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

Abstract

Multi-object tracking requires locating the targets as well as labeling their identities. Inferring identities of the targets from their appearances is a challenge when the availability and the reliability of the observation process do vary along the time and space.

The purpose of this paper is to assign identities to those appearance measurements using a graph-based formalism. Each node of the graph corresponds to a tracklet, which is defined to be a sequence of positions that very likely correspond to the same physical target. Tracklets are pre-computed, *e.g.* using [2], and our work investigates how to assign them identities, knowing the reference appearance of each target. Initially, each node is assigned a probability distribution over the set of possible identities, based on the observed appearance features. Afterwards, belief propagation is considered to infer the identities of more ambiguous nodes from those of less ambiguous nodes, by exploiting the graph constraints and the measures of similarities between the nodes. In contrast to the standard belief propagation, which treats the nodes in an arbitrary order, the proposed method uses a priority-based belief propagation, in which less ambiguous nodes are scheduled to transmit their messages first.

Validation is performed on a real-life basketball dataset. 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.

1 Introduction

Multi-object tracking is a fundamental issue in computer vision. Reliable tracking and identification of targets indeed support higher-level scene analysis and interpretation. For example, vehicle trajectories are collected to control traffic monitoring solutions [13]. People displacement analysis is important to improve the security of public places [14], or to support sport team game analysis, *e.g.*, for autonomous production [5].

We consider the very common scenario for which the information captured by a (set) of visual sensor(s) is exploited to track and assign identities to targets, based on a set of discriminant appearance features. In many practical scenarios, the features cannot always be reliably and accurately estimated from the observation of the scene. They are subject to non-stationary noise, and their ability to discriminate objects varies with the scene context. For example, color features tend to be quite noisy in presence of occlusions and clutters. In some other cases, highly discriminant features are only available sporadically. This occurs, for example, when reading the digit on players' jerseys as it is available only when the jersey is facing the camera.

In this paper, we see multi-target tracking and identification as a two-stage process.

In the first stage, the plausible target candidates are first detected independently in each frame, *e.g.*, based on [6]. Each detection is characterized by a set of features and their corresponding confidence values. Typically, the confidence assigned to a feature depends on various factors, *e.g.*, whether the detection is occluded and/or visible on the camera view or not, whether the detection is close to the camera view or not, etc. Afterwards, the detections are aggregated into tracklets, which are defined to group consecutive detections that obviously correspond to the same physical object. The practical implementation of this aggregation process is a research topic by itself. In this paper, we built on the solution described in [2] for this step, but any other alternative could be envisioned [3, 15]. The benefits obtained from such aggregation process are twofold. First, it reduces the number of entities that have to be processed in the second stage. Second, it provides more reliable and more accurate knowledge about the target appearance, since a target is observed several times along its 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 confidence assigned to the observed tracklet appearance. Typically, a low confidence in the tracklet appearance measurement, or a measurement that is similar to several targets, both result into a flat and thus ambiguous identity distribution for the tracklet. An edge between two tracklets represents that their identities are dependent, meaning that the knowledge of the identity of one tracklet brings some information about the identity of the other. This usually happens in two cases: (a) when the tracklets co-exist at the same time, and (b) when the tracklets are sufficiently close in space, time and/or appearance. The belief propagation module exploits the graph structure and the similarity between the nodes to compute the posterior probability distribution of identities at the nodes. We view this as an inference problem, looking for the most likely identity of a tracklet, given the identities of the other tracklets. As a main contribution, our paper introduces an original scheduling mechanism to order the propagation of identity beliefs along the graph. In short, the intuition behind our approach consists in propagating the less ambiguous identity information before more ambiguous identity cues.

The ability to drive the belief propagation, based on the level of ambiguity associated to the identity of each node, differentiates our work from most earlier works dealing with identity assignment. Two examples of related earlier works are [12] and [10]. In [12], the authors assume that a track-graph, denoting when targets are isolated and describing when they interact, exists. Assuming that the feature vectors of isolated tracks are reliable, they define a measure the similarity between them. Later, they formulate the identity linking as a Bayesian inference problem in order to find the most probable set of paths. For this, they use standard message passing technique. In [10], the authors use a conditional random field to identify players in broadcast sport videos. For this, they use a set of features, like SIFT interest points, MSER regions, color histograms, etc. in order to propagate easy-to-classify images of a player to other images.

We differentiate our work from the above works in the sense that the authors in both papers assume that the feature descriptors have the same reliability throughout the time. Therefore, those earlier works do not account for the ambiguity of the identity assignment during message passing.

The remainder of the paper is organized as follows. Section 2 first reviews the work of [2], to explain how the detections are aggregated into tracklets, and how each tracklet

gets an identity distribution based on its appearance. Section 3 provides a brief introduction to the standard belief propagation, and then describes the proposed priority-based belief propagation. Finally, an experimental validation is provided and discussed in Section 4.

2 Tracklet definition and prior identity distribution

In this section, we introduce how the inputs to our identity assignment problem are computed. Assuming that the candidate targets are independently detected at each time instant, we briefly explain how they are aggregated into short tracks of detections that quite likely correspond to the same physical object (called tracklets). We then define how the prior identity probability distribution of each tracklet is estimated, based on the accumulation of appearance cues observed along the tracklet.

2.1 Tracklet definition

Given a set of candidate detections, the tracking is generally formulated as a data-association problem in a graph [2, 8, 15]. As introduced earlier, the detections have a set of appearance features and their corresponding confidence values. The confidence value of a feature reflects the reliability of its measurement. To link such detections, we follow an iterative aggregation strategy, as in [2]. Starting with a graph in which each detection is a node, the strategy progressively aggregates the nodes of the graph into bigger nodes, named tracklets. Each node corresponds to a tracklet. Specifically, each iteration considers a node, named key-node, and investigates how to link it with either previous or subsequent nodes, assuming that the appearance of the key-node is the appearance of the target. This hypothesis testing procedure computes a list of shortest-paths to/from the key-node. Only the path that is significantly better than the other alternative paths defines a tracklet. The main advantage of such a strategy is that it can benefit from the appearance features that are sporadically available, or affected by a non-stationary noise, along the sequence of detections.

The outcome of the aggregation process is a set of tracklets. Formally, each tracklet v is characterized by the following features:

- The positions of the starting and ending, denoted as $\mathbf{x}_v^{(s)}$ and $\mathbf{x}_v^{(e)}$ respectively,
- The starting and ending time of the tracklet, denoted as $t_v^{(s)}$ and $t_v^{(e)}$ respectively,
- The average appearance features of the object detected along the tracklet, and their corresponding confidence values. These features give an initial estimate of its identity distribution, as discussed in the next section.

2.2 Assigning identity distribution based on appearance features

In the previous section, we have presented how the tracklets are defined. This section explains how to assign an identity distribution to each of them.

We assume that there are N targets, each of them being characterized by K appearance features, which are assumed to be known *a priori*. Nevertheless, the appearance features can be learned automatically too [16].

Let the ith feature of the jth target be denoted by $\mathbf{f}_i^{(j)}$, $1 \leq j \leq N$. Then, the feature set for the jth target is $\mathcal{F}^{(j)} = \{\mathbf{f}_1^{(j)}, ..., \mathbf{f}_K^{(j)}\}$. For example, in a basketball match, color and digit on the jersey of the player can be considered as features. In this case, $K = 2$ and $\mathcal{F} = \{\text{color}, \text{digit}\}$.

The appearance of a tracklet is defined by averaging the appearance features, measured in each detection of the tracklet. Let the average appearance features for a tracklet v be denoted by $\overline{\mathcal{F}}^{(v)} = \{\bar{\mathbf{f}}_1^{(v)}, ..., \bar{\mathbf{f}}_K^{(v)}\}$. This allows us to define the probability of the tracklet v having identity j, denoted by $p_v(j)$, as

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

where $\tau_i^{(v)}$ weighs the influence of feature i on identity assignment, and is related to the confidence assigned to the ith feature of the tracklet v. It decreases when the appearance feature becomes more accurate and reliable.

When τ_i is very large or when the appearance of a tracklet is far from all target appearances, the probability distribution, in Equation 1, tends to uniform distribution, i.e., the identity assignment becomes ambiguous as all identities are equally likely. Conversely, when the tracklet appearance is closer to a target appearance, the probability distribution becomes peaky around that identity, implying that the identity assignment is less ambiguous. Consequently, depending on the observed appearance features and their confidence values, some tracklets have less ambiguous identity distributions than others.

3 Belief propagation

In this section, belief propagation is considered to exchange identity information between the tracklets. The purpose is to compute posterior identity probabilities by merging the prior identity distributions and exploiting the graph structure. We first survey the principles of belief propagation. We then present how the belief propagation graph is constructed in our application scenario. Eventually, we introduce the main contribution of our paper, which lies in an original priority-based scheduling mechanism to select the nodes from which the identity information is propagated.

3.1 Standard belief propagation

The framework for the belief propagation technique is defined as follows [7, 10]. An undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is given, where $\mathcal{V}$ is the set of nodes in the graph and $\mathcal{E}$ represents the association between the nodes. The neighborhood of node $v \in \mathcal{V}$ is denoted by $\mathcal{N}_v$.

We assume that each node $v \in \mathcal{V}$ and each edge $(u, v) \in \mathcal{E}$ are associated with potential functions ϕ_v and ϕ_{uv} respectively. The unary potential $\phi_v(l_v)$ is the likelihood of node v having label l_v, and the pairwise potential $\phi_{uv}(l_u, l_v)$ is the likelihood that the nodes u and v have labels l_u and l_v respectively.

The purpose of belief propagation is to find a labeling function l that labels each node $v \in \mathcal{V}$ with a label $l_v \in \mathcal{L}$, $|\mathcal{L}|$ being the total number of labels, so as to maximize the joint likelihood function:

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

Generally, it is done iteratively by exchanging "messages" between the nodes. Let $\mathbf{m}_{u\to v}^{(t)} \in [0,1]^{|\mathcal{L}|}$ be the message that the node u sends to a neighboring 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:

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

where $h_u(l_u)$ is the information gathered at node u about the label l_u. It is also referred to as the "pre-message" for l_u. Alternatively, the summation term of the Equation 3 is replaced by max term and is referred as *max-product form*.

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

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

Finally, the normalized belief vector $\mathbf{b}_v^{(T)}$ is considered as the estimate of the identity distribution of the node v. The complexity of message passing is $\mathcal{O}\big(|\mathcal{V}||\mathcal{L}|^2 T\big)$. It requires $\mathcal{O}\big(|\mathcal{L}|^2\big)$ to compute each message, there are $\mathcal{O}\big(|\mathcal{V}|\big)$ messages to compute at each iteration, and there are T iterations [7].

3.2 Graph of identity beliefs and definition of potential terms

In this section, we will first explain the graph formalism for the belief propagation. Later, we will explain how the potential terms are constructed in our application scenario.

As discussed in Section 2, the output of aggregation step is a set of tracklets. Each tracklet has its starting and ending time-stamps and positions. Moreover, each tracklet is assigned an initial identity distribution, based on the observed appearance features. These 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. Thus, knowing the identity of one brings some information about the other. Specifically, the support of the tracklets can be used to enforce the constraint that two tracklets, which co-exist at the same time, should belong to two different physical targets. This defines a *mutex* edge between them. Additionally, the knowledge of the extremities of the tracklets (and their appearances) enables us to estimate the proximity between them such that if the tracklets are sufficiently close, they are likely to share the same identity. In contrast, if they are significantly far apart, they are encouraged to have different identities. This defines a *temporal* edge between them. An example is elucidated in Figure 1.

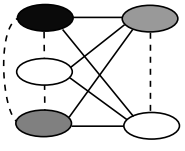

Figure 1: Each node has an identity distribution: **white** - a peaky distribution, **black** - a flat distribution, and **gray** - an intermediate distribution. Solid lines represent temporal edges whereas mutex edges are shown in dotted lines.

Now, we explain how we associate the potential terms to each node and edge. 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. As the prior identity distribution $p_v(l_v)$, computed by Equation 1, represents how likely the label l_v is, we use it as the estimate of the unary potential. That is, $\phi_v(l_v) = p_v(l_v)$, $l_v \in \mathcal{L}$.

The pairwise potential term ϕ_{uv} is defined to reflect that the identities of u and v are dependent. Typically, in our practical scenario, it is defined such that, when u and v are likely to correspond to the same physical target (*e.g.*, because they are close in appearance or in space and time), $\phi_{uv}(l_u, l_v)$ tends to zero for $l_u \neq l_v$, and to 1 for $l_u = l_v$. In contrast, when they are likely to correspond to different physical targets (*e.g.*, because they co-exist in time), $\phi_{uv}(l_u, l_v)$ should be defined so that $\phi_{uv}(l_u, l_v)$ tends to zero for $l_u = l_v$, and to 1 for $l_u \neq l_v$.

Following those general principles, we define the potentials over the mutex and temporal edges in our graph structure as follows. In case of mutex edges, u and v should have different labels. Therefore,

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

where ε is a small positive number. Setting $\varepsilon = 0$ enforces that the identities of the nodes are unique at a given time. However, it is possible that two nodes share the same identity. This, for example, happens when a tracklet incorrectly aggregates (see Section 2) detections that correspond to different physical targets, resulting in a mixed identity distribution. In such cases, imposing $\varepsilon = 0$ might lower the performance of the system. In our settings, we use $\varepsilon = 0.1$.

In case of temporal edges, we express ϕ_{uv} in terms of the *distance* d_{uv} between them. When the distance between the nodes is small, they should be encouraged to share the same label and vice versa. We define

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

where τ_{dist} is a constant. Now, we turn our attention towards the definition of the distance between the nodes. Two cases are distinguished. On the one hand, if both u and v have reliable identity estimate, then the computation of the Bhattacharyya distance between the belief vectors $\mathbf{b}_u$ and $\mathbf{b}_v$ is used to define the distance between the nodes. We denote the entropy of belief vector of a node u by $E(u)$. Then, if $E(u) < \tau_{\text{TH}}$ and $E(v) < \tau_{\text{TH}}$ then,

$$d_{uv} = \left[1 - \sum_{l_v \in \mathcal{L}} \sqrt{b_u(l_v) b_v(l_v)} \right]^{1/2} \tag{7}$$

We set $\tau_{\text{TH}} = 0.5$ and $\tau_{\text{dist}} = 0.3$. The choice of τ_{TH} is not critical, as long as it is chosen small enough, as illustrated in the supplementary material.

On the other hand, if one of the nodes does not have reliable identity estimate, then the computation of the Bhattacharyya distance between the belief vectors is irrelevant. In such cases, when the nodes are close in time (i.e., $|t_v^{(s)} - t_u^{(e)}| < \tau_{\text{max}}$), the position information is used to measure their distance d_{uv}. If they are far in space, they should have different identities, and conversely if they are close in space, they should be encouraged to share the same identity. We set $\tau_{\text{max}} = 120$ which allows to investigate nodes that are upto 6 seconds (at the frame rate of 20 fps) far apart, and the distance is computed as:

$$d_{uv} = \left\| \mathbf{x}_v^{(s)} - \mathbf{x}_u^{(e)} \right\|_2 \tag{8}$$

We use $\tau_{\text{dist}} = 450$.

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 as it does not help to disambiguate the possible labels of the nodes.

3.3 Priority scheduling of belief message exchanges

The standard belief propagation (BP) technique, described in Section 3.1, does not prioritize nodes, which means that the nodes are selected in an arbitrary order to send messages to their neighbors. Moreover, all the nodes transmit messages to their neighbors.

However, in our graph, some nodes are less ambiguous about their identities than others. Such non-uniform distribution of ambiguity over the nodes should allow faster convergence of the BP because the messages sent by less ambiguous nodes are more informative. Consequently, they can help the more ambiguous neighbors to disambiguate their labels. Therefore, it sounds natural to prioritize the nodes such that less ambiguous nodes transmit their messages first. Interestingly, the authors in [9] have followed the same intuition to solve an image completion problem. We have adapted the principle of belief propagation prioritization to our identity assignment context.

As already mentioned, priority is related to the ambiguity of the node identity. The definition of ambiguity (and hence the priority) of a node v depends on the peakedness of the current belief vector $\mathbf{b}_v$ that has been estimated by the BP algorithm. We use *entropy* of the belief vector, defined as $E(v) = -\sum_{l_v \in \mathcal{L}} b_v(l_v)\log(b_v(l_v))$, to measure the ambiguity of node v. The entropy is maximum when the belief vector has a flat distribution and decreases with the peakedness of the distribution. Therefore, the node v is assigned a higher priority if it has lower entropy and vice versa. We have tested other approaches for the priority (*e.g.*, cardinality of the confusion set [9], kurtosis, etc.). However, the results were not better than the ones obtained with the entropy measure.

Apart from priority-based scheduling, to avoid propagating confusing information, the construction of exchanged messages is also affected by the ambiguity of the information available in each node. Specifically, during the construction of message $\mathbf{m}_{u\to q}$, the node u is supposed to gather messages from all its neighbors, except q, to construct $\mathbf{h}_u$. However, since there can be some nodes in the neighborhood of u which are more ambiguous than u, the messages coming from those nodes could be considered as being uninformative, or even confusing, since they are nearly flat, meaning that all labels are equally likely. Thus, we only consider those nodes that are less ambiguous than u to compute $\mathbf{h}_u$. The practical implementation of this principle works as follows. Each node is assigned a *committed* flag, initialized to *false* in the beginning of each iteration of the message exchange process. Once a node is scheduled to exchange messages with its neighbors, its flag is set to *true*. In this way, the less ambiguous neighbors of u have their committed flags set to true as they have been scheduled before u. Similarly, the more ambiguous nodes will have their committed flags set to false. Following the same kind of idea, the message $\mathbf{m}_{u\to q}$ is constructed and sent only if q is more ambiguous than u, *i.e*, if the committed flag of q is false. This is illustrated in Figure 2.

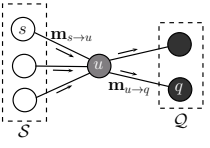

Figure 2: Message construction and dissemination at node u (in **gray**). The node u gathers information from its less ambiguous neighbors, $\mathcal{S}$ (in **white**). Afterwards, u transmits message to its more ambiguous neighbors, $\mathcal{Q}$ (in **black**).

The algorithm for the priority based belief propagation is presented in Algorithm 1. After T iterations, we assign the a label l_v to a node v, as

$$l_v = \begin{cases} l_x^* & \text{if } b_v^{(T)}(l_x^*) > \kappa b_v^{(T)}(l_y^*), \\ \text{undefined} & \text{otherwise.} \end{cases} \tag{9}$$

where $l_x^* = \arg\max_{l_x \in \mathcal{L}} b_v^{(T)}(l_x)$ and $l_y^* = \arg\max_{l_y \in \mathcal{L}\setminus l_x^*} b_v^{(T)}(l_y)$. We use $\kappa = 4$.

4 Experimental validation

We apply the proposed method on the output of the detector and tracklet aggregation process, as described in [7]. We run the algorithm on 10 minutes of a real-life basketball game captured by 7 loosely synchronized cameras, distributed around a basketball court [1]. The distribution of the cameras is not symmetrical, meaning that there are more cameras on one side than the other. Hence, the observation of appearance features is more reliable on one side of the court than on the other side. The identity and the position of the players have

Algorithm 1 Priority Belief Propagation

Initialize: $b_v(l_v) \leftarrow p_v(l_v)$ and $\phi_v(l_v) \leftarrow p_v(l_v) \qquad \forall v \in \mathcal{V}, l_v \in \mathcal{L}$
for $t = 1$ to T **do**
  v.committed $\leftarrow$ false $\forall v \in \mathcal{V}$
  $\mathcal{R} \leftarrow \mathcal{V}$
  while $\mathcal{R} \neq \emptyset$ **do**
    $u \leftarrow$**Schedule**$(\mathcal{R})$ {/* Prioritize the node u according to its level of ambiguity */}
    u.committed $\leftarrow$ true
    $\mathcal{S} \leftarrow \{s | s \in \mathcal{N}_u, s.\text{committed=true}\}$ {/* Less ambiguous neighbors of u */}
    $\mathbf{h}_u \leftarrow$ Compute the pre-message of u from $\mathcal{S}$
    $\mathcal{Q} \leftarrow \{v | v \in \mathcal{N}_u, v.\text{committed=false}\}$ {/* More ambiguous neighbors of u */}
    for $v \in \mathcal{Q}$ **do**
      $\mathbf{m}^{(t)}_{u \rightarrow v} \leftarrow$ **ComputeMessage**$(u, v, \mathbf{h}_u, \tau_{\max})$
      $\mathbf{b}_v \leftarrow$ Compute belief for node v
    end for
    $\mathcal{R} \leftarrow \mathcal{R} \setminus u$
  end while
end for
Assign identity to each node $v \in \mathcal{V}$ according to the Equation 9.

been manually defined at every second. This provides the reference ground truth used in our evaluation.

In the rest of the section, we first describe the metrics used to evaluate the proposed algorithm. Afterwards, the results for different configurations of the graph and different belief propagation algorithms are presented.

4.1 Performance metrics

Given the ground truth trajectory and its identity, we follow [4] and define the performance metrics. Let $G = \{g_j | j = 1, ..., N\}$ be the set of ground truth objects at time $t \in [0, T_{\text{obs}}]$, where T_{obs} is the entire observation interval. Each object g_j has an identity l_j and a location $\mathbf{x}_j$, i.e., $g_j = (l_j, \mathbf{x}_j)$. Similarly, let $H = \{h_i | i = 1, ..., |\mathcal{V}|\}$ be outputs of the system at time t. Each hypothesis h_i has a location $\mathbf{y}_i$ and an identity estimate $\hat{l}_i$, i.e., $h_i = (\hat{l}_i, \mathbf{y}_i)$. Thus, we have,

At each time t, we define following error metrics:

- **Missed detection** corresponds to a g_j for which no hypothesis is detected.
- **Correct detection but wrong identification** is a h_i for which $||\mathbf{x}_j - \mathbf{y_i}||_2 < \tau_{\text{dist}}$ but $l_j \neq \hat{l}_i$. We use τ_{dist} = 30 cm.
- **False positive** corresponds to a h_i for which there is no g_j such that $l_j = \hat{l}_i$.

Let, wi_t, fp_t, ms_t and gt_t be the number of wrong identifications, false positives, misses and ground truth objects at time t respectively. Then, we define the metrics as:

$$\text{FP} = \frac{\sum_\text{t} \text{fp}_\text{t}}{\sum_\text{t} \text{gt}_\text{t}}, \quad \text{WI} = \frac{\sum_\text{t} \text{wi}_\text{t}}{\sum_\text{t} \text{gt}_\text{t}}, \quad \text{MS} = \frac{\sum_\text{t} \text{ms}_\text{t}}{\sum_\text{t} \text{gt}_\text{t}}, \quad \text{Accuracy} = 1 - \text{FP} - \text{MS} - \text{WI} \tag{10}$$

4.2 Results

In order to elucidate the effect of message passing in the performance of the system, we explored the following variations of the identity assignment algorithm: (a) **No BP** (no message

passing is done, equivalent to treating each node independently); (b) **Standard BP** (nodes are chosen either sequentially or in an arbitrary order); and (c) **Priority BP** (nodes are prioritized in increasing order of ambiguities). Table 1 presents the metrics for the above message passing techniques.

	Accuracy(%)	FP(%)	WI(%)	MS(%)
No BP	68.14	0.13	1.48	30.25
Std. BP (sequential)	73.59	0.15	1.76	24.50
Std. BP (random)	83.69	0.16	2.36	13.79
Priority BP	**89.04**	**0.15**	**2.54**	**8.27**

Table 1: Comparison of performance metrics for different node scheduling approaches.

From the Table 1, we can observe that the message passing between the nodes indeed boosts the accuracy of the system from 68.14% to 89.04% at the cost of a slight increase in the wrong identification error from 1.48% to 2.54%. Interestingly, we can observe that the results are better for random access of the nodes as compared to the sequential access. More importantly, the scheduling of the nodes improves the performance drastically as compared to the arbitrary access of the nodes.

In Figure 3, we can observe how the average entropy of the system evolves at each iteration. We can see that the priority BP not only attains the lowest entropy but also converges rapidly. It shows that that the order in which the nodes transmit message indeed affects the convergence of the system, thereby, affecting the quality of the solution.

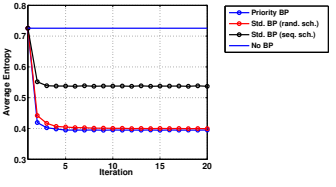

Figure 3: Evolution of average entropy for different node scheduling mechanisms.

In addition, we explored how the graphical model affects the performance of the priority BP. For this purpose, we have run priority BP on other three different variations of the full graph. They are defined as: (a) **No edges** (nodes are not connected by edges, and are hence independent); (b) **Mutex edges only** (with only the mutex edges between the nodes); (c) **Temporal edges only** (with only the temporal edges between the nodes).

	Accuracy(%)	FP(%)	WI(%)	MS(%)
No edges	68.14	0.13	1.48	30.25
Mutex edges	80.32	0.15	2.03	17.50
Temporal edges	84.49	0.15	2.71	12.65
Both edges	**89.04**	**0.15**	**2.54**	**8.27**

Table 2: Performance metrics for priority-based BP on four graphical models.

The performances of all models are shown in Table 2. We can draw two main conclusions. First, adding edges to the baseline system allows exploiting the correlation between the nodes. Second, both the temporal edges and the mutex edges help in improving the performances.

5 Conclusion

In this paper, we presented an approach to solve identity (or, label) assignment problem in a scenario for which target candidates have been detected and observed independently with various degree of reliability in the belief propagation framework. Messages are transmitted

between the detections to infer the identity of ambiguous observations, based on the reliable identities available from non-ambiguous appearance features. We show that the order in which messages are exchanged between the nodes affects the quality of the solution. We have proposed a priority-based node scheduling mechanisms to favor the transmission of information from less ambiguous to more ambiguous nodes. The above approach has been applied on a real-life basketball game to recognize the players. The correct recognition rate of 89% demonstrates the effectiveness and efficiency of the proposed approach.

Acknowledgements

The authors thank Cédric Verleysen for his help. Amit Kumar K.C. and Christophe De Vleeschouwer are supported by Belgian National Science Foundation (F.N.R.S.).

References

[1] http://www.apidis.org/dataset/.

[2] Authors. Aggregation of local shortest paths for multiple object tracking with noisy/missing appearance features. In *ECCV 2012 Submission ID 1342*.

[3] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua. Multiple object tracking using k-shortest paths optimization. *PAMI*, 33:1806–1819, September 2011. ISSN 0162-8828. doi: http://dx.doi.org/10.1109/TPAMI.2011.21. URL http://dx.doi.org/10.1109/TPAMI.2011.21.

[4] Keni Bernardin. *Multimodal probabilistic person tracking and identification in smart spaces*. PhD thesis, Universität Fridericiana zu Karlsruhe (TH), November 2009.

[5] Fan Chen, Damien Delannay, and Christophe De Vleeschouwer. An autonomous framework to produce and distribute personalized team-sport video summaries: a basket-ball case study. *IEEE Transactions on Multimedia*, pages 1381–1394, December 2011.

[6] D. Delannay, N. Danhier, and C. De Vleeschouwer. Detection and recognition of sports(wo)men from multiple views. In *ICDSC, Como, Italy*, 2009.

[7] Pedro Felzenszwalb and Daniel Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70:41–54, 2006. ISSN 0920-5691. URL http://dx.doi.org/10.1007/s11263-006-7899-4. 10.1007/s11263-006-7899-4.

[8] H. Jiang, S. Fels, and J. Little. A linear programming approach for multiple object tracking. In *CVPR*, 2007.

[9] 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.

[10] 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.

[11] Kevin P. Murphy, Yair Weiss, and Michael I. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *In Proceedings of Uncertainty in AI*, pages 467–475, 1999.

[12] Peter Nillius, Josephine Sullivan, and Stefan Carlsson. Multi-target tracking - linking identities using bayesian network inference. In *CVPR*, 2006.

[13] M. Ocakli and M. Dermirekler. Video tracker system for traffic monitoring and analysis. In *IEEE Signal Processing and Communication Applications*, 2007.

[14] C. Piciarelli, C. Micheloni, and G.L. Foresti. Trajectory based anomalous event detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2008.

[15] H. Pirisiavash, D. Ramanan, and C. C. Fowlkes. Globally-optimum greedy algorithms for tracking a variable number of objects. In *CVPR*, 2011.

[16] Cédric Verleysen and Christophe De Vleeschouwer. Recognition of sport players' numbers using fast color segmentation. In *Proc. of the SPIE-IS&T Electronic Imaging*, 2012.