2247

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Multiple small objects tracking based on dynamic Bayesian networks with spatial prior

Rui Yao[a,b,*], Yanning Zhang[b], Yong Zhou[a], Shixiong Xia[a,*]

[a] School of Computer Science and Technology, China University of Mining and Technology, China
[b] School of Computer Science, Northwestern Polytechnical University, China

ARTICLE INFO

Article history:
Received 15 May 2013
Accepted 13 October 2013

Keywords:
Multiple small objects tracking
Dynamic Bayesian Networks
Multi-cue integration
Data association

ABSTRACT

This paper proposes an end-to-end algorithm for multiple small objects tracking in noisy video using a combination of Gaussian mixture based background segmentation along with a Dynamic Bayesian Networks (DBNs) based tracking. Background segmentation is based on an adaptive backgrounding method that models each pixel as a mixture of Gaussians with spatial prior and uses an online approximation to update the model, the spatial prior is constructed for small objects. Furthermore, we create observation model with hidden variable based on multi-cue statistical object model and employ Kalman filter as inference algorithm. Finally, we use linear assignment problem (LAP) algorithm to perform the models matching. The experimental results show the proposed method outperforms competing method, and demonstrate the effectiveness of the proposed method.

1. Introduction

Visual multiple objects tracking has received tremendous attention in the video processing community due to its numerous potential applications in important tasks such as video surveillance, human activity analysis, traffic monitoring, and so forth [8,15]. In particularly, small visible space objects tracking is one of the key issues in the research of long-range early warning and space debris surveillance [14]. The basic tracking task consists in estimating the trajectory of a moving object by consistently assigning a label over frames considering noisy measurements. Multiple objects tracking for objects whose appearance is distinctive is much easier since it can be solved reasonably well by using multiple independent single-object trackers. Recently, most researchers are mainly focus on the following two problems. On one hand, it is crucial to improve tracking performances in complex environments. Typical problems are partial and complete occlusion [6], multiple objects [4,5], multiple cameras [10], nonuniform object motion, nonrigid and articulated objects [3], complex and time-variable background [2]. For example, in occlusion situation, when tracking a specific object, all the other objects can be viewed as background due to their distinct appearance. The tracker must separate the objects and assign them correct labels [8]. On the other hand, accuracy tracking result is required by many applications, such as providing location of small objects.

The traditional tracking algorithm includes two steps: object detection and tracking. Object detection method mainly regards as background segmentation, which extracting object area from image background. Background modeling is a frequently used segmentation algorithm [11]. Literature [7,13] used color and histogram to segment object and background [13] presented a feature integration method to enhance the performance of tracking. Segmentation and tracking are two complementary tasks and several previous methods aim at combining MRF segmentation with object tracking/pose estimation [6]. However, most algorithms simplified the detection step, and focused on the tracking method. Recently, many researches focus on applying probability graphical models (PGMs) to tracking [1,9,12]. The application of Bayesian filtering algorithms to PGMs can provide useful tools for multiple-object tracking.

PGMs are able to provide an appropriate theoretical framework where object dynamics and appearance can be combined and the motion estimation problem can be efficiently solved. It can be divided into two general classes, i.e., directed acyclic graphs (DAGs) and undirected graphs (UGs). DAGs such as Hidden Markov Models (HMMs), Bayesian networks (BNs), DBNs, and Kalman filter models (KFMs) have been extensively used for video analysis because of their capability of modeling temporal relationships. These relationships are characterized by directional dependencies in the time evolution. Instead, Markov networks, Boltzmann machines, and loglinear models are widely employed for image analysis and segmentation to describe spatial dependencies between image pixels. Therefore, we usually use DAGs to solve tracking problem.

* Corresponding authors.
E-mail addresses: ruiyao@cumt.edu.cn, yaorui@mail.nwpu.edu.cn (R. Yao).

0030-4026/$ – see front matter © 2013 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ijleo.2013.10.108

In summary, our main contributions are as follows: (1) we propose an end-to-end solution for multiple small objects tracking based on Gaussian mixture and Dynamic Bayesian Networks; (2) we introduce spatial prior to traditional GMM method for small object tracking. The new GMM method can reduce false positives effectively; and (3) we construct a DBN with hidden variable to integrate multi-cue in tracking, which can improve tracking performance (as shown in Fig. 3). The rest of the paper is arranged as follows: Section 2 shows the details of the proposed background segmentation algorithm with spatial prior. The DBNs based tracking algorithm is introduced in Section 3. The experiments are showed in Section 4. Finally, we give the conclusion in Section 5.

2. Background segmentation

An efficient tracking algorithm should deal with movement through cluttered areas, occlusion, shadows, lighting changes, effects of moving elements of the scene, and objects being introduced or removed from the scene. Thus, we need robust segmentation algorithm that can account for such a wide range of effects. In this study, we model the values of a particular pixel as a mixture of Gaussian [11] with spatial prior. Based on the GMM and spatial prior, we determine which Gaussian may correspond to background pixels.

2.1. Mixture background model

We define the values of a particular pixel over time as a pixel process. At any time t, the history of a particular pixel (r_i, c_j) is $\{\mathbf{s}_1, \ldots, \mathbf{s}_t\} = \{I(r_i, c_j, k) : 1 \le k \le t\}$, where I is the image sequence. We model the history of a pixel $\{\mathbf{s}_1, \ldots, \mathbf{s}_t\}$ as a mixture of K Gaussian distributions. Then the probability of observing the current pixel value is

$$P(\mathbf{s}_t) = \sum_{k=1}^{K} \lambda_{k,t} f(\mathbf{s}_t, \mu_{k,t}, \Sigma_{k,t}), \tag{1}$$

where K is the number of distribution, $\lambda_{k,t}$ is the weight of the kth Gaussian distribution, $\mu_{k,t}$ and $\Sigma_{k,t}$ are the mean value and covariance matrix of the kth Gaussian distribution, and where f is a Gaussian probability distribution density function. For computational reasons, we assume that the color space is RGB, then the covariance matrix is assumed to be of the form $\Sigma_{k,t} = \sigma_k^2 I$. The next step is to update the parameter $\lambda_{k,t}$, $\mu_{k,t}$ and $\Sigma_{k,t}$ along the time serial. The update operation is described as following: when a new pixel value $\mathbf{s}_t$ coming, it is firstly checked against the existing K Gaussian distributions, until a match is found, that is a pixel value within 2.5 standard deviations of a distribution. Then, we adjust the prior weights of the K distributions at time t as $\lambda_{k,t} = (1-\alpha)\lambda_{k,t-1} + \alpha M_{k,t}$, where α is the learning rate and $M_{k,t}$ is 1 for the matched model and 0 for the others. For the matched model and unmatched model:

- If none of the K distributions match the current pixel value, the least probable distribution is replaced with a distribution with the current value as its mean value, an initially high variance, and low prior weight.
- If the current pixel value matched a distribution, the parameters of the distribution which matches are updated as follows,

$$\mu_t = (1-\rho)\mu_{t-1} + \rho \mathbf{s}_t, \tag{2}$$

$$\sigma_t^2 = (1-\rho)\sigma_{t-1}^2 + \rho(\mathbf{s}_t - \mu_t)^T(\mathbf{s}_t - \mu_t), \tag{3}$$

where $\rho = \alpha f(\mathbf{s}_t | \mu_k, \sigma_k)$ is the learning factor for adapting current distributions.

2.2. Foreground extraction with spatial prior

In order to extract the candidate object, we need to discrete the foreground and the background. As the each pixel process is composed by a mixture model, we would like to determine which of the Gaussian of the mixture are most likely produced by background processes. In general, we are interest in the distribution which has high support weight and low variance.

According to the analysis as mentioned above, we order the Gaussian distribution by the value of λ/σ. Then the first B distributions are chosen as the background model: $B = \operatorname{argmin}_b[\sum_{k=1}^{b} \lambda_k > T]$, where T is the threshold of the minimum portion of the data that should be accounted for by the background. Thus, the rest $K-B$ distributions are foreground.

Due to far from imaging device, the small objects occupy only several pixels, without any information about structure or texture. However, the shape of most of small objects is point. Its size generally has to be equal to or larger than 4 pixels, and less than 20–30 pixels. These are useful spatial prior. In the standard GMM method each pixel is treated separately and spatial prior of small objects is not considered. Using spatial prior false positives can be eliminated.

We use morphological operation to generate a spatial mask for eliminating part of false positives. For each pixel $\mathbf{a} = (r, c)$ at tth frame, its spatial mask value is estimated according to

$$S(\mathbf{a}, t) = \begin{cases} 0 & k(\mathbf{a}) \in [1, \ldots, B] \\ I(\mathbf{a}, t) \circ \mathbf{b} & \text{else} \end{cases}, \tag{4}$$

where $\circ$ is opening operator, $I \circ \mathbf{b} = (I \ominus \mathbf{b}) \oplus \mathbf{b}$. In practical applications, opening operations can remove small details (false positives, in our case). Here $\mathbf{b}$ is a 3×3 structuring element. And the distribution of $\mathbf{a}$ is got by $k(\mathbf{a})$.

To use the spatial dependency, where the neighborhood of each pixel is considered, the sum of the pixels in a square window W is computed. By using two thresholds $T_{\min}$ and $T_{\max}$ the number of false positives is reduced, and we can get the binary segmented image BI of tth frame,

$$BI(\mathbf{a}, t) = \begin{cases} 1 & \sum_W T_{\min} < S(\mathbf{a}, t) < T_{\max} \\ 0 & \text{else} \end{cases}. \tag{5}$$

So far, we have separated foreground objects from background, the next step is to construct temporal processes, and estimate the state of small objects.

3. DBNs-based tracking

Dynamic Bayesian Networks are used to model temporal stochastic processes. DBNs generalize HMMs by representing the hidden and observed state in terms of state variables, and modeling the probabilistic dependencies between them. In this sense, they can be considered as an extension of Bayesian Networks to handle temporal models [9]. In order to enhance the effectiveness of tracking algorithm, we employ a hidden variable ω to integrate multi-cue in tracking. Fig. 1 shows the Dynamic Bayesian Networks of the proposed algorithm, which $x_t \in \mathbb{R}^n$ represents the hidden node and $z_t \in \mathbb{Z}^m$ represents the observation node.

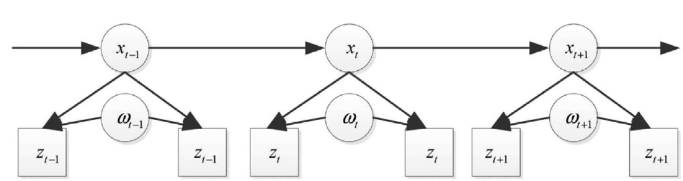

Fig. 1. Dynamic Bayesian Networks for the proposed algorithm.