A Comparative View on Exemplar ’Tracking-by-Detection’ Approaches
Elie Moussy, Alhayat Ali Mekonnen, Guilhem Marion, Frédéric Lerasle

To cite this version:

HAL Id: hal-01202882
https://hal.archives-ouvertes.fr/hal-01202882
Submitted on 23 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A Comparative View on Exemplar ‘Tracking-by-Detection’ Approaches

E. Moussy\textsuperscript{1}, A. A. Mekonnen\textsuperscript{1}, G. Marion\textsuperscript{1,2}, F. Lerasle\textsuperscript{1,2}\\
\textsuperscript{1}CNRS, LAAS, 7 avenue du Colonel Roche, F-31400 Toulouse, France\\
\textsuperscript{2}Univ de Toulouse, UPS, LAAS, F-31400 Toulouse, France\\
\{emoussy,aamekonn,marion,lerasle\}@laas.fr

Abstract

In this work, we present a comparative evaluation of various ‘tracking-by-detection’ approaches on public datasets. The work investigates popular sequential Monte Carlo and template ensemble based trackers coupled with relevant visual people detectors with emphasis on exhibited performance variation depending on tracker-detector choice. Extensive experimental results are provided on public dataset and results indicate the choice of a detector can significantly vary the performance of a tracker. Our experimental results show, depending on the choice of the detector, average tracking accuracy across three public datasets could exhibit a 45\% standard deviation with only, on average, a 6.8\% and 11.1\% standard deviation in detector recall and precision respectively.

1. Introduction

People detection and tracking is an important research area with prominent applications in video surveillance, pedestrian protection systems, human-computer interaction, robotics, and the like. As a result, it has amassed huge interest from the scientific community [2, 5, 4]. People tracking falls under Multi-Object Tracking (MOT) which deals with the process of accurately estimating the state of objects – position, identity, and configuration – over time from observations. Due to incurred challenges – scene clutter, target dynamics, intra/inter-class variation, measurement noise, frame rate – it has long been established that coupling trackers with detectors, in a paradigm called ‘tracking-by-detection’, helps better tackle these challenges [2, 9, 8]. In the context of people tracking, ‘tracking-by-detection’ approaches rely on a people detector to start, update, re-initialize, guide (avoid drift), or terminate a tracker.

In the literature, it is common to find many ‘tracking-by-detection’ approaches applied to people tracking. However the usual trend it to select a single detector and directly couple it with the tracker, e.g., [2, 9]. With the advent of various different people detection techniques with significant variations in terms of detection performance and speed, see [5], the first step should be to evaluate the effect a detector choice impacts performance, and the relevant association with filtering strategies. To the best of our knowledge no such work exits till date. To clarify, there are indeed very good experimental comparative works in detection, e.g., [5], but none that shows the inter related effects by detector and tracker choices. To bridge this gap, in this work, we present a comparative evaluation of exemplar ‘tracking-by-detection’ approaches, with different detectors, on relevant public datasets, primarily focusing on sequential Monte Carlo approach as most approaches rely on it. We consider three different trackers (filtering strategies) owing to their pervasive use in the literature and relevance: A Decentralized Particle Filter (DPF), e.g., [2, 7], Tracker Hierarchy [12], and Reversible Jump Markov Chain Monte Carlo - Particle Filter (RJMC) [8]. The DPF and RJMCMC are selected as they are the most popular Monte Carlo approaches marking two important tracker configurations: a decentralized one which assigns an independent tracker per target, and centralized one – also called a joint state tracker – in which all the states of the tracked targets are concatenated forming a single representation that captures the entire configuration. The Tracker Hierarchy is selected as it, unlike the usual implementations of DPF and RJMCMC which utilize simple single target representation, consists of a rich target representation model in the form of template ensembles. Similar to DPF, it is a decentralized approach and has shown tracking results comparable to the state-of-the-art [12].

Channel Features (ACF) based detector. The detectors mark three distinct detector ‘superiority era’ onsets as published in 2005, 2010, and 2014 consecutively. Furthermore, our choice is motivated by the fact that both ACF and DPM are amongst the current best detectors and HOG-SVM, though not currently the best itself, its features make constituents, one way or another, of current state-of-art approaches and historically has been the de facto benchmark detector.

In short, the selected trackers and detectors are quite relevant in the literature and the variations amongst them, namely: (1) tracker configuration that governs how the state space is explored, (2) variation in target representation – simple color appearance to template ensemble, (3) the detectors variation in terms of detection accuracy and precision [5], dictates the representativeness and relevance of this comparative evaluation. Hence, the contributions of this paper are: (1) systematic evaluation of ‘tracking-by-detection’ approach based on relevant combination of trackers and detectors, (2) a presentation of relevant results with insightful discussions that highlight the benefits of tracker choice, target representation, and detector choice.

This paper is organized as follows: section 2 starts with background and overview of the different trackers and detectors combined, section 3 details the experimental setting and obtained results, section 4 presents a comprehensive discussion, and the paper finally finishes off with conclusions and future works in section 5.

2. Background and overview

In this section, we will provide an overview and relevant background information about the two main ingredients of ‘people tracking-by-detection’: people detection and tracking. The different detectors and trackers selected for the comparative evaluation are briefly described.

2.1. People Detection

Depending on the type of feature, model, classifier, and learning technique adopted, visual people detectors perform variably on different datasets. Below, the three chosen detectors are briefly explained.

**HOG-SVM** [3]: This detector is one of the classical and oldest detectors. This detector computes local histograms of the gradient orientation on a dense grid and uses linear Support Vector Machine (SVM) as a classifier. The constituent HOG features have shown to be the most discriminant features to date, and in fact, a majority of detectors proposed hence-after make use of HOG or its variant one way or another [5].

**DPM** [6]: Contrary to HOG and ACF which detect a person’s full body, the DPM is a parts based detector that works by aggregating evidence of different parts of a body to detect a person in an image. Its trained model is divided in different parts. For instance, a person’s model could be made up of a head, upper body, and lower body sub-models. Each detected area has its own score. Thus, it is possible to put a threshold in order to remove detections that have low scores. Since this detector relies on parts, and not solely on a full body, it detects partially occluded people rather well, see figure 1b for example. Additionally, it also has better localization accuracy as it infers bounding box based on detected body parts. The detector uses variants of HOG features with Latent-SVM as classifier.

**ACF** [4]: This is a fast person detector that has shown state-of-the-art performance on various benchmarking datasets [4]. It is based on aggregates (summed over blocks) of features represented as channels and a variant of Boosted classifier. Examples of features channels used include: normalized amplitude of the gradient, the histograms of oriented gradients (HOG, 6 channels) and color channels (LUV).

Generally speaking, the ACF, amongst the others, does significantly better in outdoor environments, e.g., camera mounted on a vehicle, whereas the DPM outperforms whenever the dataset contains partially occluded person instances. Surprising, the HOG-SVM detector also does well with the presence of contrasting background that clearly helps delimit the boundaries of people. As a result, it is the interest in this work to show how these detector performance variations affect ‘tracking-by-detection’ based trackers, notably their filtering and target appearance model robustness, on different benchmarking public datasets. Sample detections are shown in figure 1.

2.2. Multi-Object Tracking (MOT)

For MOT, we consider the two popular tracker configurations: A decentralized and centralized tracker configuration.

2.2.1 Decentralized MOT

In decentralized MOT, each instantiated tracker has its own state vector which is independent of the others. In this class, we chose to use our variant of the classical Particle Filter (PF) – a popular choice in the literature, e.g., [2, 7] – and the more involved open-sourced Tracker Hierarchy [12] which is a template ensemble based tracker.

**Decentralized Particle Filters (DPF)**: In this approach, each target is assigned a unique instance of a Particle Filter as a tracker. In this work, this target specific tracker is
implemented based on the ICondensation [10] filter as it is the most widely used and suitable PF variant for detector integration. This is a sequential Monte Carlo approach which approximates the posterior over the target state $x_t$ given all measurements up to time $t$, $Z_{1:t}$, using a set of $N$ weighted samples, i.e., $p(x_t | Z_{1:t}) \approx \{x_t^{(i)}, w_t^{(i)}\}_{i=1}^N$. Tracking is achieved sequentially with the notion of Importance Sampling whereby the particles at time $t-1$ are propagated according to a proposal density, $q(\cdot)$, and their weights are updated in accordance with equation 1.

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(z_t | x_t^{(i)}) p(x_t^{(i)} | p(x_{t-1}^{(i)})}{q(x_t^{(i)} | x_{t-1}^{(i)}, z_t)} (1)$$

Where, $p(x_t^{(i)} | x_{t-1}^{(i)})$ is target dynamic model, $p(z_t | x_t^{(i)})$ is likelihood term, and $q(x_t^{(i)} | x_{t-1}^{(i)}, z_t)$ is the proposal density evaluated at the sampled state. To derive this filter with incoming detections, hence to realize ‘tracking-by-detection’, the proposal density shown in equation 2 is employed. According to this density, part of the particles will be sampled from the detector cues, $\pi(x_t^{(i)} | z_t)$, some from the dynamics, and some from the prior $p_0(x_t^{(i)})$, in accordance with the ratios $\alpha, \beta, \gamma$ which should sum to 1.

$$q(x_t^{(i)} | x_{t-1}^{(i)}, z_t) = \beta p(x_t^{(i)} | x_{t-1}^{(i)}) + \alpha \pi(x_t^{(i)} | z_t) + \gamma p_0(x_t^{(i)}) (2)$$

The likelihood $p(z_t | x_t^{(i)})$ is a probabilistic measure based on the Bhattacharyya similarity coefficient, as in [10], with respect to a dynamically updated simple single target color histogram, constructed in the HSV color space. During MOT, each incoming detections have to be associated with the different trackers distinctly. For that we use a greedy assignment algorithm [2] which performs comparatively to the famous Hungarian assignment algorithm. The DPF is coupled with each one of the detectors presented in section 2.1 during experimental evaluations. The exact label for the tracker is derived by appending the detector name on DPF. For example, DPF-ACF is the decentralized Particle Filter multi-object tracker using ACF as detector.

**Tracker Hierarchy (Hierarchy) [12]:** This multi-object tracker is another ‘tracking-by-detection’ decentralized MOT that assigns a single tracker per target. It is a tracker that consists of a rich appearance model of the target in the form of a template ensemble and uses hierarchy of expert and novice trackers for efficient multi-person tracking. At each time step, the correct target position is estimated by making use of a mean-shift mode estimator and a Kalman filter. We consider evaluating its performance with different detectors as it, unlike the two other trackers, consists of a rich and involved target appearance model. Please refer to [12] for further details. This tracker combined with any of the detectors is labelled as Hierarchy followed by detector name, e.g., Hierarchy-ACF.

### 2.2.2 Centralized MOT

The centralized MOT, also called joint state tracker, represents the tracked target using a joint state that captures the entire configuration. The main advantage is that should the targets interact, an interaction model can be incorporated in the tracking step. The most famous approach in this class is the Reversible Jump Markov Chain Monte Carlo - Particle Filter (RJMCMC) tracker [8]. The RJMCMC defines a Markov Chain over the state configuration so that the stationary distribution of the chain approximates the posterior distribution $p(x_t | Z_{1:t})$. It replaces the inefficient important sampling step with a trans-dimensional Metropolis Hastings (MH) algorithm to sample from the chain.

In RJMCMC tracker, the posterior at time $t$ is approximated using a set of $M$ discrete unweighted samples: $p(x_t | Z_{1:t}) \approx \{x_t^{(i)}\}_{i=1}^M$. It uses a set of moves $m$ to change the dimension of the state, i.e., adding new target, removing untracked targets, or leave it unchanged according to a prior move proposal $q_m$. Each move is associated with a move specific proposal distribution $Q_m(\cdot)$. Each move $m$ must have a reverse move $m^*$ that assures reversibility so that detailed balance will be achieved and the chain will converge to the desired stationary distribution [8]. During the iterative estimation process, at the $i^{th}$ iteration, it first samples a move from $q_m$, and proposes a new particle $x^*$ based on $Q_m(\cdot)$. It then computes the acceptance ratio $\alpha_a$ shown in equation 3 considering $Q_m(\cdot)$ and the reverse move proposal distribution $Q_{m^*}(\cdot)$. The proposed particle is accepted with probability $\alpha_a$ or otherwise rejected. Particles used both for the burn-in, $M_b$, and thin-out, $M_t$, are discarded leaving $M$ unweighted samples to represent the posterior.

$$\alpha_a = \min \left(1, \frac{p(x^* | Z_{1:t}) Q_{m^*}(x_{t-1}^{(i)}; x^*) q_m(\Psi(x^*))}{p(x_{t-1}^{(i)} | Z_{1:t}) Q_m(x^*; x_{t-1}^{(i)}) q_m(\Psi(x_{t-1}^{(i)}))} \right) (3)$$

$\Psi(\cdot)$ in equation 3 is the interaction model. Our implementation is based on [8] with the set of moves $m = \{add, delete, stay, leave, update, swap\}$, a markov random field based interaction model $\Psi(\cdot)$, a likelihood measure $p(z_t | x_t)$ based on the Bhattacharyya similarity coefficient of a simple single target color histogram, contrary to Hierarchy which has an ensemble, in the HSV color space. The different move specific proposal distributions, $Q_m(\cdot)$, are defined as in [8]. The RJMCMC is coupled with the different detectors presented and evaluated. The coupled tracker-detector is denoted using RJMCMC followed by used detector acronym, e.g., RJMCMC-ACF. Similar to DPF, the detection-track data association is handled via a greedy assignment algorithm.
3. Experiments and results

As stated, the main objective of this paper is to provide a comparative evaluation of three multi-object ‘tracking-by-detection’ based tracking techniques utilizing three different detectors. The tracker and detector combination leads to nine evaluated combinations. In this section, we present the different datasets, evaluation metrics, implementation details, and obtained results.

3.1. Datasets

To evaluate the different detector-tracker combination, we have utilized three public datasets, namely: The CAVIAR OneShopOneWait dataset1, the CAVIAR EnterExitCrossingPaths dataset1, and the PETS2009S2L12 dataset. Here onwards, these datasets are referred as OneShop, EnterExit, and PETSS2L1 respectively. These datasets are selected as they highlight different conditions: (1) the OneShop sequence consists of intermittent target occlusions, (2) the EnterExit sequence is challenging for detectors due to background clutter, and (3) the PETSS2L1 features an outdoor scene with many targets and encountered target-target interactions.

![Figure 2](image_url) Figure 2: (a) Detector Precision-Recall curve on a subset of the PETSS2L1 dataset. (b) Performance, in terms of MOTA, of the Monte Carlo trackers as a function of number of particles used to approximate the posterior.

### Table 1: Utilized public datasets along with detector performance.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Frame Rate</th>
<th>No. Frames</th>
<th>No. of Id.</th>
<th>Detector</th>
<th>Recall / Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>EnterExit</td>
<td>25</td>
<td>383</td>
<td>4</td>
<td>TP</td>
<td>58 / 76</td>
</tr>
<tr>
<td>OneShop</td>
<td>25</td>
<td>1377</td>
<td>6</td>
<td>TP</td>
<td>32 / 48</td>
</tr>
<tr>
<td>PETSS2L1</td>
<td>7</td>
<td>795</td>
<td>19</td>
<td>TP</td>
<td>68 / 92</td>
</tr>
</tbody>
</table>

Relevant descriptions of the different datasets along with the performance of the three detectors is provided in table 1. Relevant descriptions of the different datasets along with the performance of the three detectors is provided in table 1.

3.2. Evaluation metrics

To quantify the performance of the different trackers, we utilize the prevalent CLEAR-MOT metrics [1]. The CLEAR-MOT metrics are basically based on computation of two quantities: the Multi-Object Tracking Accuracy (MOTA) and the Multi-Object Tracking Precision (MOTP). The MOTA = 1−(FP + FN + l_{sw}), where FP = ∑t E_{P_{t}} quantifies total false positives, FN = ∑t E_{N_{t}} quantifies the total false negatives, and \( l_{sw} = \sum_{t} l_{sw_{t}} \) quantifies total id switches, all divided by the ground truth targets and summed over the entire dataset. The MOTP is the average bounding box overlap between the estimated target position and ground truth annotations over the correctly tracked targets. A tracker estimated rectangular po-

---

1http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/
2http://www.cvg.reading.ac.uk/PETS2009a.html
3http://opencv.org/
3.4. Results

The results obtained by running the tracker-detector combinations on the three public datasets are shown in detail in tables 2, 3, and 4, corresponding to the EnterExit, OneShop, and PETSS2L1 datasets respectively. Each tracker-detector combination is run five times to account for their stochastic nature. In all cases, the results are reported as mean/standard deviation, i.e., $\mu/\sigma$. Before analyzing these results, it is worth taking a look at the detector performance on these datasets in table 1. All the three detectors achieve high detection rates, $>80\%$, with high precision, $>90\%$, on the PETSS1L dataset with top performance from ACF detecting 88\% percent of the targets correctly. On the contrary, ACF does poorly on the other two datasets, EnterExit and OneShop, with DPM achieving the best result trailed by HOG-SVM. On average, DPM perform much better than the others with consistent high precision rates. The results also indicate that the OneShop dataset poses a great challenge for detectors followed by EnterExit.

In this dataset, the DPF is best combined with ACF, the Hierarchy with DPM, and the RJMCMC also with DPM. The best result, a 58\% MOTA, is achieved with the DPF-DPM tracker. Besides, when the detector is not reliable such as the ACF in Table 3, the DPF and the RJMCMC handle it better than the Hierarchy, i.e., have better filtering capabilities.

### Table 4: CLEAR-MOT results on the PETSS2L1 dataset. All results reported as $\mu/\sigma$ computed over five runs.

<table>
<thead>
<tr>
<th>Tracker-Detector</th>
<th>MOTA</th>
<th>MOTP</th>
<th>$T_P$</th>
<th>$F_P$</th>
<th>$F_N$</th>
<th>$I_{d_{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPF-ACF</td>
<td>.54/ .01</td>
<td>.53/ .00</td>
<td>.87/ .01</td>
<td>.34/ .11</td>
<td>.09/ .01</td>
<td>76.3/ 7.8</td>
</tr>
<tr>
<td>Hierarchy-ACF</td>
<td>.88/ .00</td>
<td>.68/ .00</td>
<td>.93/ .00</td>
<td>.05/ .00</td>
<td>.05/ .00</td>
<td>79.1/ 0.0</td>
</tr>
<tr>
<td>RJMCMC-ACF</td>
<td>.73/ .03</td>
<td>.65/ .00</td>
<td>.83/ .02</td>
<td>.10/ .02</td>
<td>.16/ .02</td>
<td>59.6/ 19.6</td>
</tr>
<tr>
<td>DPF-DPM</td>
<td>.68/ .01</td>
<td>.53/ .01</td>
<td>.77/ .01</td>
<td>.09/ .01</td>
<td>.21/ .01</td>
<td>55.7/ 12.0</td>
</tr>
<tr>
<td>Hierarchy-DPM</td>
<td>.86/ .00</td>
<td>.70/ .00</td>
<td>.90/ .00</td>
<td>.04/ .00</td>
<td>.09/ .00</td>
<td>55.8/ 0.0</td>
</tr>
<tr>
<td>RJMCMC-DPM</td>
<td>.61/ .01</td>
<td>.55/ .00</td>
<td>.62/ .01</td>
<td>.01/ .01</td>
<td>.37/ .01</td>
<td>49.8/ 8.8</td>
</tr>
<tr>
<td>DPF-HOG</td>
<td>.32/ .01</td>
<td>.30/ .00</td>
<td>.81/ .04</td>
<td>.25/ .01</td>
<td>.17/ .03</td>
<td>42.7/ 2.6</td>
</tr>
<tr>
<td>Hierarchy-HOG</td>
<td>.88/ .00</td>
<td>.65/ .00</td>
<td>.92/ .00</td>
<td>.04/ .00</td>
<td>.08/ .00</td>
<td>37.2/ 0.0</td>
</tr>
<tr>
<td>RJMCMC-HOG</td>
<td>.54/ .03</td>
<td>.46/ .01</td>
<td>.61/ .01</td>
<td>.07/ .02</td>
<td>.31/ .01</td>
<td>56.8/ 6.1</td>
</tr>
</tbody>
</table>

Table 3: CLEAR-MOT results on the OneShop dataset. All results reported as $\mu/\sigma$ computed over five runs.

<table>
<thead>
<tr>
<th>Tracker-Detector</th>
<th>MOTA</th>
<th>MOTP</th>
<th>$T_P$</th>
<th>$F_P$</th>
<th>$F_N$</th>
<th>$I_{d_{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPF-ACF</td>
<td>.53/ .01</td>
<td>.42/ .04</td>
<td>.73/ .09</td>
<td>.52/ .02</td>
<td>.21/ .07</td>
<td>19.4/ 10.2</td>
</tr>
<tr>
<td>Hierarchy-ACF</td>
<td>.31/ .00</td>
<td>.64/ .00</td>
<td>.68/ .00</td>
<td>.38/ .00</td>
<td>.31/ .00</td>
<td>6.9/ 0.0</td>
</tr>
<tr>
<td>RJMCMC-ACF</td>
<td>.30/ .02</td>
<td>.64/ .02</td>
<td>.55/ .02</td>
<td>.24/ .01</td>
<td>.43/ .02</td>
<td>32.8/ 6.7</td>
</tr>
<tr>
<td>DPF-DPM</td>
<td>.68/ .03</td>
<td>.59/ .04</td>
<td>.79/ .07</td>
<td>.12/ .07</td>
<td>.19/ .07</td>
<td>15.7/ 4.0</td>
</tr>
<tr>
<td>Hierarchy-DPM</td>
<td>.75/ .00</td>
<td>.76/ .00</td>
<td>.82/ .00</td>
<td>.08/ .00</td>
<td>.17/ .00</td>
<td>42.0/ 0.0</td>
</tr>
<tr>
<td>RJMCMC-DPM</td>
<td>.53/ .02</td>
<td>.62/ .01</td>
<td>.54/ .03</td>
<td>.01/ .01</td>
<td>.45/ .03</td>
<td>9.0/ 2.7</td>
</tr>
<tr>
<td>DPF-HOG</td>
<td>.23/ .01</td>
<td>.62/ .01</td>
<td>.66/ .04</td>
<td>.37/ .05</td>
<td>.41/ .02</td>
<td>19.1/ 1.2</td>
</tr>
<tr>
<td>Hierarchy-HOG</td>
<td>.24/ .00</td>
<td>.22/ .00</td>
<td>.36/ .00</td>
<td>.12/ .00</td>
<td>.63/ .00</td>
<td>18.0/ 0.0</td>
</tr>
<tr>
<td>RJMCMC-HOG</td>
<td>.50/ .04</td>
<td>.41/ .00</td>
<td>.58/ .05</td>
<td>.08/ .04</td>
<td>.40/ .05</td>
<td>6.9/ 5.7</td>
</tr>
</tbody>
</table>

Table 2: CLEAR-MOT results on the EnterExit dataset. All results reported as $\mu/\sigma$ computed over five runs.

Table 2 shows the detailed results on the EnterExit dataset. The best overall result, both accuracy and precision, is obtained with the Hierarchy-DPM tracker. The best results of each of the trackers is obtained when combined with the DPM detector. As the ACF performs poorly in this dataset, corresponding results are accordingly poor. In this dataset, the DPF is best combined with ACF, the Hierarchy with DPM, and the RJMCMC also with DPM. The HOG-SVM detector is better combined with the RJMCMC tracker.

For better analysis, we report two summarized results in tables 5 and 6. Table 5 reports the $\mu/\sigma$ of the MOTA obtained by the tracker-detector combination across the three datasets and the overall average MOTA, across dataset and detector. Clearly, the best result is obtained using the Hierarchy based trackers. When combined with DPM, Hierarchy-DPM, this leads to an average 67\% MOTA across all the datasets. This result is trailed by the RJMCMC, and then the DPF. Similarly, table 6 details the average MOTA and Id switch counts of the different detectors (averaged over detector combinations) on the three datasets. Results are reported as $\mu/\sigma$.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>MOTA</th>
<th>MOTP</th>
<th>$T_P$</th>
<th>$F_P$</th>
<th>$F_N$</th>
<th>$I_{d_{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EnterExit</td>
<td>.87/ .01</td>
<td>.59/ .07</td>
<td>.63/ .10</td>
<td>.57/ .41</td>
<td>.60/ .10</td>
<td>48.1/31.6</td>
</tr>
<tr>
<td>OneShop</td>
<td>.68/ .00</td>
<td>.51/ .00</td>
<td>.55/ .00</td>
<td>.19/ .00</td>
<td>.38/ .00</td>
<td>184.3/ 0.0</td>
</tr>
<tr>
<td>PETSS2L1</td>
<td>.61/ .00</td>
<td>.42/ .02</td>
<td>.31/ .02</td>
<td>.15/ .03</td>
<td>.09/ .02</td>
<td>20.8/ 5.5</td>
</tr>
</tbody>
</table>

Table 3: CLEAR-MOT results on the OneShop dataset. This is the dataset where the detectors all perform poorly. This is clearly reflected in the results. Again the best results, precision and accuracy, involve the DPM detector: accuracy with DPF-DPM and precision with Hierarchy-DPM. The best result, a 58\% MOTA, is achieved with the DPF-DPM tracker. Besides, when the detector is not reliable such as the ACF in Table 3, the DPF and the RJMCMC handle it better than the Hierarchy, i.e., have better filtering capabilities.
PETSS2L1, the Hierarchy based tracker does better. In terms of Id switch, the DPF does better on the OneShop and PETSS2L1 datasets, whereas the Hierarchy on average does better on the EnterExit dataset.

4. Discussions

As we can see in Tables 2, 3 and 4, the Hierarchy tracker has the best results in two out of three datasets. Its performance varies with the associated detector. Indeed, when a detector has a high precision/recall ratio (in a sequence) compared to the others, the Hierarchy tracker will get a better accuracy when combined with it than with the others. Therefore, even with a complex target representation model, the Hierarchy tracker is only showing improvements when combined with a reliable detector. Meanwhile, if the performance of two detectors are comparable, the accuracy of this tracker will not be affected much. The opposite is also true as the results of the tracker changes significantly with high variation in the detector performance. Moreover for the Hierarchy tracker, it can be inferred from tables 5 and 1, the average MOTA across datasets exhibits a 45% standard deviation with only, on average, a 6.8% and 11.1% standard deviation in detector recall and precision respectively.

The RJMCMC tracker is overall the second best filter behind the Hierarchy as it has the best results on one over three datasets. Even with its simple target representation (compared to the others) it is more resilient to the detector performance variation – better filtering capabilities. This is due to its target interaction model which seems more relevant and leads to better results (seven out of nine tests compared with the particle filter). In fact, as we can see in Table 5, it has 11% better average MOTA than the DPF. Besides, when the detector performance deteriorates, the RJMCMC has an accuracy higher than the Hierarchy filter (three out of five cases). The DPM has the best accuracy in two out of nine test cases which rates it behind the two other filters. These results are justified by the fact that the particle filter has a simple target representation and no target interaction model. Moreover, it has a better accuracy variation than the Hierarchy filter across the detector performance change and across the dataset change.

MOTP is also affected by the utilized detector. Indeed, a higher precision of the detector in a given sequence will increase the ones of the trackers’. We can notice that the DPM has the best precision, followed by the ACF and finally the Hierarchy tracker. If the particle filter is the best, it is because it creates less tracks than the other two filters which helps it have less Id switches. Moreover, even though the Hierarchy tracker has a more complex target representa-

tion, the target interaction model puts the RJMCMC in front the Hierarchy tracker in terms of Id switches.

5. Conclusions and future works

In this work, we have evaluated three ‘tracking-by-detection’ approaches with three different detector combinations on three public datasets. The results show that the overall performance depends on how challenging the dataset is, the performance of the detector on the specific dataset, and the tracker-detector combination. The choice of the exact detector to use in ‘tracking-by-detection’ should be carefully investigated, and if possible verified on a validation set before plugging into a tracker as state-of-the-art detector does not necessary lead to better detection performance (ACF vs DPM). A tracker using a rich target representation model, as in the Hierarchy, a precise detector, as in DPM, and utilizing an interaction model, as in the RJMCMC, will perform significantly better across datasets of varying challenges.

As a future work, we aspire to broaden the evaluation to incorporate more ‘tracking-by-detection’ approaches and more public datasets. In addition, further investigations would be oriented in coupling the RJMCMC with rich target representation model inspired from the many trackers highlighted in [11].

References


