Prioritizing the Propagation of Identity Beliefs for Multi-object Tracking

Amit Kumar K.C.
http://www.uclouvain.be/amit.kc
Christophe De Vleeschouwer
http://www.uclouvain.be/christophe.devleeschouwer

ICTEAM Institute
Université catholique de Louvain
Louvain-la-neuve, Belgium

Multi-object tracking requires locating the targets as well as labelling their identities. Inferring identities of the targets is a challenge when the availability and the reliability of the appearance features do vary along the time and the space. We see the multi-object tracking and identification as a two-stage process.

In the first stage, plausible target candidates are detected at each frame independently, and are aggregated into tracklets. The benefits obtained from such aggregation process are twofold. First, it reduces the number of entities that have to be processed later. Second, it provides more reliable and more accurate knowledge about the appearance of the target observed along the tracklet.

In the second stage, which embeds the main contributions of the paper, a graph-based belief propagation formalism is considered to estimate the identity of each tracklet. Each node in the graph corresponds to a tracklet, and is assigned a probability distribution of identities, based on the tracklet appearance, and given prior knowledge of the possible target appearances. Typically, a low confidence in the tracklet appearance measurement, or a measurement that is similar to several target appearances, both result into a flat and thus ambiguous identity distribution for the tracklet. Afterwards, belief propagation is considered to infer the identities of more ambiguous nodes from those of less ambiguous nodes, by exploiting the graph constraints. In contrast to the approaches with standard belief propagation [2], which treats the nodes in an arbitrary order, the proposed method schedules less ambiguous nodes to transmit their messages first.

From appearance features to identity distribution We assume that there are N targets, each of them being characterized by K appearance features. The feature set for the j-th target is $\mathcal{F}^{(j)} = \{\mathbf{f}_1^{(j)}, ..., \mathbf{f}_K^{(j)}\}$. Let the appearance features for a tracklet v be $\overline{\mathcal{F}}^{(v)} = \{\bar{\mathbf{f}}_1^{(v)}, ..., \bar{\mathbf{f}}_K^{(v)}\}$. Then, the probability of the tracklet v having identity j, denoted by $p_v(j)$, as

$$p_v(j) \propto \prod_{i=1}^{K} \exp\left[-\| \mathbf{f}_i^{(j)} - \bar{\mathbf{f}}_i^{(v)} \|_1 / \tau_i^{(v)}\right] \qquad \text{for } 1 \leq j \leq N \tag{1}$$

where $\tau_i^{(v)}$ monitors the influence of feature i on identity assignment. It decreases as the appearance feature observation becomes more reliable. Depending on the observed appearance features and on the estimated reliability of these observations, some tracklets have less ambiguous identity distributions than others.

Graph definition The tracklets are gathered into a graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of nodes, with each node corresponding to a tracklet; $\mathcal{E}$ is a set of edges, defining the connectivity between the nodes in $\mathcal{V}$. An edge between nodes u and v implies that their identities are dependent. For example, two tracklets, which co-exist at the same time, should belong to two different physical targets. This defines a *mutex* edge between them. Additionally, if they are sufficiently close in space, time and/or appearance, they are likely to share the same identity, whereas if they are far, they should be encouraged to have different labels. This defines a *temporal* edge between them. Each node $v \in \mathcal{V}$ and each edge $(uv) \in \mathcal{E}$ is characterized by potential functions ϕ_v and ϕ_{uv} respectively. In short, $\phi_v(l_v)$ represents how likely is the label l_v be assigned to the node v. Similarly, $\phi_{uv}(l_u, l_v)$ represents the likelihood that nodes u and v have labels l_u and l_v respectively.

Belief propagation We briefly introduce how the belief propagation formalism works. A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is given, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ represents the association between the nodes. The neighbourhood of node $v \in \mathcal{V}$ is denoted by $\mathcal{N}_v$. The purpose of belief propagation is to find a labelling function l that labels each node $v \in \mathcal{V}$ with a label $l_v \in \mathcal{L}$, $\mathcal{L}$ being the set of possible labels, so as to maximize the joint likelihood function:

$$p(l) \propto \prod_{v \in \mathcal{V}} \left[\phi_v(l_v) \prod_{u \in \mathcal{N}_v} \phi_{uv}(l_u, l_v)\right] \tag{2}$$

It is done iteratively by exchanging "messages" between the nodes. Let $\mathbf{m}_{u \to v}^{(t)}$ be the message that the node u sends to a neighbouring node v at iteration t. Intuitively, $m_{u \to v}^{(t)}(l_v)$ is the belief that node u thinks about the label l_v of node v at any iteration t. Each message is initialized uniformly. Afterwards, new messages are updated (in sum-product form) at each iteration. as:

$$m_{u \to v}^{(t)}(l_v) \propto \sum_{l_u \in \mathcal{L}} \left[\phi_{uv}(l_u, l_v)\phi_u(l_u) \prod_{s \in \mathcal{N}_u \setminus v} m_{s \to u}^{(t-1)}(l_u)\right] \tag{3}$$

After T iterations, a belief vector $\mathbf{b}_v$ is computed for node v as

$$b_v^{(T)}(l_v) \propto \phi_v(l_v) \prod_{s \in \mathcal{N}_v} m_{s \to u}^{(T)}(l_v) \tag{4}$$

from which the most likely identity is estimated.

Construction of potential terms We briefly explain how we design the potential terms in our application scenario. The unary potential term $\phi_v(l_v)$ is defined to be the likelihood of the node $v \in \mathcal{V}$ having a label l_v. It is given as: $\phi_v(l_v) = p_v(l_v)$, $l_v \in \mathcal{L}$ (ref. Eqn 1). In case of mutex edges, u and v should have different labels. Therefore,

$$\phi_{uv}(l_u, l_v) = \begin{cases} \varepsilon & \text{if } l_u = l_v \\ 1 - \varepsilon & \text{otherwise,} \end{cases} \tag{5}$$

We use $\varepsilon = 0.1$. We express ϕ_{uv} for temporal edges in terms of the distance d_{uv} as

$$\phi_{uv}(l_u, l_v) = \begin{cases} \exp(-d_{uv}/\tau_{\text{dist}}) & \text{if } l_u = l_v \\ 1 - \exp(-d_{uv}/\tau_{\text{dist}}) & \text{otherwise,} \end{cases} \tag{6}$$

where τ_{dist} is a constant. If both u and v have reliable identity estimate, then the Bhattacharyya distance between the belief vectors, $\mathbf{b}_u$ and $\mathbf{b}_v$, is used to define d_{uv}. On the other hand, if one of the nodes does not have reliable identity estimate, then the computation of the Bhattacharyya distance is irrelevant. In such cases, when the nodes are close in time, the position information is used to measure their distance. In contrast, when the nodes are far in time, even the position cannot guide the definition of the distance. In this case, no message is exchanged between the nodes.

Priority scheduling of belief message exchanges To emphasize our contribution, we make two observations about the standard belief propagation: (i) nodes are arbitrarily selected to send messages, (ii) a node gathers information from all its neighbours. However, in our graph formulation, some nodes are less ambiguous about their identities than others. The messages sent by such nodes are more informative. Hence, they help the more ambiguous neighbours to disambiguate their labels [1]. Moreover, during the message construction step, since the messages coming from more ambiguous nodes are usually uninformative and even confusing, we strictly restrict gathering of messages from less ambiguous nodes as shown in Figure 1.

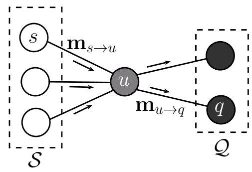

Figure 1: Message construction and dissemination at node u (in **gray**). The node u gathers information from its less ambiguous neighbors, $\mathcal{S}$ (in **white**). Afterwards, u transmits message to its more ambiguous neighbors, $\mathcal{Q}$ (in **black**).

Given the current estimate of belief vector $\mathbf{b}_v$, we use *entropy* of the belief vector to measure the ambiguity level of a node. Then, nodes are sorted in increasing order of entropies.

Results Experimental validation is performed on 10 minutes long real-life basketball video. The proposed method achieves 89% identification rate, which is an improvement of 21% and 16% compared to individual identity assignment, and to standard belief propagation, respectively. Please refer to the main paper and the supplementary paper for detailed analysis.

[1] N. Komodakis and G. Tziritas. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *Image Processing, IEEE Transactions on*, 16(11):2649 –2661, nov. 2007.

[2] Wei-Lwun Lu, Jo-Anne Ting, Kevin P. Murphy, and James J. Little. Identifying players in broadcast sports videos using conditional random fields. In *CVPR*, 2011.