Detection and Tracking of Multiple Moving Vehicles with a UAV
Min-Hyuck Lee and Seokwon Yeom
School of Computer and Communication Engineering, Daegu University, Gyeongsan, Korea

Abstract
Multiple object detection and tracking are essential for video surveillance. A drone or unmanned aerial vehicle (UAV) is very useful for aerial surveillance due to the efficiency of capturing remote scenes. This paper addresses the detection and tracking of moving vehicles with a UAV. A UAV can capture video sequences above the road at a distance. The detection step consists of frame differencing followed by thresholding, morphological filtering, and removing false alarms considering the true size of vehicles. The centroids of the detected areas are considered as measurements for tracking. Tracking is performed with Kalman filtering to estimate the state of the target based on the dynamic state model and measured positions. In the experiment, three moving cars are captured at a long distance by a drone. Experimental results show that the proposed method well detects moving cars and achieves good accuracy in tracking their dynamic state.

Keywords: UAV/drone imaging, Object detection, Multiple target tracking, Kalman filtering

1. Introduction
Multiple object detection and tracking are important for security and surveillance [1]. Recently, there is increased usage of unmanned aerial vehicles (UAVs) or drones for aerial surveillance due to the efficiency of capturing remote scenes [2]. The image resolution for objects located far from the camera is low. Further, blurring and noise can degrade image quality. Various studies have been conducted on visual detection and tracking of moving objects. Methods based on background subtraction and frame difference were studied in [3–8]. Gaussian mixture modeling (GMM) was used to analyze background and target regions in [3–5]. Scale invariant feature transform (SIFT) was adopted to extract foreground objects from the background scene in [6, 7]. In [8], background was subtracted under Gaussian mixture assumption followed by morphological filters. In [9], hypothesis fitting was adopted for vehicle detection. Long-range moving objects were detected by background subtraction in [10].

In addition to the above-mentioned research, many studies have been conducted on multiple-target tracking [11–14]. High maneuvering, closely located multiple targets, heavy clutter (false alarm), and low detection probability are often considered difficulties to overcome [11]. Kalman filtering provides a solution for estimating the state of the target in real time. It is known to be optimal under independent Gaussian noise assumption [12, 13]. When multiple measurements are detected in a frame, data association is required to assign the measurements to the established tracks [14].
Vison-based target tracking with UAVs has been researched in \cite{15, 22}. Correlation filtering-based tracking was studied in \cite{15}. Closely located-object tracking was performed with feature matching and multi-inertial sensing data in \cite{16}. Pedestrians were tracked by template-matching in \cite{17, 18}. Small animals were tracked with a freely moving camera in \cite{19}. A moving ground target was tracked in dense obstacles areas with UAVs in \cite{20}. Multiple moving targets are tracked by Kalman filtering and IMM filtering in \cite{21} and \cite{22}, respectively. In \cite{23}, position and velocity were estimated with multiple sensor fusion.

In this paper, we address the detection and tracking of multiple moving vehicles obtained by a UAV, which captures videos above the road from a long distance. To achieve this goal, we propose to integrate two systems: one is detection based on image processing, which manipulates pixels in the stationary scene, and the other is tracking based on Kalman filtering, which estimates dynamic states (position and velocity) of targets in the time domain. Object detection is performed through frame differencing and thresholding, morphological filtering, and removing false alarms based on the minimum size of vehicles. Two frames separated by a constant interval are subtracted and turned into a binary image by thresholding, which can represent the difference between the current and the past frames. Two morphological operations, erosion and dilation, are applied to the binary image to extract regions of interest (ROIs), which are the disjoint alternative areas. The ROIs are the candidate areas of moving vehicles. False ROIs are removed by the minimum ROI size, which is based on the true size of moving vehicles. The centroids of ROIs are the results of detection. The centroids represent the measured positions of targets, which will be processed in the next tracking stage. Tracking is performed by Kalman filtering to estimate the state of the target. A nearly constant velocity (NCV) model is assumed for the dynamic state of the target \cite{11}. A position gating process excludes the measurement outside the validation region. The nearest measurement-to-track association scheme assigns one measurement to one track \cite{21}. The tracks are initialized by a two-point differencing initialization method with a maximum speed gating \cite{11}. In the experiments, three moving cars are captured at a long distance by a UAV. The camera built in a UAV points directly downward to the road. Root-mean-squared errors (RMSEs) are obtained for position and velocity in $x$ and $y$ directions. Experimental results show that the location and the speed are well detected and tracked with good accuracy by the proposed method.

The remainder of the paper is organized as follows. Object detection is discussed in Section 2. Multiple target tracking is presented in Section 3. Section 4 demonstrates experimental results. The conclusion follows in Section 5.

2. Object Detection with Frame Differencing

A block diagram of object detection is described in Figure 1. First, frame difference is obtained between the current and the past frames. A thresholding step follows to generate a binary image as:

$$I_T(m, n; k) = \begin{cases} 1, & \text{if } |I(m, n; k) - I(m, n; k - k_d)| > \theta_T, \\ 0, & \text{otherwise}, \end{cases}$$

$$m = 1, ..., M, \quad n = 1, ..., N,$$  \hfill (1)

where $I(m, n; k)$ and $I(m, n; k - k_d)$ are the $k$th frame and the $(k - k_d)$th frame, respectively. $k_d$ is a constant interval for frame differencing, $\theta_T$ is a thresholding value, and $M$ and $N$ are pixel sizes in $x$ and $y$ directions, respectively. Two morphological operations, erosion and dilation, are applied sequentially to the binary image to segment moving vehicles. Erosion and dilation filtering, respectively, are defined as \cite{24}:

$$I_E = \{l \mid E_l \cap I_T = \varphi\},$$  \hfill (2)

$$I_D = \{l \mid D_l \cap I_T \neq \varphi\},$$  \hfill (3)

where $E_l = \{e + l \mid e \in E\}$ and $D_l = \{d + l \mid d \in D\}; E$ and $D$ are the structuring elements for erosion and dilation, respectively. It is noted that $l$ is an integer value less than the image size. Morphological filters extract ROIs, which are the disjoint alternative areas. At last, false ROIs are removed by the minimum ROI size as:

$$O_i(m, n) = \begin{cases} 1, & \text{Size } O_i \leq \theta_S, \\ 0, & \text{otherwise} \end{cases},$$

where $O_i$ is the $i$th ROI and $\theta_S$ is the minimum size of ROI. The center of each ROI is considered as a measurement for target tracking in the next section.

Figure 1. Block diagram of moving object detection.
3. Target Tracking with Kalman Filtering

3.1 System Modeling

The dynamic state of the target is modeled as a NCV model; the target’s maneuvering is modeled by the uncertainty of the process noise, which is assumed to follow the Gaussian distribution. The following is the discrete state equation of a target:

\[
x_t(k+1) = F(\Delta)x_t(k) + q(\Delta)v(k), \quad t = 1, ..., N_T, \tag{5}
\]

\[
F(\Delta) = \begin{bmatrix}
1 & \Delta & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & \Delta \\
0 & 0 & 0 & 1
\end{bmatrix}, \quad q(\Delta) = \begin{bmatrix}
\Delta^2/2 & 0 \\
\Delta & 0 \\
0 & \Delta^2/2 \\
0 & \Delta
\end{bmatrix},
\]

where \(x_t(k) = [x_t(k) \ v_{tx}(k) \ v_y(k) \ v_{ty}(k)]^T\) is the state vector at frame \(k\), \(x_t(k)\) and \(y_t(k)\) are the positions in \(x\) and \(y\) directions, respectively, \(v_{tx}(k)\) and \(v_{ty}(k)\) are the velocities in \(x\) and \(y\) directions, respectively, \(T\) denotes matrix transposition, \(\Delta\) is the sampling time, and \(v(k)\) is a process noise vector, which is Gaussian white noise with the covariance matrix \(Q = \text{diag}([\sigma^2_x \ \sigma^2_y])\), and \(N_T\) is the number of targets. A measurement vector is composed of two components in \(x\) and \(y\) directions. The following is the measurement equation:

\[
z_t(k) = \begin{bmatrix}
z_{tx}(k) \\
z_{ty}(k)
\end{bmatrix} = Hx_t(k) + w(k), \tag{7}
\]

\[
H = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix},
\]

where \(w(k)\) is a measurement noise vector, which is Gaussian white noise with the covariance matrix \(R = \text{diag}([r^2_x \ r^2_y])\).

3.2 Two-Point Differencing Initialization with Velocity Gating

The initial state of each target is calculated by the two-point differencing method with a maximum speed gating [14]. The initial state and covariance matrix of target \(t\) become:

\[
\hat{x}_t(k|k) = \begin{bmatrix}
\hat{x}_t(k|k) \\
\hat{v}_{tx}(k|k) \\
\hat{v}_y(k|k) \\
\hat{v}_{ty}(k|k)
\end{bmatrix} = \begin{bmatrix}
z_{tx}(k) \\
\frac{z_{tx}(k) - z_{tx}(k-1)}{\Delta} \\
z_{ty}(k) \\
\frac{z_{ty}(k) - z_{ty}(k-1)}{\Delta}
\end{bmatrix},
\]

\[
P_t(k|k) = \begin{bmatrix}
r^2_x & \frac{r^2_x}{\Delta} & 0 & 0 \\
\frac{r^2_x}{\Delta} & \frac{r^2_x}{\Delta^2} & 0 & 0 \\
0 & 0 & r^2_y & \frac{r^2_y}{\Delta} \\
0 & 0 & \frac{r^2_y}{\Delta} & \frac{r^2_y}{\Delta^2}
\end{bmatrix}.
\]

The state is confirmed as an initial state of a track if the following speed gating is satisfied:

\[
\sqrt{\hat{v}_{tx}(k|k)^2 + \hat{v}_{ty}(k|k)^2} \leq V_{\text{max}}.
\]

3.3 Prediction and Filter Gain

The state and covariance predictions are iteratively computed with the initial values as:

\[
\hat{x}_t(k|k-1) = F\hat{x}_t(k-1|k-1), \tag{10}
\]

\[
P_t(k|k-1) = FP_t(k-1|k-1)F^T + Q, \tag{11}
\]

where \(\hat{x}_t(k|k-1)\) and \(P_t(k|k-1)\), respectively, are the state and the covariance prediction for target \(t\) and frame \(k\). The residual covariance \(S_t(k)\) and the filter gain \(W_t(k)\), respectively, are obtained as:

\[
S_t(k) = H P_t(k|k-1) H^T + R, \tag{12}
\]

\[
W_t(k) = P_t(k|k-1) H^T S_t(k)^{-1}. \tag{13}
\]

3.4 Data Association

Data association is the process of assigning multiple measurements to established tracks. Measurement gating is a preprocessing that reduces the number of candidate measurements for data association. The measurement gating is performed by chi-square hypothesis testing with the assumption of the Gaussian measurement residuals [11]. Measurements become valid for target \(t\) at frame \(k\) if they exist in the validation region as

\[
Z_t(k) = \{z_{m}(k)\nu_{tm}(k)^T[S_t(k)]^{-1}\nu_{tm}(k) \leq \gamma, \quad m = 1, ..., m_t(k)\},
\]

\[
\nu_{tm}(k) = z_{m}(k) - Hx_t(k|k-1),
\]

where \(z_{m}(k)\) is the \(m\)th measurement vector at frame \(k\), \(\gamma\) is the gating size, and \(m_t(k)\) is the number of candidate measurements for target \(t\) at frame \(k\). The nearest neighbor rule associates a measurement with track \(t\) as:

\[
\hat{m}_t = \arg\min_{m=1, ..., m_t(k)} ||\nu_{tm}(k)^T[S_t(k)]^{-1}\nu_{tm}(k)||.
\]
3.5 State Estimate and Covariance Update

The state estimate and the covariance matrix of targets are updated as follows:

$$\begin{align*}
\hat{x}_t(k|k) &= \hat{x}_t(k|k - 1) + W_t(k)\nu_t m_t(k), \\
P_t(k|k) &= P_t(k|k - 1) - W_t(k)S_t(k)W_t(k)^T.
\end{align*}$$

(17)  (18)

If $m_t(k)$ is equal to zero, i.e., no measurement is associated with target $t$ at frame $k$, they merely become the predictions of the state and the covariance as:

$$\begin{align*}
\hat{x}_t(k|k) &= \hat{x}_t(k|k - 1), \\
P_t(k|k) &= P_t(k|k - 1).
\end{align*}$$

(19)  (20)

3.6 Performance Evaluation

Several metrics are used for performance evaluation, position error, velocity error, and RMSE of position and velocity. The position error of target $t$ at frame $k$ is obtained as:

$$\begin{align*}
e_{px}^t(k) &= |\hat{x}_t(k|k) - x_{true}^t(k)|, \\
e_{py}^t(k) &= |\hat{y}_t(k|k) - y_{true}^t(k)|,
\end{align*}$$

(21)

where $x_{true}^t(k)$ and $y_{true}^t(k)$ are the ground truth of position of target $t$ at frame $k$, in $x$ and $y$ directions, respectively. The ground truth of positions of targets is obtained manually at each scene. The RMSE of position is obtained as:

$$\begin{align*}
RMSE_p(t) &= \frac{1}{K_t(f) - K_t(s) + 1} \left\{ \sum_{k=K_t(s)}^{K_t(f)} \left\{ e_{px}^t(k)^2 + e_{py}^t(k)^2 \right\} \right\}.
\end{align*}$$

(22)

where $K_t(s)$ and $K_t(f)$ are the first and last frame where target $t$ is estimated. The velocity error of target $t$ at frame $k$ is obtained in both $x$ and $y$ directions as:

$$\begin{align*}
e_{vx}^t(k) &= |\hat{v}_x(k|k) - v_{true}^x(k)|, \\
e_{vy}^t(k) &= |\hat{v}_y(k|k) - v_{true}^y(k)|,
\end{align*}$$

(23)

where the ground truth of the velocity is approximated as:

$$\begin{align*}
v_{true}^x(k) &= \frac{1}{2\Delta} \left[ x_{true}^t(k + \delta) - x_{true}^t(k - \delta) \right], \\
v_{true}^y(k) &= \frac{1}{2\Delta} \left[ y_{true}^t(k + \delta) - y_{true}^t(k - \delta) \right].
\end{align*}$$

(24)

4. Results

4.1 Scenario Description

A drone (Phantom 4 Advanced) captures three moving vehicles at 30 frames per second. A total of 372 frames are captured for more than 12 seconds. The size of one frame is $4096 \times 2160$ pixels; the frame is reduced to 20% size, $820 \times 432$ pixels, for efficient image processing. The drone hovers (stays still in the air) while capturing video sequences. One pixel corresponds to 0.1344 m. There are three cars moving, with different initialization positions and maneuverings. The first car (Car 1) is located at the center of the scene from the first frame and moves to the right until it disappear after the 155th frame. The second car (Car 2) moves to the left from the first frame to the 289th frame. The third car (Car 3) moves in the same direction as the second car from the 74th to the 372nd frame. Figure 2(a) shows Cars 1 and 2 at the first frame, Figure 2(b) shows Cars 1-3 at the 150th frame; Figure 2(c) shows Car 3 at the 300th frame. In the figures, the red circles indicate the moving vehicles.

4.2 Detection of Moving Objects

A total of 368 frames from the 5th frame to the 372nd frame is processed to detect moving objects, since $k_d$ in Eq. (1) is set at 4, $\theta_T$ in Eq. (1) is set at 30, and $E$ and $D$ in Eqs. (2) and (3) are set at $[1]_{9 \times 9}$ and $[1]_{15 \times 15}$, respectively. The parameters for thresholding and morphological operations are determined heuristically when better results are produced; $\theta$
Figure 3. (a) Frame differencing; (b) thresholding; (c) morphological filtering and centroids; (d) segmented region by ROI windows after false alarm removal.

Figure 4. Measurements including false alarms for 368 frames.

Table 1. Detection rates

| Car   | Number of frames | Number of detections | Detection rate (%) |
|-------|------------------|----------------------|--------------------|
| Car 1 | 151              | 151                  | 100                |
| Car 2 | 285              | 282                  | 98.95              |
| Car 3 | 299              | 273                  | 91.3               |
| Total/average | 735              | 706                  | 96.05              |

Figure 5. Tracking results when $\Delta = 0.033$ s: (a) Car 1, (b) Car 2, (c) Car 3.

4.3 Multiple Target Tracking

The centroids in Figure 4 become input measurements for tracking. The sampling time $\Delta$ in Eq. (6) is 0.033 s since the frame rate is 30 fps. The standard deviation of process and the measurement noise are set at $\sigma_x = \sigma_y = 30$ m/s$^2$ and $r_x = r_y = 1.5$ m, respectively; $\gamma$ for the gating test is set at 4; and $\sigma_x$ and $\sigma_y$ are conventionally set between 0.5$a_{max}$ and $a_{max}$ where $a_{max}$ is the maximum acceleration of the target [14]; and $r_x$ and $r_y$ are set at half the length of the vehicle. The validation region in Eq. (14) is chi-square distributed with the degree of freedom is equal to the dimension of measurement vectors, thus the probability mass corresponds to 86.5% when $\gamma$ is set at 4. The maximum speed for speed gating of two-point differencing initialization is set at 30 m/s. The initial-state vectors of Cars 1 and 2 are initialized by the measured position in the 5th and 6th frames; the initial state of Car 3 is initialized by the 74th and 75th frames. Cars 1, 2, and 3 are estimated until 155th, 289th, and 372nd frame for 150, 284, and 298 frames, respectively, including initialization. Figure 5(a)-(c) are the results of tracking Cars 1-3, respectively. No false tracks are established, and no track termination scheme is applied.

Figure 6(a) shows the ground truth of the position of Car 1 from the first to the 155th frame. Figures 6(b) and c are the position vectors of Cars 2 and 3, respectively, from the first frame to the 289th frame, and from the 74th frame to the 372nd frame. Figure 7(a)-(c) shows the approximated ground truth of the velocity in $x$ and $y$ directions of Car 1, 2, and 3, obtained by Eq. (24); $\delta$ is set at 26 when the least average velocity RMSE is produced. Figures 8 and 9 show the position in Eq. (21) and velocity errors in Eq. (23), respectively. Large position errors in $x$ direction are mainly caused by measurement errors. High speed generates greater measurement errors. Therefore, measurement errors during detection are bigger in the direction...
of the car’s movement. Velocity errors are generated during tracking. It is noted that the velocity error differences in $x$ and $y$ directions are small compared with those of position errors.

The sampling time is set differently in the following experiments. Figure 10 shows the tracking result with $\Delta = 0.1$ s, that is, measurements are acquired every third frame from the original data. Cars 1 and 2 are initialized by the 5th and 8th frames and Car 3 is initialized by the 74th and 77th frames. Cars 1, 2, and 3 are estimated until 155th, 287th, and 371th frame, respectively, thus Cars 1, 2, and 3 are estimated for 50, 94, and 99 frames, respectively, including initialization. Figure 11 is the tracking results with every sixth frame ($\Delta = 0.2$ s) updated. Cars 1 and 2 are initialized by the 5th and 11th frames and Car 3 is initialized by the 77th and 83rd frames. Cars 1, 2, and 3 are estimated for 25, 47, and 49 frames, respectively, until 155th, 298th, and 371st frame. The relative computational times are 1, 0.51, and 0.34 for $\Delta = 0.033$, $\Delta = 0.1$, and $\Delta = 0.2$ s, respectively.

Table 2 shows the RMSE of position in Eq. (22). The large position RMSEs of Cars 2 and 3 are due to the high maneuvering of the circular motion. The average position RMSE is bigger over many sampling times, as shown in Table 2. Finer sampling steps can reduce position errors in Car 1 and 2, costing more computational resources. Table 3 shows the RMSE of velocity in Eq. (25). The velocity RMSE is also bigger for Cars 2 and 3 than Car 1. It is shown that the average velocity RMSE decreases as the sampling time increases.
5. Conclusion

In this paper, multiple moving objects are captured over long distances by a UAV. The locations of objects are detected using a background scene. Three targets are tracked with Kalman filtering. Experimental results show that the proposed algorithm detects and tracks moving objects with good accuracy. Position and velocity errors are obtained for different maneuvering targets and sampling times. It is shown that the proposed system can adjust the sampling time to build a more efficient system in computational resources. This property is favorable for drone applications because the computational resources in a drone are very limited. This study also proposes new applications for drones, such as traffic control and smart car as well as security and defense. Tracking a large number of targets with high maneuvering remains for future study.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant Number: 2017R1D1A3B03031668).

References

[1] A. Hampapur, L. Brown, J. Connell, S. Pankanti, A. W. Senior, and Y. L. Tian, “Smart surveillance: applications, technologies and implications,” in Proceedings of the 2003 Joint Conference of the 4th International Conference on Information, Communications and Signal Processing and 4th Pacific Rim Conference on Multimedia, Singapore, 2003, pp. 1133-1138. https://doi.org/10.1109/ICICS.2003.1292637

[2] R. Kumar, H. Sawhney, S. Samarasekera, S. Hsu, H. Tao, Y. Guo, et al., “Aerial video surveillance and exploitation,” Proceedings of the IEEE, vol. 89, no. 10, pp. 1518-1539, 2001. https://doi.org/10.1109/5.959344

[3] W. Chung, Y. Kim, Y. J. Kim, and D. Kim, “A two-stage foreground propagation for moving object detection in a non-stationary,” in Proceedings of 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Colorado Springs, CO, 2016, pp. 187-193. https://doi.org/10.1109/AVSS.2016.7738024

[4] C. H. Yeh, C. Y. Lin, H. E. Lai, K. Muchtar, and M. T. Sun, “Three-pronged compensation and hysteresis thresholding for moving object detection in real-time video surveillance,” IEEE Transactions on Industrial Electronics, vol. 64, no. 6, pp. 4945-4955, 2017. https://doi.org/10.1109/tie.2017.2669881

[5] Y. Wu, X. He, and T. Q. Nguyen, “Moving object detection with a freely moving camera via background motion subtraction,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 2, pp. 236-248, 2017. https://doi.org/10.1109/tcsvt.2015.2493499

[6] M. Mousse, C. Motamed, and E. Ezin, “An adaptive algorithm for fast moving object detection from dynamic background based on codebook,” in Proceedings of 2016 4th International Conference on Robotics and Mechatronics (ICROM), Tehran, Iran, 2016, pp. 584-588. https://doi.org/10.1109/ICRoM.2016.7886808

[7] Y. Zhang, X. Huang, J. Li, X. Liu, H. Zhang, and X. Xing, “Research of moving object detection algorithm in transmission lines under complex background,” in Proceedings of 2016 International Conference on Condition Monitoring and Diagnosis (CMD), 2016, pp. 176-179. https://doi.org/10.1109/cmd.2016.7757799

[8] A. Olugboja and Z. Wang, “Detection of moving objects using foreground detector and improved morphological filter,” in Proceedings of the 3rd International Conference on Information Science and Control Engineering, Beijing, China, 2016, pp. 329-333. https://doi.org/10.1109/icisce.2016.80

[9] M. Cheon and H. Lee, “Vision-based vehicle detection system applying hypothesis fitting,” International Journal of Fuzzy Logic and Intelligent Systems, vol. 17, no. 2, pp. 58-67, 2017. http://dx.doi.org/10.5391/IJFIS.2017.17.2.58

[10] S. Yeom, M. H. Lee, and I. J. Cho, “Long-range moving object detection based on background subtraction,” in Proceedings of the 18th International Symposium on Advanced Intelligent Systems (ISIS), Daegu, Korea, 2017, pp. 1082-1085.
[11] Y. Bar-Shalom and X. R. Li, Multitarget-Multisensor Tracking: Principles and Techniques. Storrs, CT: YBS Publishing, 1995.

[12] L. D. Stone, R. L. Streit, T. L. Corwin, and K. L. Bell, Bayesian Multiple Target Tracking, 2nd ed. Boston, MA: Artech House, 2014.

[13] K. R. Hamid, A. Talukder, and A. K. M. Ehtesanul Islam, “Implementation of fuzzy aided Kalman filter for tracking a moving object in two-dimensional space,” International Journal of Fuzzy Logic and Intelligent Systems, vol. 18, no. 2, pp. 85-96, 2018. https://doi.org/10.5391/IJFIS.2018.18.2.85

[14] S. Yeom, T. Kirubarajan, and Y. Bar-Shalom, “Track segment association, fine-step IMM, and initialization with Doppler for improved track performance,” IEEE Transactions on Aerospace and Electronic Systems, vol. 40, no. 1, pp. 293-309, 2004. https://doi.org/10.1109/TAES.2004.1292161

[15] K. Chaudhary, M. Zhao, F. Shi, X. Chen, K. Okada, and M. Inaba, “Robust real-time visual tracking using dual-frame deep comparison network integrated with correlation filters,” in Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017, pp. 6837-6842. https://doi.org/10.1109/IROS.2017.8206604

[16] P. Chen, Y. Dang, R. Liang, W. Zhu, and X. He, “Real-time object tracking on a drone with multi-inertial sensing Data,” IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 1, pp.131-139, 2018. https://doi.org/10.1109/TITS.2017.2750091

[17] C. Bian, Z. Yang, T. Zhang, and H. Xiong, “Pedestrian tracking from an unmanned aerial vehicle,” in Proceedings of 2016 IEEE 13th International Conference on Signal Processing (ICSP), 2016, pp. 1067-1071. https://doi.org/10.1109/ICSP.2016.7877993

[18] K. Haag, S. Dotenco, and F. Gallwitz, “Correlation filter based visual trackers for person pursuit using a low-cost quadrotor,” in Proceedings of the 15th International Conference on Innovations for Community Services (I4CS), 2015, pp. 1-8. https://doi.org/10.1109/I4CS.2015.7294481

[19] B. Risse, M. Mangan, L. Del Pero, and B. Webb, “Visual tracking of small animals in cluttered natural environments using a freely moving camera,” in Proceedings of IEEE International Conference on Computer Vision Workshops, Venice, Italy, 2017, pp. 2840-2849. https://doi.org/10.1109/ICCVW.2017.335

[20] J. Kim and Y. Kim, “Moving ground target tracking in dense obstacle areas using UAVs,” in Proceedings of the 17th IFAC World Congress, Toulouse, France, 2008, pp. 8552-8557.

[21] M. H. Lee and S. Yeom, “Vehicle tracking based on Kalman filtering with drone imaging,” Presented in the 6th International Conference on Green and Human Information Technology (ICGHT), Chiang Mai, Thailand, 2018.

[22] M. H. Lee and S. Yeom, “Multiple target detection and tracking on urban roads with a drone,” Journal of Intelligent & Fuzzy Systems, 2018. https://doi.org/10.3233/JIFS-169847

[23] Y. Cho and H. Lee, “A position and velocity estimation using multifarious and multiple sensor fusion,” International Journal of Fuzzy Logic and Intelligent Systems, vol. 17, no. 2, pp. 121-128, 2017. https://doi.org/10.5391/IJFIS.2017.17.2.121

[24] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed. Upper Saddle River, NJ: Pearson Prentice-Hall, 2008.

**Minhyuck Lee** is currently pursuing a B.S. degree in School of Computer and Communication Engineering at Daegu University in South Korea. He has been working on researches related to image processing using UAV. His research interests include intelligent image processing and target tracking. E-mail: dool0331@daegu.ac.kr

**Seokwon Yeom** received the M.S. and B.S. degrees in Electronics Engineering from Korea University and the Ph.D. degree in Electrical and Computer Engineering from the University of Connecticut. He is currently a professor of School of Computer and Communication Engineering at Daegu University in Korea. He is now performing several research projects related to image and signal processing. His research interests include image processing, target tracking, machine learning, and optical information processing. E-mail: yeom@daegu.ac.kr