Supplementary Material for "Online Multi-Object Tracking and Segmentation with GMPHD Filter and Simple Affinity Fusion"

Young-min Song
Gwangju Institute of Science and Technology
Gwangju 61005, South Korea
{sym}@gist.ac.kr

In this supplementary material, we present the Gaussian mixture probability hypothesis density (GMPHD) filtering for Multi-object tracking and segmentation (MOTS).

1. The GMPHD filter for MOTS

The GMPHD filter [6] has been widely used for online approach in state-of-the-art 2D box MOT methods [1, 2, 4, 5]. Thus, we exploit it for online multi-segment tracking i.e., MOTS. The GMPHD filtering process requires data Association for multi-object tracking. In this section, we address how those key modules works with input/output of data association.

1.1. The GMPHD filter based tracking process

The main steps of the GMPHD filtering based tracking includes *Initialization*, *Prediction*, *Update* where observations (instance segmentation) and states (segments tracks) at time t are represented by:

$$X_t = \{\mathbf{x}_t^1, ..., \mathbf{x}_t^{N_t}\}, \tag{1}$$

$$Z_t = \{\mathbf{z}_t^1, ..., \mathbf{z}_t^{M_t}\}, \tag{2}$$

where a state vector $\mathbf{x_t}$ is composed of $\{x, y, vx, vy\}$ with track ID, and segment mask. x, y, and vx, vy indicate the center coordinates of the mask's bounding box, and the velocities of x and y directions of the object, respectively. An observation vector $\mathbf{z_t}$ is composed of $\{x, y\}$ with segment mask. $\{x, y\}$ of one state initializes one Gaussian model.

Initialization:

$$v_t(\mathbf{x}) = \sum_{i=1}^{N_t} w_t^i \mathcal{N}(\mathbf{x}; \mathbf{m}_t^i, P_t^i), \tag{3}$$

where the Gaussian mixture models are initialized by using the initial observations from detection responses. Besides, when an observation is failed to find the association pair, i.e., updating target state, the observation initializes a new Gaussian model. We call that kind of initialization as *birth*. N_t means the number of Gaussian models. Each Gaussian $\mathcal{N}$ represents each state model with weight w, mean vector m, input state vector x, and covariance matrix P. At this step, we set the initial velocities of mean vector to zeros. Each weight is set to the normalized confidence value of corresponding detection response. Also, how to set covariance matrix P is shown in Subsection 1.3.

Prediction:

$$v_{t-1}(x) = \sum_{i=1}^{N_{t-1}} w_{t-1}^i \mathcal{N}(\mathbf{x}; \mathbf{m}_{t-1}^i, P_{t-1}^i), \tag{4}$$

$$\mathbf{m}_{t|t-1}^i = F\mathbf{m}_{t-1}^i, \tag{5}$$

$$P_{t|t-1}^i = Q + FP_{t-1}^i(F)^T, \tag{6}$$

where we assume that we have already the Gaussian mixture of the target states at the previous frame $t-1$ as shown in (4). F is the state transition matrix and Q is the process noise covariance matrix. Those two matrices are constants in our tracker. Then, we can predict the state at time t using the Kalman filtering. In (5), $\mathbf{m}_{t|t-1}^i$ is derived by using the velocity at time $t-1$. Covariance P is also predicted by the Kalman filtering method in (6).

Update:

$$v_{t|t}(\mathbf{x}) = \sum_{i=1}^{N_{t|t}} w_t^i(\mathbf{z}) \mathcal{N}(\mathbf{x}; \mathbf{m}_{t|t}^i, P_{t|t}^i), \tag{7}$$

$$q_t^i(\mathbf{z}) = \mathcal{N}(\mathbf{z}; H\mathbf{m}_{t|t-1}^i, R + HP_{t|t-1}^i(H)^T), \tag{8}$$

$$w_t^i(\mathbf{z}) = \frac{w_{t|t-1}^i q_t^i(\mathbf{z})}{\sum_{l=1}^{N_{t|t-1}} w_{t|t-1}^l q_t^l(\mathbf{z})}, \tag{9}$$

$$\mathbf{m}_{t|t}^i(\mathbf{z}) = \mathbf{m}_{t|t-1}^i + K_t^i(\mathbf{z} - H\mathbf{m}_{t|t-1}^i), \tag{10}$$

$$P_{t|t}^i = [I - {K_t}^i H]P_{t|t-1}^i, \tag{11}$$

$$K_t^i = P_{t|t-1}^i(H)^T(HP_{t|t-1}^i(H)^T + R)^{-1}, \tag{12}$$

where the goal of update step is deriving (7). First, we should find an optimal observation $\mathbf{z}$ at time t to update a

Gaussian model. The optimal $\mathbf{z}$ among observation set Z makes q_t into the maximum value in (8). In the perspective of application, the update step involves data association. Finding the optimal observations and updating the state models is equal to finding the association pairs. R is the observation noise covariance. H is the observation matrix to transit a state vector to an observation vector. Both matrices are constants in our application. After finding the optimal $\mathbf{z}$, the Gaussian mixture is updated in order of (9), (10), (11), and (12).

1.2. Data association

In Detection-to-Track Association (D2TA), detection can be 2D box, 3D box, and 2D segments. Anyway, for data association, cost matrix should be computed between observations and states. By using the GMPHD filter, we can calculate position and motion affinity to build the cost matrix. In this paper, we assume that detection results is instance segmentation results.

Segment-to-track association. Inputs are equal to observations in (2) and states in (1). By using these inputs in data association, affinity (cost) matrices are computed and the Hungarian method [3] can be used to solve the cost matrices. Then, the segments of observations are assigned to associated tracks.

Affinity calculation. In fact, the GMPHD filter includes Kalman filtering that designs prediction by using linear motion with noise. Therefore, position and motion affinity between i_{th} state and j_{th} observation indicates the probabilistic value $w \cdot q(\mathbf{z})$ by the GMPHD filter as follows:

$$A_{pm}^{(i,j)} = w^i \cdot q^i(\mathbf{z}^j), \tag{13}$$

which is acquired in Update step of the GMPHD filter, see (9).

Finally, the affinities can create a cost matrix as follows:

$$Cost(\mathbf{x^i_{t|t-1}}, \mathbf{z^j_t}) = -\alpha \cdot \ln A_{pm}^{(i,j)}, \tag{14}$$

where α is a scale factor empirically set to 100. If one of affinities is close to zero value like 10^{-39}, the cost is set to 10000 to avoid that final cost becomes infinity value. Then, the final costs ranges 0 to 10000.

1.3. Parameter settings

The matrices F, Q, P, R, and H are used in *Prediction* and *Update*. Experimentally, We set the parameter matrices for the GMPHD filter's tracking process as follows:

$$F = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, Q = \frac{1}{2}\begin{pmatrix} 5^2 & 0 & 0 & 0 \\ 0 & 10^2 & 0 & 0 \\ 0 & 0 & 5^2 & 0 \\ 0 & 0 & 0 & 10^2 \end{pmatrix},$$

$$P = \begin{pmatrix} 5^2 & 0 & 0 & 0 \\ 0 & 10^2 & 0 & 0 \\ 0 & 0 & 5^2 & 0 \\ 0 & 0 & 0 & 10^2 \end{pmatrix}, R = \begin{pmatrix} 5^2 & 0 \\ 0 & 10^2 \end{pmatrix},$$

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

References

[1] N. L. Baisa and A. Wallace. Development of a N-type GM-PHD filter for multiple target, multiple type visual tracking. *Journal of Visual Communication and Image Representation*, 59:257–271, 2019.

[2] Z. Fu, F. Angelini, J. Chambers, and S. M. Naqvi. Multi-level cooperative fusion of GM-PHD filters for online multiple human tracking. *IEEE Trans. Multimed.*, 21(9):2277–2291, 2019.

[3] R. Jonker and A. Volgenant. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38(4):325–340, 1987.

[4] T. Kutschbach, E. Bochinski, V. Eiselein, and T. Sikora. Sequential sensor fusion combining probability hypothesis density and kernelized correlation filters for multi-object tracking in video data. In *AVSS*, 2017.

[5] Y. Song, K. Yoon, Y. Yoon, K. C. Yow, and M. Jeon. Online multi-object tracking with GMPHD filter and occlusion group management. *IEEE Access*, 7:165103–165121, 2019.

[6] B.-N. Vo and W.-K. Ma. The Gaussian mixture probability hypothesis density filter. *IEEE Trans. Signal Process.*, 54(11):4091–4104, 2006.