Maintaining accurate multi-target tracking under frequent occlusion

Yizheng Cai
Department of Computer Science
University of British Columbia
Vancouver, V6T 1Z4
Email:*yizhengc@cs.ubc.ca*
Homepage: *www.cs.ubc.ca/~yizhengc*

Abstract

Multi-target tracking is a very active research topic because of its various applications. Accurately maintaining the identity of each target in the field is the most challenging task in that domain. This project is an extension to the published work on hockey player tracking [1]. It tries to maintain accurate tracking when players cross over each other and result in significant occlusions. It adopts a dynamic Bayesian network (DBN) with two layers of hidden variables to map the track from image to the standard rink coordinate so that the camera motion could be compensated. In addition, the project implements both extended auction algorithm and track-oriented multiple hypotheses tracking (MHT), which are data association approaches to associate the boosting proposal [1] for each track. Results show that the new dynamic model of each target is a good complement when visual information is absent. The extended auction algorithm is a simple and effective approach to solve data association problems in this application, while MHT is tedious to implement, computationally expensive and no more effective than the simple extended auction theory if the tracking is maintained by particle filters.

1 Introduction

Multi-target tracking has attracted attention in the field of signal processing for decades. Mostly, its application is for target tracking using radar or laser sensor signals. Visual tracking has also been studied extensively in recent years. However, most of the related woks either assume that the feature correspondence or data association is already solved or trivial so that nearest neighbor approach is enough. Normally, because the visual information from the features or targets is much better than the radar or laser signals, correlation windows or nearest neighbor approach are indeed adequate. However, in some applications, for example the hockey players tracking by Okuma et al. [1], multiple tracks would confuse with each other or migrate onto only one target with highest likelihood. Therefore, data association techniques are required to maintain accurate identity of each track when visual information is indistinguishable or it is absent because of occlusion. Gating is the basic technique to eliminate very unlikely observation-to-track pairs. However, the tracking in [1] is performed in the image coordinate so that the zooming effect will significantly affect the gating radius' range. Therefore, some preliminary works are required to compensate the camera motion. Details would be presented in next section.

Global nearest neighbor (GNN), joint probability data association (JPDA) and multiple hypotheses tracking (MHT) are most widely used data association approaches in literature. Standard GNN method only considers current observation-to-track assignment and only maintains the most likely association at each step. Thus, error association at any time step would not be corrected at later stages and would result in subsequent errors as well. JPDA is more widely used these days with Particle Filters [2] and Kalman Filters [3] as well. The main difference between GNN and JPDA is that the latter one tries to deal with multiple data association hypotheses. Actually, JPDA is a special case of MHT but relatively simple. One big disadvantage of JPDA is that it does not explicitly handle track birth and death, which is one of the basic requirements for this hockey tracking system. MHT keeps a history of the data association and dynamically update them when new observations are fed in. Thus, error association could be corrected when evidences are updated. Also, the algorithm handles track birth and death explicitly. Investigation of comparison work has been done by [4] to compare the single hypothesis (GNN) and multiple hypotheses. Results show that MHT is much better than GNN. Thus, MHT is the most suitable one for the task we would like to accomplish. However, both JPDA and MHT are computationally expensive, especially for particle filtered tracks. JPDA for Particle Filter has been studied in [2], the application was only to track two people rather than up to 16 people in this project. Therefore, it would be challenging to combine MHT with particle filtered tracks. The remaining of this report will be organized as follows: section 2 would describe the preliminary work that compensates the camera motion; section 3 describes the details of the two data association approaches—extended auction algorithm and track-oriented MHT—that are implemented for comparison; last two sections presents the experiment results of the two approaches and give some discussion.

2 Preliminary Work

The published work [1] on hockey players tracking combined mixture particle filter (MPF) for tracking and boosting detection for particle filter proposal. The tracks are accurate when players are far away from each other. However, when players cross over each other, which is quite frequent in the hockey games, multiple tracks will merge into one track. Therefore, when players split, the system can not tell if it is a new track or an old one. Even with the merging and splitting part of MPF turned off, problems still exist when cross over happens. Different tracks would stick on to the one with the most likelihood because no data association or exclusion principle is applied to keep accurate tracking in those situations.

As is mentioned in previous section, gating is the basis for all the typical data association approaches, it is important to compensate the camera motion and the zooming effect so that the gating range could be constant and easier to apply. Therefore, instead of tracking in the video image coordinate, the particle filter tracking is applied to the standard hockey rink coordinate. This is equivalent to building up a dynamic Bayesian network with two layers of hidden variables with a deterministic mapping between them at each time step. The graphical model can be described as follows, where *X'* is the location of the player in the rink coordinate, *X* is the location in the image coordinate, *Y* is the observation which is evaluated in the image coordinate. There is deterministic way to map the point from the image to the rink and vice versa.

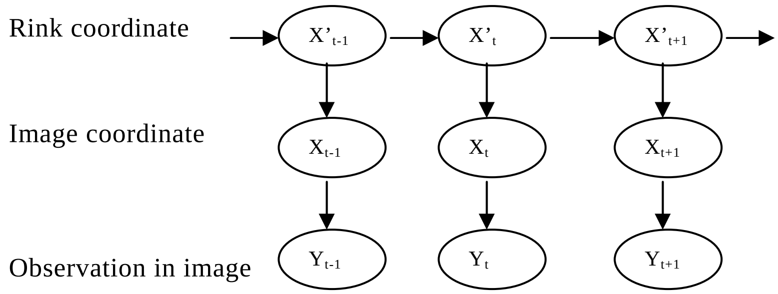


Figure 1: DBN with two layers of hidden variables

With this approach, all the camera motions and zooming effects can be compensated. Therefore, both the visual information and the motion information can be combined to solve the problem during crossovers. Dynamics would be more robust because it's in accordance with physical model which is easy to predict. Thus, it would significantly improve the tracking when visual information is not available during occlusion.

3 Data association for boosting proposal

The classical data association approaches try to find the observation-to-track assignments that can correctly update all the tracks in the field. In this project, because the observation likelihood is evaluated at the location of each particle of the target, there is no need to perform any data association at this stage. However, because the boosting detection result is used as a proposal for each track at each time step, the association is required to assign the boosting detection to the corresponding track. There are various approaches to achieve this goal. Gating is normally the basis for various data association techniques. With the approach mentioned in the previous section, the actual position of each target is in a standard rink coordinate system so that the gating can be implemented with constant radius all over the rink despite of the zooming effect of the cameras. For multiple observations within one gating ellipsoid or within multiple ellipsoids, other techniques are required. This project implemented an extended auction algorithm, which is one of the solutions to GNN approach. In addition, a track-oriented MHT is also implemented for comparison with the simple extended auction algorithm.

3.1 Extended auction algorithm

	Tracks		
Observations	T1	T2	T3
O1	a_{11}	a_{12}	x
O2	a_{21}	x	a_{23}
O3	x	x	x

Table 1: General global assignment matrix

The auction algorithm is originally designed to solve the assignment problem which stem from economic theory. The data association problem can also be considered as an assignment problem with the objective to minimize cost (or maximize profit). The extended auction algorithm is to find the best assignment at each time step based on the generalized global assignment matrix shown above.

In the table, a_{ij} is the cost (or profit) of each observation-to-track pair, x means impossible pairing (the observation is not in the gate of the track). An observation that has impossible assignment to all the existing tracks indicates a new track. The above global assignment matrix can be decomposed into two matrices, one comprises of existing tracks and all the observations within their gates. The other one comprises observations that are not in any existing track's gate, thus indicates a new track. With this decomposition, the first matrix can be solved as a typical assignment problem and the second matrix can be used to create new tracks using some clustering technique. In this project, the entries of the first matrix are considered as profit which is evaluated according to the color likelihood and the distance between the boosting detection and the predicted center of the corresponding tracks:

$$a_{ij} = \frac{p(y_i \mid x_j)}{dist(L_{yi}, L_{xj})}$$

where $p(y_i \mid x_j)$ is the likelihood of the color histogram of boosting detection compared to the reference color model of the corresponding track [1]. L_{yi}, L_{xj} are the location of the boosting detection and predicted center of the track. For the impossible pairing entries, they are assigned negative infinite values.

In standard assignment problem, the assignment matrix is normally symmetric and the assignments are mostly one-to-one. There are various extensions to this standard version, for example the multiple-to-one assignment [3]. In this project, because of the robustness of the boosting detection and the possible preliminary processing of the detection results, the problem can be simplified by assuming that all the observations are not clutter or noise so that they must be assigned to one track. However, during cross over, because it is very likely that some of the occluded players would not be detected by boosting algorithm, it should be allowed that some of the tracks would not have any observation assigned to them. In addition, according to the exclusion principle by Blake et al. [4], any observation can be assigned to only one track but one track might produce multiple observations. Therefore, given the first assignment matrix mentioned in the last paragraph the solution is to find $X = \{x_{ij} \mid x_{ij} \in \{0,1\}\}$ such that $C = \sum_{i=1}^{n} \sum_{i=1}^{m} a_{ij} x_{ij}$ is maximized subject to:

$$\sum_j x_{ij} = 1 \forall i$$

$$\sum_i x_{ij} \geq 0 \forall j$$

This is another extension to the standard auction algorithm with loosened constraint, but the solution is easier:

$$x_{ij'} = \begin{cases} 1 & if \;\; j' = \arg\max_j a_{ij} = 1 \forall i \\ 0 & otherwise \end{cases}$$

The extended version of auction algorithm is easy to implement and effective in real application, although it is not tolerant to the error assignment, especially when error assignment continuous for several time steps.

3.2 Multiple Hypotheses Tracking

MHT is a deferred decision logic in which alternative data association hypotheses are formed whenever there are uncertainties about observation-to-track assignments. "The hypotheses are propagated in anticipation that subsequent data will resolve the uncertainty"[4]. MHT is first denoted as Reid's algorithm. An efficient implementation of Reid's approach is later presented by Cox and Hingorani [5]. The high level logic of MHT can be represented as follows [4]:

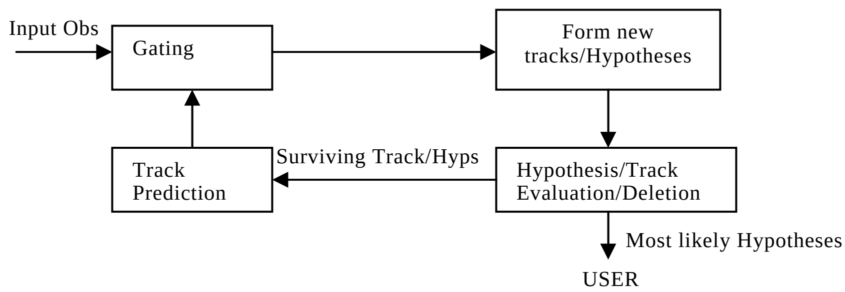


Figure 2: Flow chart of the MHT logic

In the gating step, all the observations and tracks are clustered so that the total number of possible hypotheses would decrease significantly. In the second step of the above chart, new hypotheses will be generated based on the previous ones. It is easy to notice that this step would result in an explosion of the number of hypotheses. The N-best solutions to the assignment problem, which is described in detail by Cox et al. [5], could maintain a stable growth rate of the newly generated hypotheses and make the MHT more computationally feasible. In addition, in order to sustain a reasonable number of hypotheses, N-Scan pruning is used to eliminate hypotheses that are of very low probability. The details of N-Scan pruning will be described in the next section. So far, almost all the implementations of the MHT approach use Kalman Filter for tracking rather than particle filters. Schulz et al. [6] used Rao-Blackwellised particle filter to sample from the distribution of hypotheses for each scan while the posterior distribution of each target is still propagated by Kalman filters. In this project, because the boosting detections over time steps are independent, the association distribution of the observations to tracks is uniform. Therefore, it would not be profitable enough to adopt RBPF in this project. More importantly, because it would be interesting to investigate the combination of MHT logic with the particle filtered tracks, track oriented implementation of MHT is adopted in this project.

3.3 Track Oriented Solution to MHT

The standard MHT implementation maintains and propagates a huge set of hypotheses. In that set, most of the hypotheses are of low probability so that it would be computationally expensive and inefficient to maintain the whole set. In addition, there would be many duplicated tracks in different hypotheses which would result in much redundant space and computation. One alternative implementation is the track-oriented solution, which only maintains a set of possible tracks in a tree structure. Therefore hypotheses are formed only at each stage from tracks and those from the previous stages are discarded right away. The track-oriented approach would be suitable for this project because for each track created, a set of particles need to be generated as well to propagate the distribution. It would cost quite a lot of computation and storage space. Therefore, the track-oriented approach could guarantee a feasible storage requirement. The high level flow chart of the track-oriented approach is shown below [4].

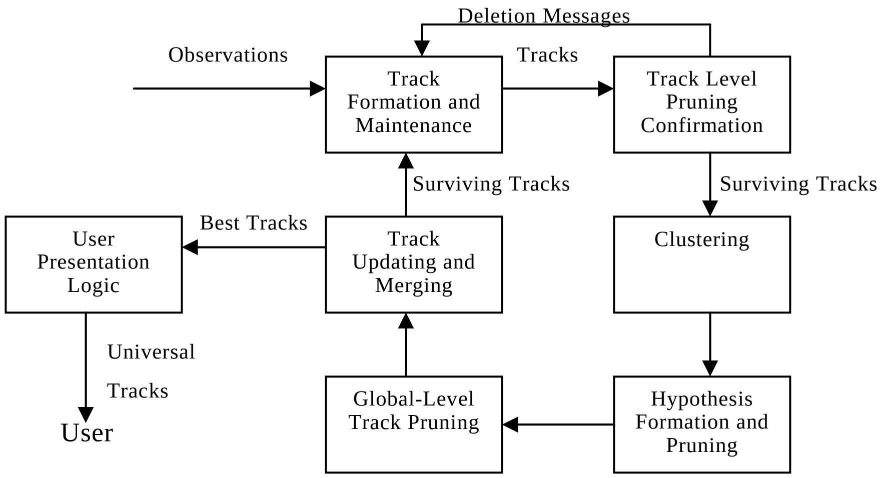


Figure 3: High level flow chart of track-oriented approach to MHT

The first step uses basic clustering technique to initialize or create new tracks. In track level pruning confirmation, the standard way is to evaluate the tracks according to their log likelihood ratio, which is the log of the ratio of the probability that the track is valid to the probability that all observations are false alarms. To simplify the problem, in this project, tracks are evaluated according to the variance of the particle distribution of the posterior or the track scores. Track scores are the basic evaluation for the validity of each track. The way of computing the value would be presented later. In addition, the larger the variance, the more uncertain the track is. Thus, those tracks with large variance should be pruned in the track level as well. During clustering, tracks are linked by the same observations. Tracks in the same cluster are incompatible because they share common observations. Each track maintains a list of tracks that are incompatible with it so that during the formation of hypotheses, the incompatibility can easily be checked. The hypotheses are created each scan using a breadth-first approach. Each hypothesis starts with one track and expands by adding more tracks into it. Although, theoretically, there could be any number of tracks in one hypothesis, in this project, it is simplified that all hypotheses must contain exactly one track in each target family, which is reasonable in application. In addition, any track that is added to the hypothesis must be compatible with all the other tracks in the existing hypothesis. The actual implementation of the first four steps in this project can be represented in Figure 4.

According to Figure 4, at each time step k, track-level pruning is performed before new tracks are created. Tracks with new observations associated are considered as a new track and assigned a new track ID, for example Track 5 in Figure 4. Tracks without any assigned observations (no observation in their gates) would evolve according to their dynamics and maintain the old ID. In this project, a creation of new track would result in creation of a set of particles which propagates from the previous step with the update of the new observations. Meanwhile, the particles of tracks that are from the previous tracks are also maintained for the dynamics model. Maintaining a reasonable number of tracks in the tree is critical to the MHT implementation in this project.

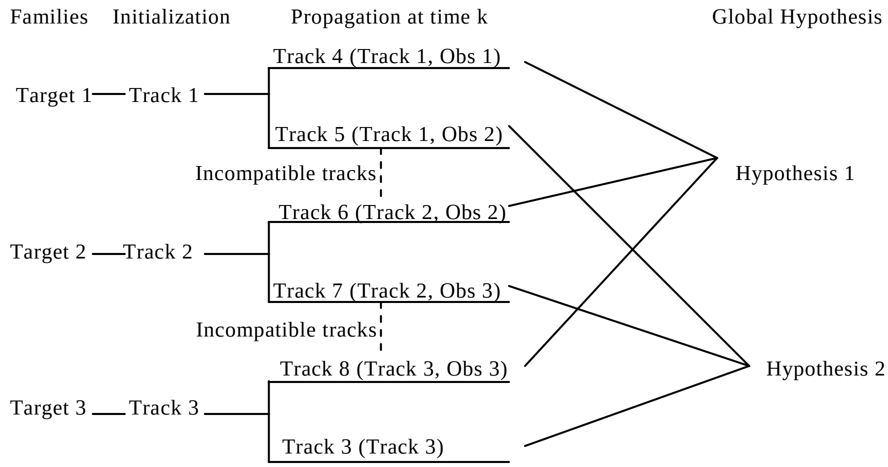


Figure 4: Formations of tracks and hypotheses

The global-level track pruning requires computation of the probability of the tracks and hypotheses from their corresponding scores. The track score is computed in the same way as is mentioned in section 3.1, which could be used for pruning tracks locally before the formation of hypothesis. In this project, it is only used for computing the hypotheses scores.

$$L_i(k) = \frac{p(y_i \mid x_j)}{dist(L_{yi}, L_{xj})}$$

The score of any given hypothesis H_j is just the sum of the scores of all component tracks:

$$L_{H_j} = \sum_{T_i \in H_j} L_i$$

With this score, the probability $p(H_j)$ of the hypothesis j can be computed using all J hypotheses:

$$p(H_j) = \frac{\exp(L_{H_j})}{1 + \sum_{i=1}^{J} \exp(L_{H_i})}$$

Because any track could be contained in multiple hypotheses, the probability of each track is the sum of the probability of hypotheses that contains it. With the probability of hypotheses and tracks, the pruning can be easily performed according to certain thresholds. As is defined, MHT defers the decision when new data arrives to resolve the uncertainty. N-Scan pruning is the most widely used approach for this logic. It first identify the most likely hypotheses in scan *k*. Pruning is accomplished by tracing back N scans from each track in the most likely hypothesis and make that node the new root. Branches that do not have the same new root will be deleted. Figure 5 shows an instance of the N-Scan pruning algorithm, where N=2. Assume, at time k, track 9 is in the most likely hypotheses. 2 scan tracing back makes the track 2 the new root so that the left branch in Family 1 is pruned. New families would be created as a new tree each as new targets are detected. In standard N-scan pruning, it is possible that none of the leaf node of a family tree belongs to the most likely hypothesis. Therefore the whole family would be removed. However, in this project, it is simplified that all new families are valid so that at least one of the leaf node belong to the most likely hypothesis. It would significantly simply both the hypotheses formation and pruning.

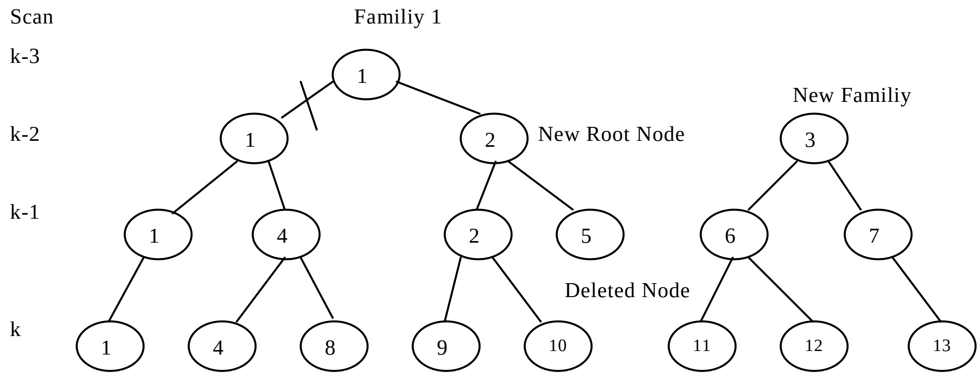


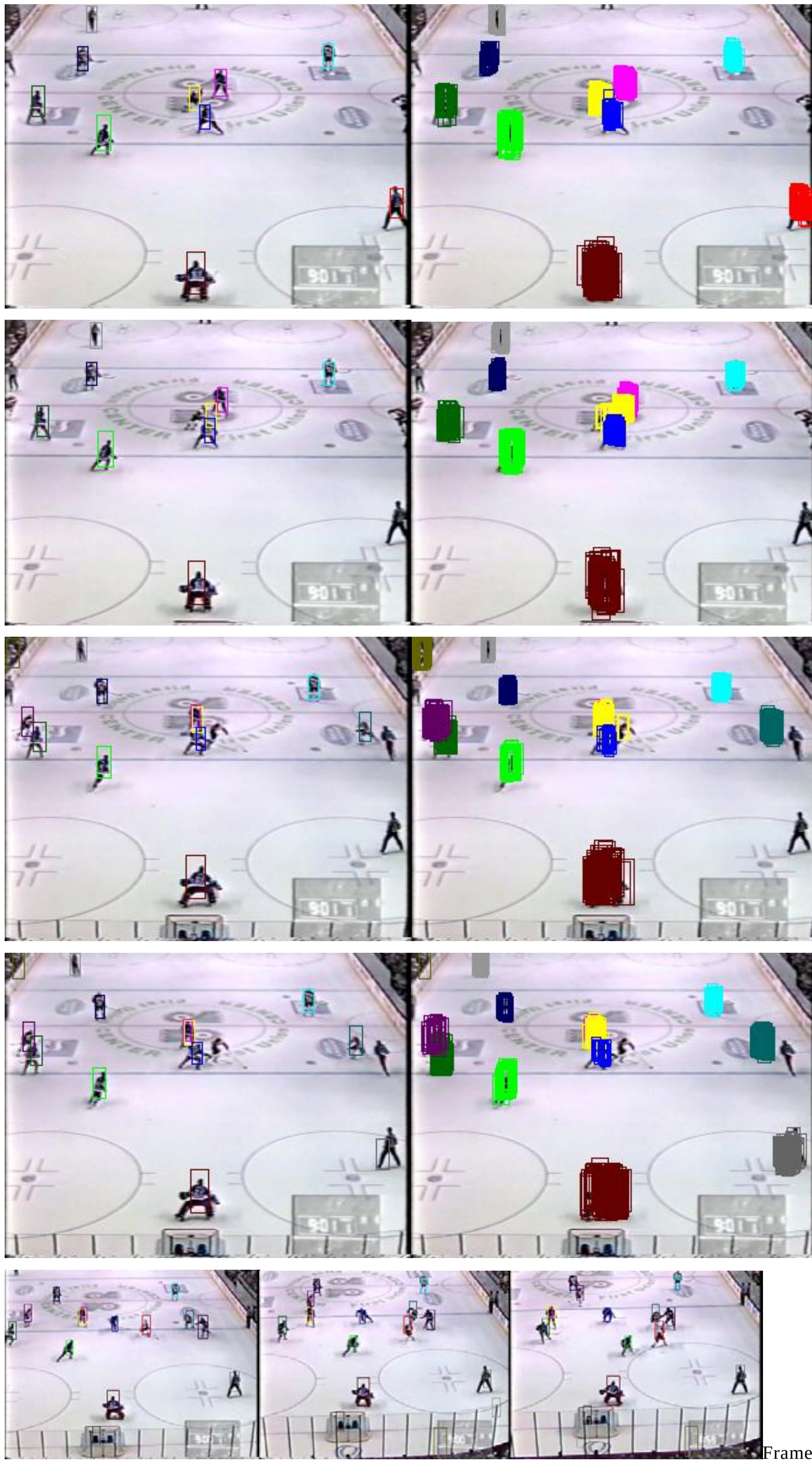
Figure 5: N-Scan Pruning

Finally, the update of the tracks is completed at the same time as the formation of tracks. The dynamics of each player are used to predict the possible position from the last track and associate them to different observations to create different new children in the track tree in the family. Merging is not performed because, in real situation, even if two players overlap for a while they would split at later stage so that merging would lead to the loss of existing tracks. To present the multiple hypotheses, all the best tracks among one target family would be presented to the user as is indicated in figure 3.

4 Experiment Results

Figure 6 shows the tracking result of the extended auction algorithm for the data association. Most of the targets are well tracked except for the yellow and pink players in the middle when three players cross over and overlap significantly. As is seen in frame 36, 48 and 50, the particle distribution on the right shows that multi-modality appears when observations far from the predicted track location are associated. Without support of sufficient visual information, two tracks migrate on the same player. Although exclusion principle is included in the extended auction algorithm, the yellow and pink tracks would still stick together because it is allowed that tracks do not have any observation associated to it. Therefore, one of the tracks could evolve by itself without any boosting proposal so that the exclusion principle is not violated, which is a big disadvantage of the extended auction algorithm. Also, as is stated previously, such error could not be recovered at later stages.

Figure 7 shows the results of the track-oriented implementation of the MHT on the particle filtered tracks. It is tedious to implement all the heuristic algorithms for this project. The original work of this project uses MPF which maintains a constant number of particles for all the tracks. However, in the track-oriented MHT, each track need to maintain a constant number of particles for propagation, which means each track maintains an independent set of particles which would have a number of child sets according to the updates of different observation association hypotheses. In the implementation, the number of observations in the same gate is normally at most three. Therefore, the number of children at each track node is quite limited. The computation, however, is still approximately 4 to 5 times the extended auction algorithm. Dynamically generating and removing particle sets while maintaining the tree structure are computationally expensive. Due to the limit of time, no extensive experiments are done to fine tune the parameters. Most thresholds are directly adopted from [4]. From figure 7, it can be seen that tracks are deleted because of the significant divergence of the particle distribution, which means a whole family of a target tree is deleted because of the large variance or the low target score of all the possible tracks.



Frame 33

Frame 36

Frame 48

Frame 50

Frame 66, 83, 92

Figure 6: Tracking results of the extended auction theory

Theoretically, there should be at least one track in the family that does not have large divergence and has a reasonable high track score. In addition, one big problem with the current implementation is that segmentation fault occurs when processing more than about 30 frames, which might be caused by the exceedingly large storage space required for all the leaf node tracks in the tree structure. Due to the limit of time, no in-depth investigation is done to analyze the exact underlying problem in implementation. It is difficult to verify if it is the limit of the algorithm or bugs in implementation.

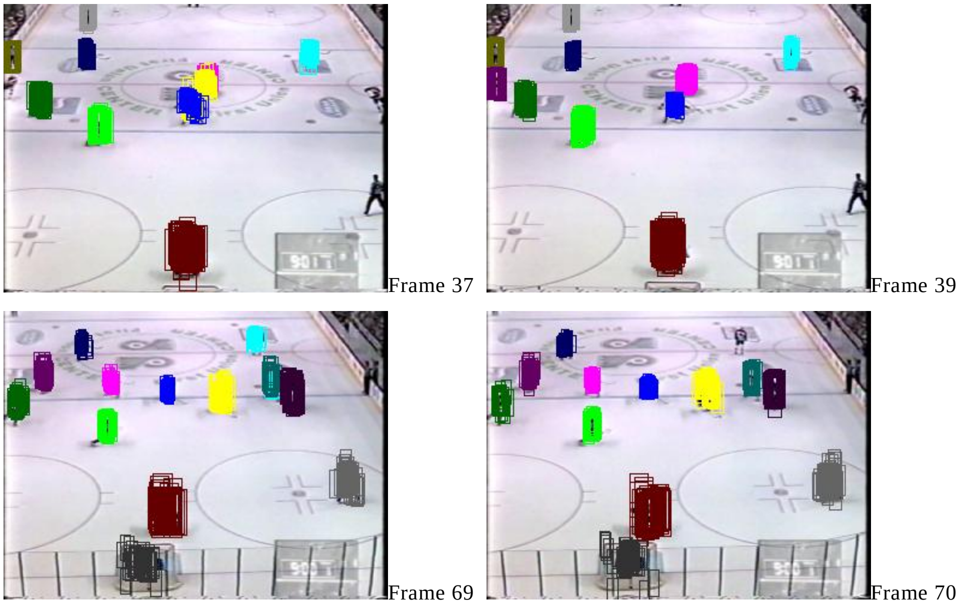


Figure 7: Results of the track-oriented MHT

5 Discussion and Future work

According to the experiment results, the extended auction algorithm is a reasonable choice for this application because of its simplicity and efficiency. Another advantage of the extended auction algorithm is that it would be easy to handle the situation when boosting detection gives many more detections on the same target. Meanwhile, it would be extremely expensive for track-oriented approach to generate that much number of particle sets for each association. Clustering of those multiple observations needs to be explicitly handled, which is also a challenging topic. Because of the insufficient investigation of the track-oriented MHT application, its advantage of recovering error association at later stages is not shown in experiment results. Further experiments and efforts are required to look into the implementation to fine tune the parameters and find out the bug stated in previous section, which is crucial to verify the feasibility of the track-oriented MHT for particle filtered tracks.

References

[1] K. Okuma, N. de Freitas, J. Little and D.G. Lowe. *"Boosted Particle Filters: Multitarget Detection and Tracking"* in European Conference of Computer Vision. 2004.

[2] D. Schultz, W. Burgard, D. Fox, and A.B. Cremers, *"Tracking multiple moving targets with a mobile robot using particle filters and statistical data association"* in IEEE International Conference on Robotics and Automation, 2001.

[3] J. Kang, I. Cohen, and G. Medioni, *"Soccer Player Tracking across Uncalibrated Camera Streams"*, in IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (PETS'03) in conjunction with ICCV, 2003

[4] S. Blackman and R. Popoli, *"Design and Analysis of Modern Tracking Systems"*, Artech House Press

[5] I. J. Cox and S. L. Jones. *"An Efficient Implementation of Reid's Multiple Hypothesis Tracking Algorithm and Its Evaluation for the Purpose of Visual Tracking"* in IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 18, No. 2, 1996

[6] D. Schulz, D. Fox, and J. Hightower, *"People Tracking with Anonymous and ID-Sensors Using Rao-Blackwellized Particle Filters"*, in Proc. of the International joint Conference on Artificial Intelligence (IJCAI), 2003