

PAPER

Multiple Hypothesis Tracking with Merged Bounding Box Measurements Considering Occlusion

Tetsutaro YAMADA^{†a)}, Masato GOCHO[†], Kei AKAMA[†], Ryoma YATAKA[†], and Hiroshi KAMEDA[†], *Members*

SUMMARY A new approach for multi-target tracking in an occlusion environment is presented. In pedestrian tracking using a video camera, pedestrians must be tracked accurately and continuously in the images. However, in a crowded environment, the conventional tracking algorithm has a problem in that tracks do not continue when pedestrians are hidden behind the foreground object. In this study, we propose a robust tracking method for occlusion that introduces a degeneration hypothesis that relaxes the track hypothesis which has one measurement to one track constraint. The proposed method relaxes the hypothesis that one measurement and multiple trajectories are associated based on the endpoints of the bounding box when the predicted trajectory is approaching, therefore the continuation of the tracking is improved using the measurement in the foreground. A numerical evaluation using MOT (Multiple Object Tracking) image data sets is performed to demonstrate the effectiveness of the proposed algorithm.

key words: camera, bounding box, target tracking, multiple hypothesis tracking

1. Introduction

In pedestrian tracking using a video camera, early track confirmation and accurate track maintenance are required. On the other hand, in a crowded environment, the conventional tracking algorithm has a problem that tracking does not continue when the tracking target is hidden behind the foreground object. There is a simple tracking technique using SORT [1]. In SORT, high-speed processing can be realized by filtering the position and size of the detected bounding box using a Kalman filter. The tracking accuracy of DeepSORT [2], which takes into account the visual appearance of the image, is high compared to SORT. There is also a tracking algorithm, MHT (Multiple Hypothesis Tracking) [3], which expresses the correspondence between measurements and tracks as a hypothesis. MHT-DAM [4] has been studied as a method applying MHT to image set. In MHT-DAM, tracking performance can be improved by track-oriented multiple hypotheses and image appearance score. On the other hand, RFS (Random Finite Set) tracking algorithms have been studied [5] and Ref. [6] describes tracking algorithms with merged measurements. Reference [7] is an example of applying an RFS-based multi-objective Bayesian filter to an image, however no explicit treatment in an occlusion environment is mentioned.

Manuscript received September 21, 2021.

Manuscript revised February 20, 2022.

Manuscript publicized May 9, 2022.

[†]The authors are with Mitsubishi Electric Corp., Kamakura-shi, 247-8501 Japan.

a) E-mail: Yamada.Tetsutaro@cb.MitsubishiElectric.co.jp

DOI: 10.1587/transinf.2021EDP7197

In order to cope with occlusion, studies have been made to predict the appearance of pedestrian objects after occlusion using LSTM (Long Short Term Memory) [8], however the learning-based method assumes a large amount of training data sets. Reference [9] is a description of dealing with occlusion for two vehicles based on the assumption of DWT (Discrete wavelet transform) detection and car size prediction.

In this paper, we propose a robust pedestrian tracking method to deal with occlusion that introduces partial information association on the bounding box and a degeneration hypothesis that relaxes the MHT one measurement to one track constraint. The proposed method relaxes the hypothesis that one measurement and multiple trajectories are associated when the predicted trajectory of the tracking approaches, therefore it improves the continuity of multi-target tracking even in an environment where occlusion occurs. The rest of this paper is organized as follows. In Sect. 2, we provide an overview of background and MHT. Next, in Sect. 3, the proposed method is explained. In Sect. 4, numerical evaluation is performed to demonstrate the effectiveness of the proposed algorithms. Finally, we summarize this work in Sect. 5.

2. Background

2.1 Concept of MHT

In hypothesis-oriented multiple hypotheses tracking, multiple hypotheses are generated as the possibility of combinations assuming one-to-one correspondence between measurements and tracks at each sampling time. A measurement is considered as a false alarm, an existing track, or a new target in MHT. The probability of each hypothesis is evaluated by measurement likelihood functions. The tracks in the hypothesis with the highest hypothesis reliability at each sampling time are displayed as a result of the track-to-measurement association. In addition, multiple hypotheses at the current time are constructed based on multiple hypotheses at the previous time. As a result, even if the association result is not correct at a certain sampling time, it is possible to correct the association result after the following sampling time.

2.2 Process Flow of MHT

Figure 1 shows the flowchart for each sampling time of

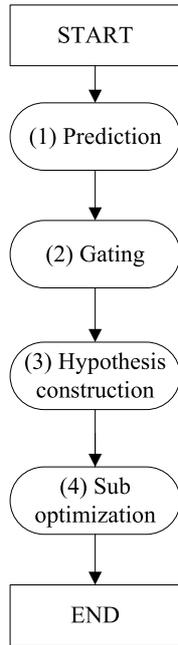


Fig. 1 Flowchart for each sampling time of MHT

MHT. The details of MHT can be found in Ref. [10], pp.402–408, Ref. [11], and Ref. [12]. MHT divides the processing for each cluster. A cluster is a subset of a group of tracks that classifies a track based on whether the track shares measurements or not. Tracks that do not share measurements can be processed separately; hence dividing into clusters can make the processing scale smaller than processing all tracks. In this paper, the explanation of clusters is omitted [10]. The outline of each processing block in Fig. 1 is described below.

(1) Prediction

For all existing tracks, calculate prediction vectors and prediction error covariance matrices at the current time.

(2) Gating

For all existing tracks, software gates centered on the prediction vectors are generated at the current time. The software gate of a track is an ellipsoid that determines whether or not there is a correlation between measurements and the track based on the residual error covariance matrix. If a measurement enters a software gate, the track that generates the gate is associated with the measurement, and if not, it is judged that there is no track-to-measurement association.

(3) Hypothesis construction

For all existing tracks before the current time (which we call “parent tracks”), “child tracks” are derived by connecting measurements at the current time. The child track calculates an updated state vector and an updated error covariance matrix based on the correspondence between the parent track and the measurement which falls in a gate generated by the parent track. Alternatively, the child track is obtained from

the memory track of the parent track. Next for all existing hypotheses, a child hypothesis whose parent is an existing hypothesis is generated. The child hypothesis is a set of tracks satisfying the following conditions 1 to 3. Condition 1 is that the parent track belonging to the parent hypothesis corresponds to at most one of the measurement groups that can correspond; if the corresponding measurement is 1, it is called an update track, and when it is 0, it is a memory track. Condition 2 is that when multiple parent tracks belonging to the parent hypothesis share a measurement, the measurement corresponds to at most one parent track. Condition 3 is that the new track is a measurement that does not correspond to the parent track belonging to the parent hypothesis (first detection).

Then, the likelihood of the child hypothesis is evaluated by hypothesis reliability. The hypothesis reliability of a child hypothesis is calculated as follows:

$$\beta_i = \frac{\gamma_i}{\sum_{n=1}^N \gamma_n} \quad (1)$$

$$\gamma_i = \beta_{p(i)} \prod_{j=1}^{N_{UT}} P_d \cdot g_j (1 - P_d)^{N_{MT}} \beta_{FT}^{N_{FT}} \beta_{NT}^{N_{NT}} \quad (2)$$

The meaning of each symbol is as follows. N : number of parent hypotheses; $\beta_{p(i)}$: parent hypothesis reliability; N_{FT} : number of false signals; N_{UT} : new target number (number of measurements); P_d : detection probability (parameter); N_{UT} : update track number; N_{MT} : memory track number of tracks (parent track number); g_j : likelihood (calculated with filter) of position measurement; β_{FT} : false alarm density (parameter); β_{NT} : new target density (parameter).

The likelihood of the position measurement in Eq. (2) is calculated by

$$g_j = \frac{1}{(2\pi)^{3/2} \sqrt{\det S_k}} \cdot \exp\left(-\left(z_j - z_{k|k-1}\right)^T S_k^{-1} \left(z_j - z_{k|k-1}\right)\right) \quad (3)$$

The meaning of each symbol is as follows. $z_{k|k-1}$: prediction position vector; k : sampling time number; z_j : j th measurement vector; S_k : residual covariance matrix.

(4) Sub optimization

In step 3, all possible child hypotheses are generated; however, there is a problem in that the number of hypotheses explodes and the processing load increases. Therefore, the processing load is reduced by sequentially applying the following semi-optimization process (N best solution search) to this problem. In the N best solution search, top N hypotheses with high reliability are extracted by considering the two-dimensional track-to-measurement association matrix as the allocation matrix and solving the allocation problem using Murty’s algorithm. For details of this process, see [3], [10]. In the MHT, the association result of the best hypothesis (the hypothesis with the highest reliability) is output as the tracking result.

However, there is a problem in that several erroneous tracks can occur when false measurements are caused by

occlusion.

3. Occlusion Countermeasure Multiple Hypothesis Tracking Method

3.1 Principle of Occlusion Countermeasure Multiple Hypothesis Tracking Method

In this section, we explain the concept and process flow of the occlusion countermeasure multiple hypothesis tracking method.

An example of a bounding box in an occlusion environment and the proposed method of filtering using partial measurements during occlusion by the foreground is shown in Fig. 2. In the proposed method, first, gate judgment is performed for each endpoint (xmin, xmax, ymin, ymax) of the bounding box, and after excluding the endpoints that are not associated, the state is updated using a Kalman filter. In the example shown in the figure, the lower side of the bounding box is hidden by the foreground. Therefore, using the endpoints of (xmin, xmax, ymax), the target state can be continuously estimated by performing Kalman filtering.

In Fig. 3, by introducing the hypothesis that associate multiple tracks with one measurement when multiple people overlap in the proposed method, one measurement due to detection failure can be tracked even when multiple people overlap. A merged prediction bounding box is created with the predicted bounding boxes of multiple tracks, and the association is determined based on the difference from the bounding box measurement. When multiple tracks are close to a single measurement, a hypothesis that relaxes the restriction of one-to-one correspondence between measurements and predictions (the degeneration hypothesis) can be used to maintain tracks.

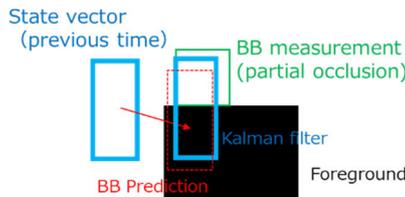


Fig. 2 Concept of partial occlusion filter

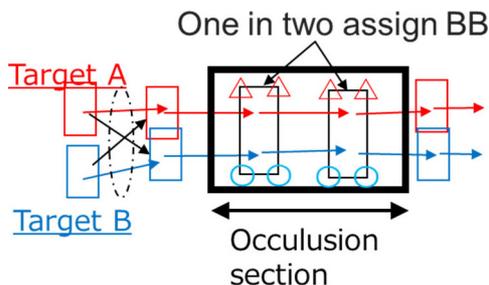


Fig. 3 Association example between existing tracking and measurements with degeneration hypothesis

3.2 Process Flow of the Proposed Method

Figure 4 shows the processing flowchart of the proposed method.

As shown in Fig. 4, the process is the contents modified by modifying step (3) “Making hypothesis with merged measurements” to the conventional MHT and step (4) “Sub optimization.” The details of the processing block with the changes are briefly described.

(3) Hypothesis construction with merged measurements

The state vector is composed of the xy-pixel position and the width and length of the bounding box and the motion model is a constant velocity model. When calculating the likelihood, gate judgment is performed for each endpoint of the bounding box separately with α times the standard deviation of the residuals as the threshold. For the sake of simplicity, in this study, the gate of the equation is determined for the endpoint with the largest error among the four endpoints of the bounding box. The log-likelihood is calculated after constructing the measurement excluding one dimension outside the gate, and the value obtained by multiplying the normal log-likelihood by $4/3$ is calculated as the score when degeneration occurs.

(4) Sub optimization (including the degeneration hypothesis)

Searching for all hypotheses, including those with degeneracy, has the problem of increased computational load. Therefore, in the proposed method, degeneracy judgment is performed based on the proximity of the track and the measured value of the high score hypothesis after using the N-best search algorithm based on the usual one-to-one con-

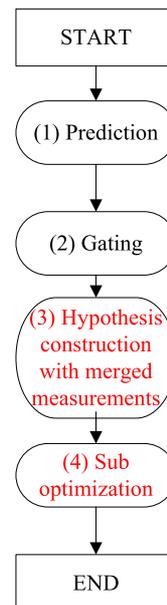


Fig. 4 Flowchart of the proposed method

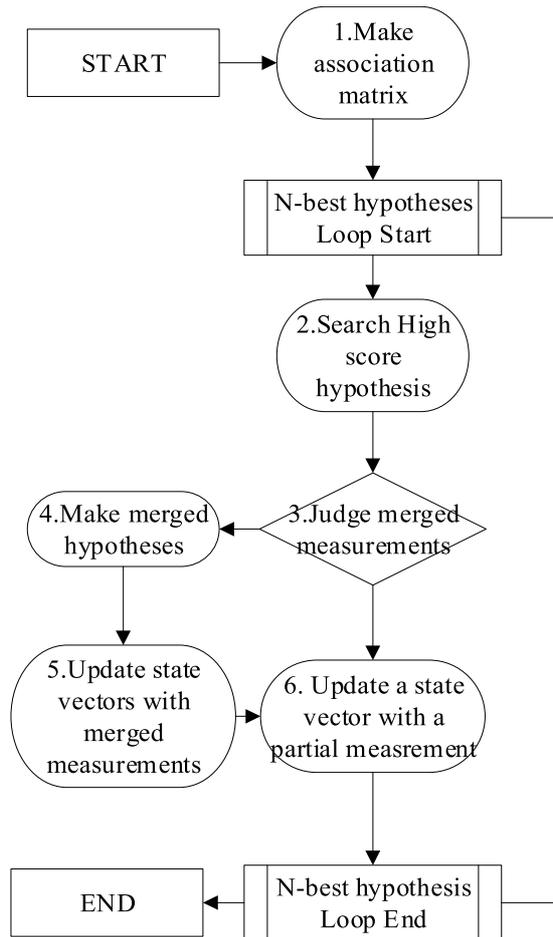


Fig. 5 Flowchart of the proposed method in detail

straint, the feature is that the degeneration hypothesis is derived and generated.

The details of the making hypothesis with merged measurements and the sub optimization method are described later. The processing flow of the proposed method is shown in Fig. 5.

The details of the processing flow are explained below.

First, in step 1 of the figure, a matrix in which scores are arranged when the measurement and track are associated is created as an association matrix. The proposed method determines the association of the endpoints as described in (3) of 3.2. As for the appearance feature, the appearance score which is calculated from the log-likelihood of identifying the same person by a cosine metric such as Deep SORT [2] is added.

Next, in step 3 of the figure, among the normal hypotheses extracted in step 2, a hypothesis with a measurement that may degenerate is determined. Specifically, the hypothesis that the IOU (Intersection Over Union) of the track is associated with the measurement and the memory track included in the hypothesis is more than a certain value (IOU threshold) is determined as a hypothesis that may degenerate.

Then, in step 4 in the figure, the score of the hypothesis with the possibility of degeneration is recalculated using the merged measurement. If the score is higher than the score of the original hypothesis, a hypothesis with the possibility of degeneration is adopted. The most convenient endpoint is selected from the endpoints of the bounding box predicted from multiple tracks and the likelihood of the merged prediction bounding box is compared with the original likelihood.

Then, in step 5 in the figure, the hypothesis that may degenerate is updated using the merged measurement. If there is degeneration, filtering is performed with the measurement error set larger than usual, considering that the observation error of the measurement is large. The visual appearance score is also used with a large standard deviation.

Finally, in step 6 in the figure, the track update processing is included in the normal hypothesis of the assumption that the track and the measurement correspond one by one. In the proposed method, in the association judgment at the endpoints in Fig. 2, filtering is performed using only the information of the associated endpoints.

Compared with the conventional MHT, the tracking performance can be improved by devising merged bounding box measurements and filtering considering the occlusion as described above. The novelty of the proposed method is that it can cope with the degradation of the bounding box, and it is characterized by the following three points.

First, as shown in (3) Making Hypothesis in Fig. 1, the conventional MHT assumes that one measurement corresponds to at most one track. The proposed method is novel in that it generates the hypothesis that multiple tracks correspond to a single measurement (degeneration hypothesis) and defines degenerated the bounding box measurement model. Second, compared with the conventional method of (4) suboptimization in Fig. 1, it efficiently generates N hypotheses with high confidence, including the degeneration hypotheses mentioned above. Finally, it is robust to occlusion compared with the conventional method of the filtered state vector in (3) of Fig. 1. This is because the endpoints of the bounding box can be used to update the state vector in (3) of Fig. 4.

4. Numerical Evaluation

4.1 Simulation Scenario

The effectiveness of the proposed method is verified using two scenarios. In Scenario 1.1, artificial data is created to simulate that the lower side of the bounding box becomes smaller due to occlusion while one target is in progress. The details of scenario 1 are described. Assuming a person who moves in the + X axis direction at a speed of 3 pixels/second with $X, Y = 100$ pixels, 100 pixels as the initial position, the size of the bounding box is $W, L = 50, 200$ pixels. Random noise is added with a standard deviation of 3 pixels. Measurements are made for 10 seconds, and between 5 and 7

seconds, the detection bounding box is reduced because occlusion due to the foreground obscured the upper side of the bounding box (the area where $y > 100$). In scenario 1.2, to investigate the limitation of the proposed method, the statistical evaluation results when the observation ratio (which is the ratio of the size of the observed partial bounding box to the total size of the bounding box that is not obscured by obstacles.) varied from 60% to 100% (no occlusion) are described. The other settings are the same as in Scenario 1.1.

In Scenario 2, the performance of the proposed method is evaluated using public open image data. The open data used training data of MOT16-02, MOT16-04, and MOT16-09 [16], which is data in a crowded environment of multiple people assuming a fixed camera. Pedestrian detection and the visual appearance features [13]–[15] used the same data as Reference [2]. For comparison, a comparison with the literature method [2] is performed.

In summary, scenario 1 is a simple simulation scenario to verify the principle, and to validate the effectiveness of the proposed method when the bounding box is missing due to occlusion, assuming a single target. Scenario 2 is a realistic scenario in a multi-target environment. A practical evaluation of the proposed method is performed using a benchmark of multi-target tracking in a fixed-point camera.

4.2 Parameters of the Proposed and Conventional Methods

Table 1 shows the setting parameters of the proposed method in this study. Measurement noise standard deviation is calculated as a ratio to the size of the bounding box measurement.

4.3 Simulation Results

The tracking center of the bounding box estimated in Scenario 1.1 is shown. The upper diagram in Fig. 6 shows the estimated track (“Conv” in the legend), the true center position (“True” in the legend), and the apparent center position of the bounding box (“BB-Center” in the legend) using the conventional method (normal MHT). The estimated track is pulled to the apparent center of the bounding box after 5 seconds, and it can be seen that the error between “Conv” and “True” increases. In contrast, the figure below shows the center of the bounding box estimated by the proposed method marked “Prop” in the legend. Since the association is performed at the endpoints, it can be seen that the true center position can be estimated even when the bounding box is hidden by occlusion. Figure 7 shows the results of the statistical evaluation for Scenario 1.2, where the horizontal axis is the observation ratio and the vertical axis is the RMSE of the BB estimation. “Proposed” in the legend indicates the error of the proposed method, and “Obs” in the legend indicates the error of the bounding box measurement. Figure 7 shows that if the observation ratio is less than 80%, the proposed method is able to estimate the occlusion and suppress the observation error of the BB. However when the

Table 1 MHT configuration parameters

Item	
Measurement noise standard deviation ratio (x,y,w,l)	(0.1,0.1,0.1,0.1)
Frame per second	30
Process noise power spectrum density	80
Gate size threshold	20
Detection threshold (Scenario 2)	0.3
Detection probability	0.7
Degeneration judgment threshold (IOU)	0.5
Measurement error magnification for update when degenerate	2
False alarm rate	1.0×10^{-10}
New target rate	1.0×10^{-15}
N best hypothesis number	30
Number of measurements to determine track confirmation	5
Hypothesis reliability to determine track confirmation	0.9
Average of appearance score	0.8
Standard deviation of appearance score	0.1

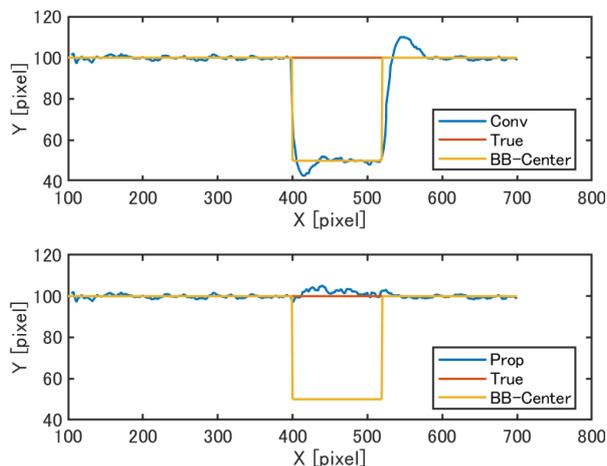


Fig. 6 Tracking result of center position of BB

observation ratio is between 85% and 95%, the observation error is not sufficiently suppressed.

Next, Fig. 8 shows the RMSE of the BB estimation when the observation ratios are 80, 90, and 100. The horizontal axis indicates the frame number and the vertical axis shows the RMSE. The top graph shows the simulation results for an observation ratio of 80, the middle graph for

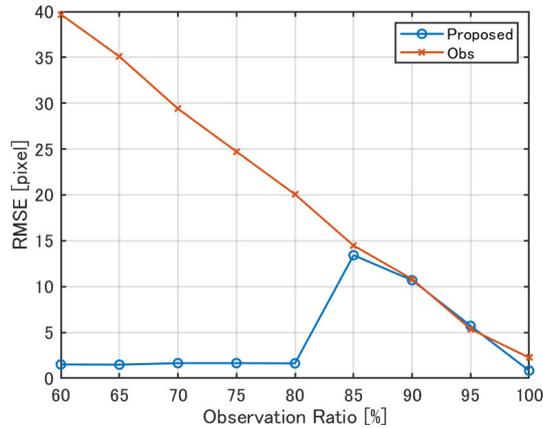


Fig. 7 RMSE of Tracking Performance

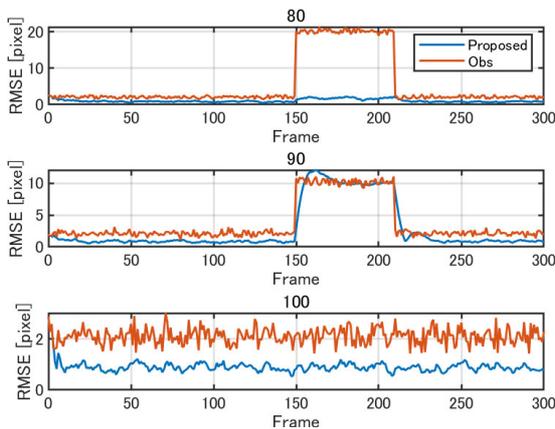


Fig. 8 RMSE of Tracking Performance in detail

Table 2 Performance evaluation (MOT16-02)

	MOTA	MOTP	MT	ML	ID	FM	FP	FN
Prop	16.5	77.4	6	36	20	78	345	14517
Conv	15.2	76.6	3	37	31	114	260	14831

Table 3 Performance evaluation (MOT16-04)

	MOTA	MOTP	MT	ML	ID	FM	FP	FN
Prop	35.3	80.5	6	34	76	389	1566	29122
Conv	33.3	80	6	36	106	561	1023	30584

an observation Ratio of 90, and the bottom graph for an observation ratio of 100 (no occlusion). In the top graph, the observation error is suppressed when occlusion occurs (after 150 frames). However, in the middle graph with an observation ratio of 90, the occurrence of occlusion cannot be estimated, and errors occur due to tracking delays in the frames from 150 to 170, just after the occurrence of occlusion. These results indicate that the proposed method is particularly effective when the observation ratio is less than 80%.

The evaluation results for Scenario 2 are presented in Tables 2, 3, and 4. The tracking evaluation indicators are

Table 4 Performance evaluation (MOT16-09)

	MOTA	MOTP	MT	ML	ID	FM	FP	FN
Prop	51.6	75.2	6	3	43	93	225	2275
Conv	45.7	75.4	4	4	39	119	224	2589

Table 5 Calculation time

	MOT16-02	MOT16-04	MOT16-09
Proposed	18(7.01) ms	89(39.50) ms	18(7.39) ms
Conventional	3(0.56) ms	13(3.46) ms	3(0.51) ms

listed in Ref. [17]. The specific calculation method of the evaluation indexes in the MOT16 benchmark is given in [17]. The index summary is as follows. MOTA (Multiple Object Tracking Accuracy) is the most widely used metric to evaluate a tracker’s performance and it provides a good indication of overall performance. MOTP (Multiple Object Tracking Precision) is the average dissimilarity between all true positives and their corresponding ground truth targets. MT is the number of targets which have been successfully tracked for at least 80% of their lifetime. ML is the number of tracks which are recovered for less than 20% of its total length. ID is a mismatch error, where each of the two tracks is associated with an incorrect measurement corresponding to another track. FM (Track fragmentations) count the number of times the ground-truth trajectory is interrupted, FP (False Positive) represents the number of false alarms, and FN (False Negative) indicates the number of targets that are outliers in any hypothesis. “Prop” in the tables indicates the proposed method and “Conv” indicates the conventional method (Deep-SORT [2]). It is confirmed that the proposed method has a high MOTA score in all scenarios and could reduce ID by approximately 20% while maintaining tracking performance, as shown in Tables 2 and 3. The MT, ML, FM and FN scores are higher than those of the conventional method, and only FP is inferior. This indicates that the proposed method tends to produce numerous false tracks. However, the proposed method outperforms the conventional method in terms of the overall evaluation indices (MOTA and ID). Therefore, it is confirmed that the tracking performance is improved even in a real environment.

The computation time of the proposed method and the conventional method for each scenario are shown in Table 5. The values in the table show the peak computation time for all frames in each scenario, and the values in parentheses show the average computation time. In spite of the complexity of the algorithm, the peak computation time of the proposed method is about 6.0 – 12.5 times less than that of the conventional method. The computation time of the conventional method is measured using OSS in literature [18]. The computer environment is as follows. The CPU is Intel (R) Core @ 3.2 GHz, the installed memory capacity is 8 GB, and the simulator is implemented using C++.

4.4 Discussion

The proposed method is effective if the prediction is reliable to some extent; however, the proposed method has the drawback that it is ineffective where some predictions cannot be trusted using a moving camera. Further evaluation and improvement studies using various scenarios will be performed in the future.

5. Conclusion

In this study, we propose a robust tracking method for occlusion that introduces a degeneration hypothesis that relaxes the MHT measurement to a one-track constraint. The proposed method relaxes the hypothesis that one measurement and multiple trajectories are associated when the predicted trajectory of the tracking approaches; therefore the continuation of the tracking is improved using the measurement in the foreground. The effectiveness of the proposed algorithm is confirmed through numerical evaluation.

References

- [1] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, pp.3464–3468, 2016. doi: 10.1109/ICIP.2016.7533003.
- [2] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, pp.3645–3649, 2017. doi: 10.1109/ICIP.2017.8296962.
- [3] D.B. Reid, "An Algorithm for Tracking Multiple Targets," IEEE Trans. Autom. Control, vol.AC-24, no.6, pp.843–854, Dec. 1979.
- [4] C. Kim, F. Li, A. Ciptadi, and J.M. Rehg, "Multiple Hypothesis Tracking Revisited," IEEE ICCV, pp.4696–4704, 2015.
- [5] R.P.S. Mahler, "Multitarget Bayes Filtering via First-Order Mutitarget Moments," IEEE Trans. Aerospace and Electric Systems, vol.39, no.4, pp.1152–1178, Oct. 2003.
- [6] M. Beard, B.T. Vo, and B. Vo, "Multi-target tracking with merged measurements using labelled random finite sets," 17th International Conference on Information Fusion (FUSION), Salamanca, pp.1–8, 2014.
- [7] N.L. Baisa, "Online Multi-object Visual Tracking using a GM-PHD Filter with Deep Appearance Learning," 2019 22th International Conference on Information Fusion (FUSION), Ottawa, ON, Canada, pp.1–8, 2019.
- [8] M. Babae, Z. Li, and G. Rigoll, "Occlusion Handling in Tracking Multiple People Using RNN," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, pp.2715–2719, 2018. doi: 10.1109/ICIP.2018.8451140
- [9] H. Lu, S. Fei, J. Zheng, and T. Zhang, "An occlusion tolerant method for multi-object tracking," 2008 7th World Congress on Intelligent Control and Automation, Chongqing, pp.5105–5110, 2008. doi: 10.1109/WCICA.2008.4593758
- [10] S.S. Blackman, Multiple Target Tracking with Radar Applications, Artech House, Dedham, 1986.
- [11] T. Yamada, T. Ito, H. Izumi, S. Murayama, Y. Sawayama, and Y. Obata, "Multi-dimensional multiple hypothesis tracking with a Gaussian mixture model to suppress grating lobes," 2017 20th International Conference on Information Fusion (Fusion), Xi'an, pp.1–8, 2017.
- [12] S.S. Blackman, R.J. Dempster, and T.J. Broida, "Multiple hypothesis track confirmation for infrared surveillance systems," IEEE Trans. Aerosp. Electron. Syst., vol.29, no.3, pp.810–824, July 1993.
- [13] F. Yu, W. Li, Q. Li, Y. Liu, X. Shi, and J. Yan, "POI: Multiple object tracking with high performance detection and appearance feature," ECCV Workshops, vol.9914, pp.36–42, 2016.
- [14] N. Wojke and A. Bewley, "Deep Cosine Metric Learning for Person Re-identification," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, pp.748–756, 2018.
- [15] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "MARS: A video benchmark for large-scale person re-identification," ECCV, vol.9910, pp.868–884, 2016.
- [16] A. Milan, L. Leal-Taix'e, I. Reid, S. Roth, and K. Schindler, "MOT16:A benchmark for multi-object tracking," arXiv:1603.00831 [cs], March 2016, arXiv: 1603.00831.
- [17] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The CLEAR MOT metrics," EURASIP J. Image Video Process, vol.2008, 2008.
- [18] A. Pennisi, "Deep Sort algorithm C++ version," https://github.com/apennisi/deep_sort, accessed 2021-12-27.



Tetsutaro Yamada received a B.E. and M.E. degree from Tokyo Institute of Technology, Tokyo, Japan, in 2010 and 2012, and Ph.D. degrees from Tsukuba University in 2019. Since April 2012, he has been with Mitsubishi Electric Corporation, Kamakura, Japan, and has worked on the development of target tracking systems. He is a member of IEEE.



Masato Gocho received the B. Eng. and M. Eng. degrees in information engineering from the University of Waseda, Japan, in 2009 and 2011, respectively. He joined Mitsubishi Electric Corporation in 2011. His research interests include design and implementation of parallel processing techniques and high-performance applications.



Kei Akama received a B.E. and M.E. degree from Waseda University in 2017 and 2019. In 2019, he joined the Information Technology R&D Center, Mitsubishi Electric Corporation, Kamakura, Japan, where he has been engaged in the research and development of target tracking.



Ryoma Yataka received M.S. degree in Information Engineering from University of Tsukuba in 2017. Since 2017, he has worked mainly for the Information Technology Research and Development center (IT R&D), MITSUBISHI Electric Corporation. His research interests are machine learning and computer vision. He is a member of IEICE.



Hiroshi Kameda received a B.E. degree from Yamaguchi University, Yamaguchi, Japan, in 1989, and M.E. and Ph.D. degrees from Kyushu University in 1991 and 2001, respectively. Since April 1991, he has been with Mitsubishi Electric Corporation, Kamakura, Japan, and has worked on the development of target tracking systems. He is a member of IEEE.