Multiple Object Tracking using Joint Probability Data Association

Jonne Matthijs van Haastregt (Author)
KTH Royal Institute of Technology
jmvh@kth.se

Loizos Hadjiloizou (Group Partner)
KTH Royal Institute of Technology
loizosh@kth.se

***Abstract*—Various methods exist for multiple target tracking. The most common approaches are Nearest Neighbour (NN), Multiple Hypothesis Tracking (MHT) and Joint Probability Data Association (JPDA). JPDA is especially promising because it is shown to be more robust in case of missing measurement and high clutter densities compared to NN approaches, while being computationally more efficient than MHT. This is because instead of making hard associations as done for approaches such as NN, it uses soft associations which are used as weights to update the estimates. Furthermore, when being extended with a M/N heuristic, the JPDA Filter can be used to track a varying number of targets, however it can sometimes be lacking during the initialisation of new targets when the measurements are noisy. The JPDA Filter in combination with the M/N heuristic has been demonstrated on a video feed of a crossroad to track cars and showed some trouble with track management but very good results once targets were confirmed. For a varying number of targets, other approaches such as the Probability Hypothesis Density Filter or multi-Bernoulli filters could be used to achieve better performance.**

***Index Terms*—JPDA, Multiple Object Tracking,**

I. Introduction

Estimation is an important part of robotics. In order for autonomous processes to function correctly, good estimations of its states and its environment are needed. A variety of estimation methods exist, but most of them rely on two steps: a *predict* step where the new state is predicted using a dynamical model, and an *update* step where the state is updated according to a measurement.

It is not always trivial to know what measurement to use for the update step, since there can be multiple available measurements, including clutter. In order to make sure that the correct measurements are used, data-association methods need to be applied. This can be done by taking e.g. the most likely measurement given the data $p(Z|X)$.

For tracking of multiple objects, there are some additional challenges to overcome. For starters it is usually not known what target was the source of a specific measurement. Moreover a measurement could be the result of two merged targets (and thus have a unresolved source). Also, targets could (temporarily) be appearing or disappearing.

If the data-association is improved, the estimation accuracy will also improve. Therefore the drive to innovate on data-association techniques to increase their accuracy is valuable for countless applications.

A. Related Work

Various methods for multiple target data association already exist. One of the most straight-forward methods is the Global Nearest Neighbour (GNN) method, where the likelihoods of individual associations are computed and using optimization the combination of associations with the biggest total likelihood is found [11].

Another method is Joint Probability Data Association (JPDA). JPDA approximates the joint probability space over all associations using a single gaussian. This means that no hard associations are made, only soft weighted associations. In [6] a JPDA Filter is used for sonar estimation. In [4], JPDA is used with a particle filter to track traffic. In [1], a method is proposed to incorporate past measurements into the JPDA technique.

A third method is Multiple Hypothesis Tracking (MHT) [12]. MHT keeps track of all possible hypothesis. This method is considered to be the most optimal association method, but due to its computational complexity it is not as viable in practice.

JPDA is specifically promising because it has a better accuracy and is more robust compared to GNN as explained in [7], while being less computationally expensive than MHT. However, since JPDA can still get computationally expensive (it increases exponentially with both the number of targets and the number of measurements), gating techniques are often applied to reduce to amount of possible associations, or approximations of the JPDA technique are used that reduce computational time such as in [2] or [7]

These methods can't natively handle a dynamic number of targets. Some methods exist that can 'update' the number of targets in between iterations of the estimation process. JPDA is used in combination with a M/N heuristic for adding and removing targets in [9]. In [3], the tracking is managed by marginalizing the probabilities over all association events where a combination of targets is present and selecting the most probable hypothesis. In [8] JPDA is used with a combination of a Poisson Point Process (PPP) and a multi-Bernoulli filter in order to achieve dynamic targets.

Some alternative estimation methods exist that do natively support dynamic targets, such as the Probability Hypothesis Density (PHD) and multi-target multi-Bernoulli which are both based on Random Finite Sets [10].

B. Own Contribution

In this paper, a JPDA filter is applied with a M/N heuristic to a video feed of traffic. Since various implementations of the JPDA Filter have already been shown to be viable, there is no explicit algorithmic innovation performed in this paper. Instead, this paper will focus on comparisons with GNN associations, visualization of the concept behind JPDA, and incorporating track management.

C. Outline

In section II, JPDA will be explained. In section III the idea of JPDA will be extended to an iterative JPDA Filter that can be applied directly for estimation purposes. In section IV, the complete method for tracking traffic from video footage will be laid out, including the retrieval of measurements and incorporating track management. Finally, in section V three experimental results are shown: A comparisson of JPDA to GNN, A visualisation of the concept behind JPDA, and finally the tracking of traffic from video footage.

II. JPDA

JPDA is an extension of PDA [5] to multiple targets. JPDA computes the joint probability space over all associations and approximates the resulting posterior using a single gaussian. This means that only soft weighted associations are made, instead of hard associations are made such as for e.g. GNN. While the GNN puts a lot of confidence in the most probable association, JPDA keeps the modes of the other possible associations into account. This makes JPDA more robust to unlucky measurement streaks and results in a higher accuracy in cases where there are many clutter measurements around the target.

A. Defining association events

Let τ be the number of targets and n_z be the number of measurements. The variable θ denotes an association event, where the entries θ_t^j for $t \in \{1, 2, ..., \tau\}$ and $j \in \{0, 1, 2, ..., z_n\}$ denote an association between the target t and measurement j. This variable is equal to 1 in case of an association and 0 otherwise. Measurement $j = 0$ is an auxiliary variable that denotes a misdetection.

The set of legal association events Θ cointains all association events θ that conform to the following two rules:

1) Each target t is either detected or misdetected.
2) No pair of targets can be associated to the same measurement.

The number of association events increase exponentially with τ and z_n. The number of legal association events can be computed using equation 1 and it is evaluated for some values in table I. In practise this means that evaluating all possible events becomes computationally intractable and therefore often gating techniques are required. In this paper, gating has not been applied.

$$\sum_{N_D}^{min(\tau, n_z)} \frac{\tau! n_z!}{N_D!(\tau - N_D)!(n_z - N_D)!} \tag{1}$$

z_n \ τ	1	2	3	4	5	6	7
1	2	3	4	5	6	7	8
2	3	7	13	21	31	43	57
3	4	13	34	73	136	339	358
4	5	21	73	209	501	1045	1961
5	6	31	136	501	1546	4051	9276
6	7	43	229	1045	4051	13327	37633
7	8	57	358	1961	9276	37633	130922
8	9	73	529	3393	19081	93289	394353
9	10	91	748	5509	36046	207775	1047376

TABLE I: Number of legal association events

B. Computing likelihoods of association events

Let $\hat{x}_k = \{\hat{x}_{k,1}, \hat{x}_{k,2}, ..., \hat{x}_{k,\tau}\}$ be the state estimates of each target and $z_k = \{z_{k,1}, z_{k,2}, ..., z_{k,n_z}\}$ be the measurements at time k. Using Bayes conditionality the likelihood of an association event can be computed as

$$p(\theta_k | z_k, \hat{x}_k) = \frac{1}{c} p(z_k | \theta_k, \hat{x}_k) p(\theta_k | \hat{x}_k) \tag{2}$$

where c is a normalization constant. The probability of the measurements conditioned on the associations and estimates is given by

$$\begin{aligned} p(z_k | \theta_k, \hat{x}_k) &= \prod_{\theta_t^j \in \theta_k} p(z_{k,j} | \hat{x}_{k,t}) \\ &= \prod_{\theta_t^j \in \theta_k} \frac{1}{(2\pi)^{\frac{M}{2}} |S_{k,t}|^{\frac{1}{2}}} \exp(-\frac{1}{2} (\nu_{k,t}^j)^T S_{k,t}^{-1} \nu_{k,t}^j) \end{aligned} \tag{3}$$

where M is the dimensionality of the states and ν_k, S_k are the innovations and innovation covariances of the Kalman Filter respectively. The probability of the event conditioned on the targets is given by

$$p(\theta_k | \hat{x}_k) = P_D^{\tau - n} (1 - P_D)^n P_{FA}^{(n_z - (\tau - n))} \tag{4}$$

where it is assumed that targets are detected independently. Here P_D is the probability of detection, P_{FA} is the number of false alarm measurements and n is the number of misdetected targets.

It is possible to obtain probabilities for a specific association between a target and a measurement by marginalising over each association event where this association occurs, as shown in equation 5. These probabilities are used as weights in the JPDA Filter in order to update the states and covariances.

$$\beta_t^j = \sum_{\theta : \theta_i^j \in \theta} p(\theta | z, \hat{x}) \tag{5}$$

III. The JPDA Filter

The JPDA Filter can be seen as an extension of the Kalman Filter. Each time step k, it has a prediction step, it performs the data association, and finally uses the marginalized association probabilities from equation 5 in an update step.

For each target the prediction step of the JPDA filter predicts a state estimate $\hat{x}$ and estimation covariance P as follows:

$$\begin{aligned} \hat{x}_{k|k-1} &= F \hat{x}_{k-1|k-1} + G u_{k-1} \\ P_{k|k-1} &= F P_{k-1|k-1} F^T + Q \end{aligned} \tag{6}$$

where u is the input to the system, F, G are the dynamics of the target, and Q is a gaussian process noise.

In order to compute the likelihoods in equation 3, an innovation ν and innovation covariance S are computed for each independent association using

$$\begin{aligned} \nu^j_{k,t} &= z_{k,j} - H\hat{x}_{k|k-1,t} \\ S_{k,t} &= HP_{k|k-1,t}H^T + R \end{aligned} \tag{7}$$

where H is the measurement model and R is a Gaussian measurement noise.

After retrieving the marginalized association probabilities from 5, the update step is performed for each target as follows:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k\mathcal{V}_k \\ P_{k|k} &= \beta^0_{k,t}P_{k|k-1} + (1-\beta^0_{k,t})\bar{P}_k + \tilde{P}_k \end{aligned} \tag{8}$$

where $\beta^0_{k,t}$ is the marginalized probability that the target was not associated to any measurement and $K_k, N_k, \bar{P}_k$ and $\tilde{P}_k$ are given by

$$\begin{aligned} K_k &= P_{k_k-1}H^T(HP_{k|k-1}H^T + R)^{-1} \\ \mathcal{V}_k &= \sum_{j=1}^{n_z}\beta^j_{k,t}\nu^j_{k,t} \\ \bar{P}_k &= P_{k|k-1} - K_k(HP_{k|k-1}H^T + R)K_k^T \\ \tilde{P}_k &= K_k\left(\left[\sum_{j=1}^{n_z}\beta^j_{k,t}\nu^j_{k,t}(\nu^j_{k,t})^T\right] - \mathcal{V}_k\mathcal{V}_k^T\right)K_k^T \end{aligned} \tag{9}$$

So practically the JPDA filter updates the state estimate by weighting the innovations based on the marginalized association probabilities. The state covariance update consists of three parts. $P_{k|k-1}$ is the predicted covariance, weighted by the probability of misdetection, since there is no measurement to update with in that case. $\bar{P}_k$ is the covariance resulting from updating with a measurement, weighted by the probability of a detection. Finally $\tilde{P}_k$ is a covariance term added based on how similar the innovations are. If they are similar, $\tilde{P}_k$ will be small, otherwise there will be some additional uncertainty added.

IV. METHODOLOGY

In order to perform the JPDA Filter on a video of traffic, a dynamical model and measurement model need to be defined. Furthermore, since traffic will enter and leave the camera-view, a method needs to be implemented to achieve a dynamic number of targets.

A. Dynamical Model

It is attempted to estimate the $[x, y]$ states of the traffic participants. Since the camera position and orientation is unknown, the screen-space coordinates will be used and $[x, y]$ will denote the pixels within the video where the traffic participant is present.

Futhermore, since the inputs to the dynamics are unknown, essentially only a random-walk model is viable with a high process noise to compensate for a missing input. In order to direct the random walk in the general direction that the traffic participant is moving, also velocity states are included resulting in the state matrix $x = [x, y, v_x, v_y]^T$. The dynamical model is as follows:

$$x_{k+1} = \begin{bmatrix} 1 & 0 & \delta t & 0 \\ 0 & 1 & 0 & \delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x_k + Q \tag{10}$$

B. Measurement Model

There are various ways to retrieve a location measurement from a video. One way is to use a neural network such as YOLO, however this can be computationally slow. Another way is to look for a color histogram, and retrieve measurement based on screen-space locations that have a similar color histogram compared to the reference [14]. This method was applied together with a JPDA particle filter in [**?**].

In this paper, measurements are taken by a background removal process based on Gaussian mixture models [13] followed by a morphological process to remove noise and find foreground objects with a minimal size. The measurements are then $[x, y]$ screen space coordinates of the centroids of all morphologically disconnected object. The processing steps are shown in figure 1. This method was chosen due to its simplistic implementation in MatLab [15]. The corresponding measurement model is as follows:

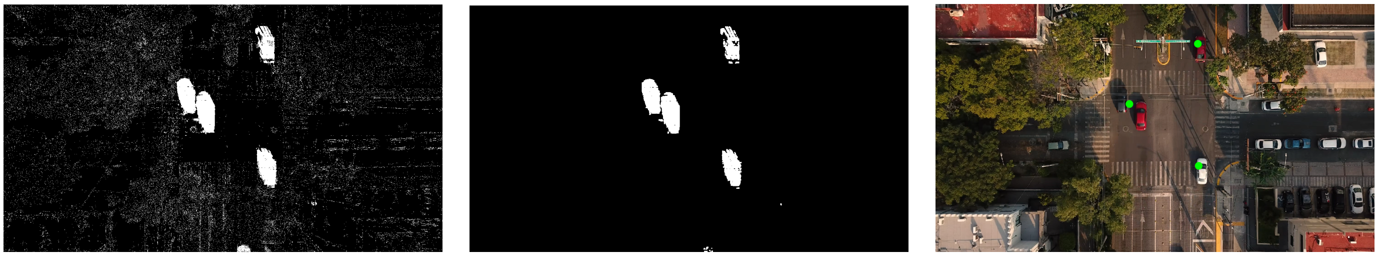
$$z_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} x_k + R \tag{11}$$

C. Dynamic number of targets

While various sophisticated methods exist for tracking a dynamic number of objects, as summarized in section I-A, in order to maintain simplicity a M/N heuristic will be used. The heuristic is based on the idea that in order to differentiate a new target from clutter in case of a measurement without strong associations, a new target should have multiple consecutive measurements. A new target will go on trial bases and be added to the list of confirmed targets when at least M out of N measurements got a sufficiently strong measurement association. If this threshold is not achieved, the new target gets removed from the trial. In a similar way, confirmed targets can be discarded if they have not been associated to any measurement too often.

Targets on trial basis only get processed with measurements that have sufficiently non-probable connections to the confirmed targets. New targets get put on trial basis only after a second measurement, which also has been discarded by the trial targets. Any resulting discarded measurements are turned into new targets. In figure 2 the process is shown schematically.

M and N should be chosen such that $P_{FA} \leq \frac{M}{N} \leq P_D$. Furthermore the heuristic can be tuned by defining what threshold $1 - \beta^0_t > c_{success}$ counts as a success, what threshold $\sum_{t=1}^{\tau}\beta^t_{k,i} < c_{discard}$ results in a measurement being discarded and passed on to the next level, and what radius r_{init}



(a) Background Removal (b) Clutter Removal (c) Resulting measurements (green dots)

Fig. 1: Processing steps to retrieve the measurements from video. This example is taken at frame 140.

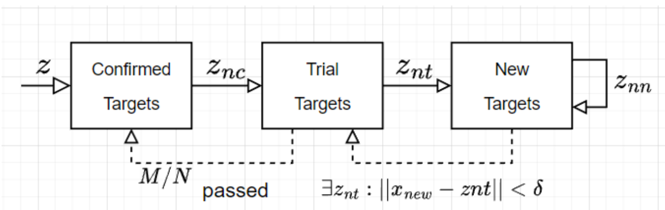


Fig. 2: Flow of measurements and new targets

Simulation Parameters	
Q	0.02
R	0.02
$P_{0\|0}$	0.02
P_D	0.8
P_{FA}	0.5
Results	
MSE_{JPDA}	0.0848
MSE_{GNN}	0.1609

TABLE II: Parameters used for 1D random walk simulation

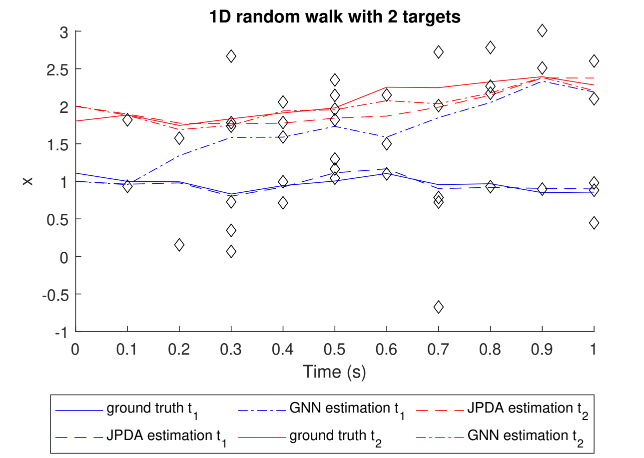


Fig. 3: JPDA compared to GNN on a noisy 1D random walk

is the threshold for a second measurement to promote a new target to a target on trial.

V. Experimental Results

Three experiments are performed. The first and second experiment are used to show some properties of the JPDA Filter in simulation. In the third experiment the JPDA Filter is applied to a video of traffic to track cars.

A. 1D random walk

For the first experiment, a JPDA Filter is used to estimate the one-dimensional random walk of two targets with noisy measurements, and compared to a GNN Filter implementation. The parameters used for the simulation are shown in table II. The MSE performance of both targets after running the simulation 1000 times are shown in table II as well. As can be seen, JPDA outperforms GNN association, as expected.

Furthermore, JPDA indeed proves to be more robust than the GNN method, as shown by an example in figure 3. A streak of unlucky measurements manages to divert the GNN estimation, while the JPDA keeps other association hypothesis in mind and manages to stay on the correct track.

B. 2D simulated trajectories

In this experiment JPDA is applied to three two-dimensional trajectories, that cross each other. Similar parameters are used as for the 1D random walk except extended to two dimensions. The simulation result is shown in figure 4. As can be seen, the JPDA tracks the various targets well, and since it uses the inputs to the trajectories, it also does not have a problem with distinguishing different targets when crossing.

The 2D experiment is a useful simulation to show visually how the associations work. In figure 5 the associations are shown for a single time instance. The associations are visualised using a transparent black line between each target and measurement. The more opaque the line is, the stronger the association probability. As can be seen, multiple measurements can be given a probable association probability, instead of one measurement being given a hard association. When the uncertainty of a targets position is high, a bigger range of measurements is considered, as can be seen at the bottom target. At the same time, it can be seen that measurements that are far outside of the 90% confidence interval are discarded.

C. Traffic tracking

For the final experiment, the JPDA filter has been applied to a video of a crossroad to track cars. The video has been

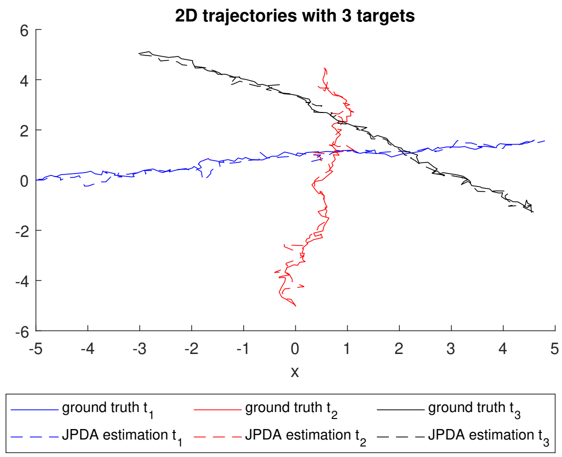


Fig. 4: JPDA used to estimate 2D trajectories

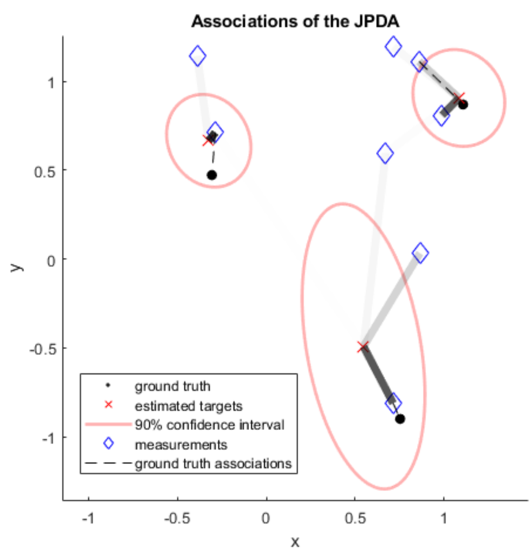


Fig. 5: Associations of JPDA visualized

retrieved from [16]. The experiment is set up as explained in section IV. The specific parameters used are given in table III.

In figure 6, a sequence of frames is highlighted to show the performance of the filter. The full video can be accessed from the github repository [1]. Since there is no ground truth available for this experiment, only a qualitative discussion of the results can be made.

First off, we can see that the performance of the JPDA Filter is quite consistent as it tracks most cars very well, such as the cars with ID 1, 2, and 4 in the frames.

Also, generally the M/N heuristic works adequately to identify new targets and remove targets, however this process can be sensitive to measurement quality. An example of a problematic birthing process is shown by the cars with ID 3

[1]https://github.com/haastregt/JPDA_Filter

JPDA Parameters	
Q	$\begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 50 & 0 \\ 0 & 0 & 0 & 50 \end{bmatrix}$
R	$200I$
$P_{0\|0}$	Q
P_D	0.8
P_{FA}	0.5
M/N heuristic parameters	
N	8
$M_{promote}$	0.6
M_{demote}	2
c_{succes}	0.5
$c_{discard}$	0.1
r_{init}	20

TABLE III: Parameters used for tracking of traffic

and 5. In frame 101, a new target is initialised (cyan color) due to an unassociated measurement. In frame 102, the target is put on trial (magenta color) after a second measurement in the neighbourhood, but note that this is a different car. Since the two cars are very close, occasionally the measurements get merged into a single measurement. Therefore, the new target eventually gets confirmed in frame 110, in between the two cars with a high covariance because the measurements it got during its trial period had high variance. In frame 120, a new target is attempted to be initialised at the car but this fails because of to many missed measurements and in frame 127 this new target has been rejected from confirmation. In frame 143 again a new target is put on trial, and this time it succeeds as can be seen in frame 151.

In general difficulties with birthing appear once a target needs to be initialised close to another existing target, since this existing target may get a small association to the measurements of the new target and prevent this measurement to be passed on to the trial pool. At the same time, one has to strike a balance since if measurements are rejected to easily, clutter near targets may become their own target as can be seen in frame 429 where a clutter measurement has been put on trial near car 31.

VI. Conclusion

It has been shown that JPDA is a very useful and well-performing technique to estimate multiple targets. Especially in cases where there are many missing measurements and clutter, its performance surpasses the performance of a nearest-neighbour approach significantly. However, in order to apply the JPDA Filter to a variable number of targets, extensions are needed. Simple heuristics like the M/N heuristic can prove to have difficulties with noisy measurements. Therefore, other methods such as methods based on Random Finite Sets might be better suited in case of a dynamic number of targets.

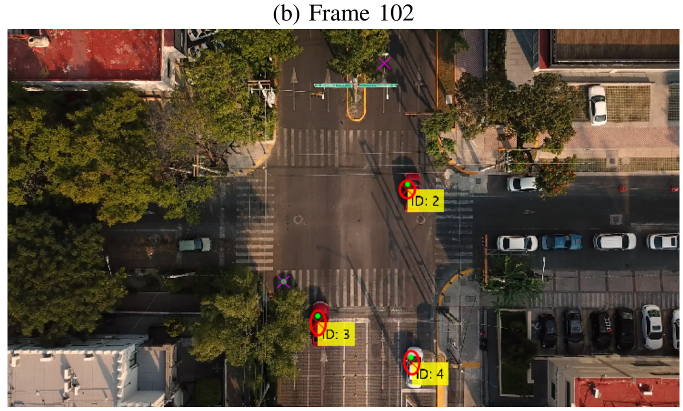
References

[1] S. H. Rezatofighi, A. Milan, Z. Zhang, Q. Shi, A. Dick and I. Reid, "Joint Probabilistic Data Association Revisited," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3047-3055, doi: 10.1109/ICCV.2015.349.

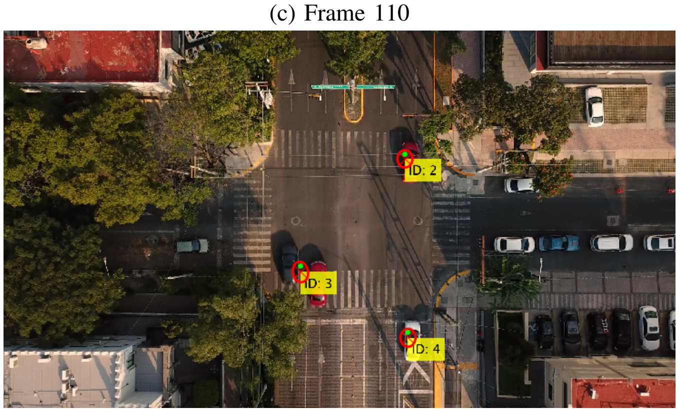
(a) Frame 101

(b) Frame 102

(c) Frame 110

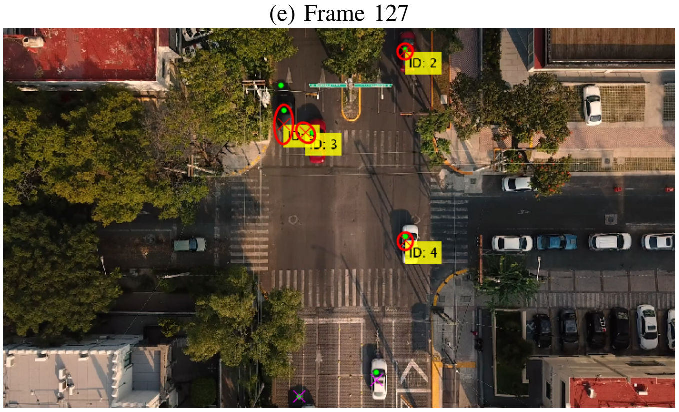


(d) Frame 120



(e) Frame 127

(f) Frame 143



(g) Frame 151

(h) Frame 429

Fig. 6: Sequence of frames from the JPDA traffic tracking

[2] Ashraf M. Aziz, A joint possibilistic data association technique for tracking multiple targets in a cluttered environment, Information Sciences, Volume 280, 2014, Pages 239-260, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2014.04.055.

[3] M. Jaward, L. Mihaylova, N. Canagarajah and D. Bull, ”Multiple object tracking using particle filters,” 2006 IEEE Aerospace Conference, 2006, pp. 8 pp.-, doi: 10.1109/AERO.2006.1655926.

[4] N. T. Pham, K. Leman, M. Wong and F. Gao, ”Combining JPDA and particle filter for visual tracking,” 2010 IEEE International Conference on Multimedia and Expo, 2010, pp. 1044-1049, doi: 10.1109/ICME.2010.5583098.

[5] Y. bar-shalom, F. Daum, J. Huang, ”The probabilistic data association filter,” (2010). Control Systems, IEEE. 29. 82 - 100. 10.1109/MCS.2009.934469.

[6] T. Fortmann, Y. Bar-Shalom and M. Scheffe, ”Sonar tracking of multiple targets using joint probabilistic data association,” in IEEE Journal of Oceanic Engineering, vol. 8, no. 3, pp. 173-184, July 1983, doi: 10.1109/JOE.1983.1145560.

[7] Y. Bar-Shalom, Multitarget-Multisensor Tracking: Advanced Applications. Volume I, Artech-House, Norwood, MA, 1990

[8] He S, Shin HS, Tsourdos A. Joint Probabilistic Data Association Filter with Unknown Detection Probability and Clutter Rate. Sensors (Basel). 2018 Jan 18;18(1):269. doi: 10.3390/s18010269. PMID: 29346290; PMCID: PMC5795933.

[9] A. Benavoli, L. Chisci, A. Farina, S. Immediata, L. Timmoneri and G. Zappa, ”Knowledge-based system for multi-target tracking in a littoral environment,” in IEEE Transactions on Aerospace and Electronic Systems, vol. 42, no. 3, pp. 1100-1119, July 2006, doi: 10.1109/TAES.2006.248193.

[10] J. L. Williams, ”Hybrid Poisson and multi-Bernoulli filters,” 2012 15th International Conference on Information Fusion, 2012, pp. 1103-1110.

[11] Konstantinova, Pavlina & Udvarev, Alexander & Semerdjiev, Tzvetan. (2003). A study of a target tracking algorithm using global nearest neighbor approach. 10.1145/973620.973668.

[12] D. Reid, ”An algorithm for tracking multiple targets,” in IEEE Transactions on Automatic Control, vol. 24, no. 6, pp. 843-854, December 1979, doi: 10.1109/TAC.1979.1102177.

[13] Stauffer, Chris & Grimson, W.. (2007). Adaptive background mixture models for real-time tracking. Proceedings of IEEE Conf. Computer Vision Patt. Recog, vol. 2. 2.

[14] Katja Nummiaro, Esther Koller-Meier, Luc Van Gool, An adaptive color-based particle filter, Image and Vision Computing, Volume 21, Issue 1, 2003, Pages 99-110, ISSN 0262-8856, https://doi.org/10.1016/S0262-8856(02)00129-4.

[15] MatLab, ”Detecting cars using gaussian mixture models”, https://se.mathworks.com/help/vision/ug/detecting-cars-using-gaussian-mixture-models.html, retrieved 09-01-2023.

[16] Free Stock Video, ”Crossroads Traffic Aerial View”, https://www.youtube.com/watch?v=Mp6klx9oeZs&ab_channel=FreeStockVideo, retrieved 09-01-2023.