Boosted Markov Chain Monte Carlo Data Association for Multiple Target Detection and Tracking

Qian Yu, Isaac Cohen, Gerard Medioni and Bo Wu
IRIS, Computer Vision Group, University of Southern California
Los Angeles, CA 90089 - 0273
{qianyu, icohen, medioni, bowu} @usc.edu

Abstract

In this paper, we present a probabilistic framework for automatic detection and tracking of objects. We address the data association problem by formulating the visual tracking as finding the best partition of a measurement graph containing all detected moving regions. In order to incorporate model information in tracking procedure, the posterior distribution is augmented with Adaboost image likelihood. We adopt a MRF-based interaction to model the inter-track exclusion. To avoid the exponential complexity, we apply Markov Chain Monte Carlo (MCMC) method to sample the solution space efficiently. We take data-oriented sampling driven by an informed proposal scheme controlled by a joint probability model combining motion, appearance and interaction among detected regions. Proposed data association method is robust and efficient, capable of handling extreme conditions with very noisy detection.

1. Introduction

Multiple targets tracking is a critical component of video surveillance systems. Environments of interest usually contain an unknown and varying number of moving targets. Automatic detection and tracking of multiple targets involves the detection of moving regions, the initialization of tracks, the association of regions across time and the filtering of erroneous detections or tracks. Instead of separating the detection and tracking as two separate procedures, we propose a probabilistic framework for automatic detection and tracking of objects, which combines the detection and tracking together. This allows object detection to make use of temporal consistency and facilitates robust tracking of the object. In our method, each preliminary detection derived by the motion segmentation is assigned a model likelihood provided by a real-valued Adaboost classifier trained offline.

In our framework, we formulate the multiple targets tracking as an association problem, in which the purpose is to find the best association between observations (*i.e.* detected moving regions) and targets while maximizing the posterior association probability. We represent the association problem in a deferred logic way like MHT [5] where association is defined between targets and a set of latest observations within a sliding window. This allows the association decision is made when enough observations are acquired. As the size of sliding window is extending, the scale of the problem grows exponentially. To avoid the enumeration of association hypothesis and solve this combinatorial optimization problem efficiently, we propose an Markov Chain Monte Carlo (MCMC) [1] method to sample the solution space. This MCMC sampling is driven by an informed proposal scheme controlled by a joint probability model combining motion, appearance and interaction among detected regions. We test our detection and tracking framework on videos captured on the moving platform, Unmanned Aerial Vehicles (UAV).

Recently in [6], the authors introduced a MCMC based sampling method to address the association of punctual observations. The posterior distribution assume a *prior* knowledge on the detection and the targets' behavior and consider only dynamics likelihood. In [4], a MCMC-based particle filter simulates the distribution of the association probability, which allows multiple temporal associations between observations and targets, but cannot deal with a varying number of targets. In [7, 8], a pairwise Markov random field (MRF) motion prior is introduced to model the interaction between targets, however these MRFs are built only on the observations at the same time instant, and therefore require good detection of moving regions.

The paper is organized as follows: We formulate the general multiple targets tracking problem and introduce Adaboost image likelihood in section 2. In section 3 we present our MCMC data association algorithm for efficiently searching for the trajectory of moving object. Section 4 provides some experimental results UAV data set.

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00

2. Multiple-Target Tracking Problem

Let $T \in Z^+$ denote the duration of tracking. From time $[0, T]$, there are K unknown number of targets in the monitored scene. Let $y_t = \{y_t^i : i = 1, ..., n_t\}$ denote the observations at time t, $Y = \cup_{t\in\{1,...,T\}} y_t$ is the set of all the observations in duration $[0, T]$. The tracking problem can be formulated as maximizing a posterior (MAP) of a partition $\omega = \{\tau_0, \tau_1, \tau_2, ..., \tau_K\}$ given the set of observations Y over time T such that:

$$\omega^* = \arg\max(p(\omega|Y)) \tag{1}$$

where τ_0 is the set of false alarms, τ_k is the track k among K tracks from the given partition.

We utilize a graph representation of all measurements within the time frame $[0, T]$. The partition can be explicitly drawn from this measurement graph (V, E), where each measurement y_t^i is represented by a node in V, and each edge corresponds to a temporal association reflecting spatial properties such as spatial overlap between detected regions. We define a neighborhood in the graph (V, E) where edges are defined between any two neighboring nodes:

$$N = \{(y_{t_1}^i, y_{t_2}^j) : \left\| y_{t_1}^i - y_{t_2}^j \right\| < t \cdot v_{max})\} \tag{2}$$

where $\|\cdot\|$ is the Euclidean distance, $t = |t_1 - t_2| \in [0, ..., t_{\max}]$ and v_{max} is the maximum speed of targets.

The posterior distribution for the partition with unknown number of targets and observations over T frames can be modeled as:

$$P(\omega|Y) = \frac{1}{z} \prod_{k=1}^{K} \psi(\tau_k)\varphi(\tau_k, Y) \prod_{j\neq k} \phi(\tau_k, \tau_j) \tag{3}$$

where $\psi(\tau_k)$is the temporal compatibility within one track, $\varphi(\tau_k, Y)$ denote the local evidence (likelihood) for one target and $\phi(\tau_k, \tau_j)$ is the spatial compatibility between different targets respectively. The posterior distribution in Eq. 3 can be viewed as having three distinct components: (i) $\psi(\tau_k)$ controlling the inner-smoothness for each track encoded by the joint motion and appearance likelihood (ii) $\varphi(\tau_k, Y)$ incorporating a local image likelihood from model information (iii) $\phi(\tau_k, \tau_j)$ encoding the interaction between different tracks. We will now discuss each one of these in turn.

2.1. Motion and Appearance Model

In this paper, the targets are represented by image blobs. Once a partition ω is chosen, the tracks $\{\tau_1, ..., \tau_K\}$ and false alarms τ_0 are determined and for each track the assigned observations are determined. To make full use of the observations for target tracking, we consider a joint probability framework for incorporating both motion and appearance information. Therefore the $\psi(\tau_k)$ in Eq.3 can be represented as follows.

$$\psi(\tau_k) = \prod_{l=1}^{|\tau_k|-1} P_{motion}(\tau_k(t_{l+1})|\bar{\tau}_k(t_l))P_{app}(\tau_k(t_{l+1}|t_l)) \tag{4}$$

For the motion likelihood, we consider a linear kinematic model:

$$\begin{aligned} x_{t+1}^k &= A^k x_t^k + w_t^k \\ y_t^k &= H^k x_t^k + v_t^k \end{aligned} \tag{5}$$

where x_t^k is the kinematic state, y_t^k the measurement, and w_t^k, v_t^k are Gaussian process noise and observation noise for target k at time t. Let $\bar{\tau}_k(t_i)$ denote the prior estimated states at time t_i of τ_k and $\bar{P}_{t_{i+1}}(\tau_k) = A^k \hat{P}_{t_i}(\tau_k)(A^k)^T + Q_{t_{i+1}}^k$ denote the prior covariance of τ_k estimated at time t_i. The motion likelihood of track τ_k at time t can then be written as:

$$P_{motion}(\cdot) = \frac{1}{(2\pi)^{N_s} \left|P_{t_{i+1}}(\tau_k)\right|} \exp\left(\frac{-e^T P_{t_{i+1}}^{-1}(\tau_k)e}{2}\right)$$

where $e = \tau_k(t_{i+1}) - \bar{\tau}_k(t_{i+1})$.

In order to model the appearance of each detected regions while preserving localization of the features, we adopt the appearance model proposed in [3]. This descriptor is invariant to 2D rigid transformations and scale change, and tolerates small shape variations.

2.2. Interaction model

The motion and appearance likelihood models provide the inner-smoothness constraint for each track independently. However, without an *a priori* knowledge of the number of targets, the inner-smoothness constraint will favor shorter paths, and therefore will split a trajectory into a large number of sub-tracks. To overcome this overfitting problem, commonly a *prior* knowledge on the detection and the targets' behavior (such as detection and false alarm rate, termination and birth rate etc.) is assumed known [6, 9].

We propose to use an interaction model that penalizes object overlapping based on Markov random fields (MRFs) [10, 8] defined on the neighborhood graph. The joint interaction between all existing nodes over time is factored as the product of local potential functions at each node. In this MRF, the cliques are pairs of nodes that are connected in the graph (V, E). The interaction potential between τ_k and τ_j is defined by:

$$\phi(\tau_k, \tau_j) = \prod_{l=1}^{|\tau_k-1|} \prod_{m=1}^{|\tau_j-1|} \exp(-\lambda\rho(\tau_k(l), \tau_j(m)) \tag{6}$$

where ρ is the spatial overlap between two observation nodes. The interacting potential is minimum when the observations have a large spatial overlap and maximum when they do not overlap. The introduction of the inter-track exclusion will prevent from splitting tracks into smaller tracks when a good overlapping of the regions exists.

2.3. Adaboost Image likelihoods

Due to the erroneous motion segmentation, the persistent false alarms may exist, especially when parallax, illumination changing, inaccurate motion compensation happen. In order to reduce the false alarm by using the

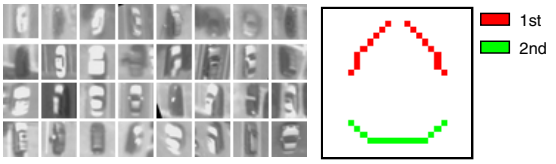
model information, we introduce the likelihood $\varphi(\tau_k, Y)$ in Eq.3, which models the probability of observing the image Y conditioned on the τ_k given one possible partition, i.e.$\varphi(\tau_k, Y) = \prod_{l=1}^{|\tau_k|} p_{model}(y_i)$. We build our likelihood model using a boosted classifier. In our experiment, we adopt the Edgelet features [11] though other simpler features, like Haar may be used. Given a set of labeled patterns (x, y), the AdaBoost procedure learns a weighted combination of base weak classifiers, $H(x) = \sum_{i=1}^{T} \alpha_i h_i(x) - b$, where x is an image pattern, and $h_i(x)$ is the weak classifier chosen for the round i of boosting, and α_i is the corresponding weight.. In order to provide each moving region a likelihood instead of boolean classification, we use a real-valued version of Adaboost algorithm [12]. Following [11], we divide the range of edgelet feature $f_{edgelet} \in [0, 1]$ into n bins. The weak classifier can be formulated as:$h_i(I) = \frac{1}{2} \sum_{j=1}^{n} \ln(\frac{W_{+1}^j + \xi}{W_{-1}^j + \xi}) \delta_j(f_{edgelet}(x))$, where $W_l^j = P(f_{edgelet}(x) \in bin_j, y = l)$, δ_j is the impulse function for bin j.

Considering the diversity of viewpoint in UAV video, we rectify the targets' heading when we collect positive samples and make samples' orientation cover other degree of freedom except heading. The positive samples are shown in Figure1(a). Given a possible track τ_k, we approximate the target's orientation by the direction of the trajectory. By applying the real-valued Adaboost detector on the possible orientation, we can assign each preliminary moving blob a likelihood which is the maximum response located in the image blob at different scale.The first two features trained are shown in Figure 1(b). After the images are stabilized using an affine motion model [2], the motion segmentation are shown in Figure 2(a) and binary Adaboost detection are shown in Figure 2.

3. MCMC Data association Algorithm

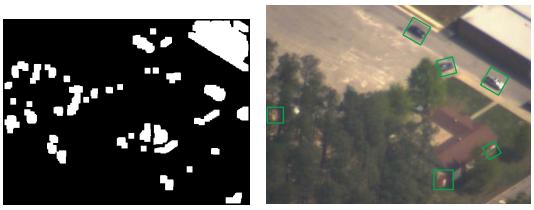
The proposed tracking approach defined in Section 2, formulates the tracking as a Bayesian inference, whose purpose is to find the maximum a posteriori estimate (MAP) given by Eq 1. We use a data-driven MCMC for estimating the best partition of the space Ω. The sampling will be guided by the posterior distribution defined in Eq.3, and convergence will be guaranteed by the Markov chain properties. Since the number of tracks is a unknown *priori*, we propose to drive the sampler, in a probabilistic manner, using the measurement graph. Moreover, in order to make the sampler more efficient, we draw samples in both temporal directions: looking forward and backward in time. This bidirectional sampling gives more flexibility and reduces significantly the total number of samples.

We use the following notations on the graph structure: $N(\cdot)$ is the neighbor set of an observation, i.e. $N(y_{t_1}^i) = \{y_{t_2}^j, (y_{t_1}^i, y_{t_2}^j) \in E\}$; Observation $y_{t_2}^j \in N(y_{t_1}^i)$ belongs to the parent set $N^c(y_{t_1}^i)$, child set $N^p(y_{t_1}^i)$ exclusively, when $t_2 < t_1$ or $t_2 > t_1$.



(a) Positive training samples (b) The first two features trained

Figure 1. Edgelet Adaboost training



(a) Foreground region (b) Binary detection

Figure 2. Motion and pattern detection

Extension/Reduction: The purpose of the extension/reduction move is to extend or shorten the estimated trajectories given a new set of observations. For the forward extension, we select uniformly at random (*u.a.r*) a track τ_k from K available tracks, $\tau_1, ..., \tau_K$. Let $\tau_k(end)$ denote the last node in the track τ_k. For each node $y \in N^c(\tau_k(end))$, we have the association probability $p(y) = \frac{p(y|\bar{\tau}_k(end))}{\sum_z p(y|\bar{\tau}_k(end))}$. We associate y and track τ_k according to this normalized probability, and then append the new observation y to τ_k with a probability γ, where $0 < \gamma < 1$. Similarly, for a backward extension, we consider a node $y \in N^p(\tau_k(start))$ and use reverse dynamics for estimating the association probability $p(y)$.

The reduction move consists of randomly shortening a track τ_k (*u.a.r* from K available tracks, $\tau_1, ..., \tau_K$), by selecting a cutting index r *u.a.r* from $2, ..., |\tau_k| - 1$. In the case of a forward reduction the track τ_k is shortened to $\{\tau_k(t_1), ..., \tau_k(t_r)\}$, while in a backward reduction we consider the sub-track $\{\tau_k(t_r), ..., \tau_k(t_{|\tau_k|})\}$.

Birth/Death: This move controls the creation of new track or termination of an existing trajectory. In a birth move, we select *u.a.r* a node $y \in \tau_0$, associate it to a new track and increase the number of tracks $K' = K + 1$.

The birth move is always followed by a extension move. From the node y we select the extension direction forward or backward *u.a.r* to extend the track $\tau_{K'}$. Similarly, in a death move we choose *u.a.r* a track τ_k and delete it. The nodes belonging to the deleted track are added to the unassigned set of measurements τ_0.

Split/Merge: In a split move, we *u.a.r* select a track τ_k, and a split point t_s, which is selected according to the normalized joint probability between two consecutive connected nodes in the track:

$$(1 - p(\tau_k(t_{i+1})|\bar{\tau}_k(t_i))) \,/\, \sum_{i=1}^{|\tau_k|-1} (1 - p(\tau_k(t_{i+1})|\bar{\tau}_k(t_i))).$$

And we split τ_k into two new tracks $\tau_{s_1} = \{\tau(t_1),...,\tau(t_s)\}$ and $\tau_{s_2} = \{\tau(t_{s+1}),...,\tau(t_{|\tau_k|})\}$.

Often, due to missing detection or erroneous detection, trajectories of objects are often fragmented. The merge move provides the ability to link these fragmented sub-tracks according to their joint likelihood of appearance and motion and the interaction based on spatial overlapping. The merge move operates on the candidate set of track pairs, for which the start node of one track is the child node of the end node of the other track and is defined by the set: $C^t_{merge} = \{(\tau_{k_1}, \tau_{k_2}) : \tau_{k_2}(start) \in N^c(\tau_{k_1}(end))\}$. We select u.a.r pairs of tracks from C^t_{merge} and merge the two tracks into a new track $\tau_k = \{\tau_{k_1}\} \cup \{\tau_{k_2}\}$

Switch: In a switch move, we are probing the solution space for better labeling of nodes that belong to multiple tracks. We consider the following candidate set of track pairs.

$$C^t_{switch} = \{(\tau_{k_1}(t_p), \tau_{k_2}(t_q)) : \tau_{k_1}(t_p) \in N^p(\tau_{k_2}(t_{q+1})), \tau_{k_2}(t_q) \in N^p(\tau_{k_1}(t_{p+1})), k_1 \neq k_2\}. \quad (7)$$

We *u.a.r* select a candidate node from C^t_{switch} and define two new tracks as:
$\tau'_{k_1} = \{\tau_{k_1}(t_1), ..., \tau_{k_1}(t_p), \tau_{k_2}(t_{q+1}), ..., \tau_{k_2}(t_{|\tau_{k_2}|})\}$ and
$\tau'_{k_2} = \{\tau_{k_2}(t_1), ..., \tau_{k_2}(t_q), \tau_{k_1}(t_{p+1}), ..., \tau_{k_1}(t_{|\tau_{k_1}|})\}$.

Although the switch move can be implemented by two times temporal and split and two times temporal moves, the switch move is introduced to reduce the number of samplings and thus accelerate the convergence.

4. Experimental Results

We have implemented the proposed association algorithm as an online algorithm within a sliding window which contains the latest 45 frames and only observations within this sliding window are stored in the measurement graph. The partition of the current graph is initialized with the solution obtained at $t-1$, *i.e.* $\omega^t_{init} = \omega^{t-1}_{best}$. The partition of graph at $t = 0$ is initialized by a greedy criteria. The MCMC sample was run for a total of 250 iterations.

Our training set contains 1,200 vehicle figures, which are from UAV data set. The figures are aligned according to the center and heading of vehicles. The samples are scaled to have a resolution of 24x24 pixels. Our negative image set contains 500 negative images from UAV video without vehicles. The classifier used in our system has 80% detection rate with 10^{-4} false alarm rate. Although more positive samples of vehicle will lead to better Adaboost result, the real-valued Adaboost classifier, combining with the motion segmentation and tracking, significantly decreases the false alarm rate in our UAV target tracking. Some illustrations of the obtained trajectories are shown in Figure 3, where color corresponds to object ID.

Figure 3. Tracking result.

5. Conclusions

In this paper, we propose a multiple hypothesis framework to find a global optimal partition of observations in the measurement graph. A data driven MCMC method is used to solve this combinatorial optimization problem. The proposal distribution utilizes a mixture model that incorporates information from the dynamics, appearance of each player and the pattern model trained by Adaboost.

References

[1] W. Gilks, S. Richardson, and D.J.Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman and Hall, 1996.

[2] J. Kang, I. Cohen, and G. Medioni. Continuous tracking within and across camera streams. In *CVPR*, volume 1, pages 267–272, Jun 2003.

[3] J. Kang, I. Cohen, and G. Medioni. Object reacquisition using invariant appearance model. In *ICPR*, pages 759–762, 2004.

[4] Z. Khan, T. Balch, and F. Dellaert. Multitarget tracking with split and merged measurements. In *CVPR*, volume 1, pages 605–610, 2005.

[5] I. Cox and S. Hingorani. An efficient implementation of reids multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. In *ICPR*, pages 437–443, 1994.

[6] S. Oh, S. Russell, and S. Sastry. Markov chain monte carlo data association for general multiple-target tracking problems. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, 2004.

[7] Z. Khan, T. Balch, and F. Dellaert. An mcmc-based particle filter for tracking multiple interacting targets. In *ECCV (4)*, pages 279–290, 2004.

[8] K. Smith, D. Gatica-Perez, and J.-M. Odobez. Using particles to track varying numbers of interacting people. In *CVPR*, pages 962–969, 2005.

[9] Y. Bar-Shalom, T. Fortmann, and M. Scheffe. Joint probabilistic data association for multiple targets in clutter. In *Proc. Conf. on Information Sciences and Systems*, 1980.

[10] Z. Khan, T. Balch, and F. Dellaert. Mcmc-based particle filtering for tracking a variable number of interacting targets. *IEEE PAMI*, (11):1805–1918, 2005.

[11] B. Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In *In ICCV*, pages IV: 90– 97, 2005.

[12] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, pages 297–336, 1999.