Sector Based Scanning and Adaptive Active Tracking of Multiple Objects

Shung Han Cho[1], Yunyoung Nam[1], Sangjin Hong[1] and Weduke Cho[2]
[1]Mobile Systems Design Laboratory, Dept. of Electrical and Computer Engineering,
Stony Brook University-SUNY, Stony Brook, NY 11794 - USA
[e-mail: {shcho, yynam, snjhong}@ece.sunysb.edu]
[2]Dept. of Electrical and Computer Engineering, Ajou University,
Suwon, 443-749 - South Korea
[e-mail: wdukecho@gmail.com]
*Corresponding author: Sangjin Hong

Received March 21, 2011; revised May 19, 2011; accepted June 10, 2011; published June 28, 2011

Abstract

This paper presents an adaptive active tracking system with sector based scanning for a single PTZ camera. Dividing sectors on an image reduces the search space to shorten selection time so that the system can cover many targets. Upon the selection of a target, the system estimates the target trajectory to predict the zooming location with a finite amount of time for camera movement. Advanced estimation techniques using probabilistic reason suffer from the unknown object dynamics and the inaccurate estimation compromises the zooming level to prevent tracking failure. The proposed system uses the simple piecewise estimation with a few frames to cope with fast moving objects and/or slow camera movements. The target is tracked in multiple steps and the zooming time for each step is determined by maximizing the zooming level within the expected variation of object velocity and detection. The number of zooming steps is adaptively determined according to target speed. In addition, the iterative estimation of a zooming location with camera movement time compensates for the target prediction error due to the difference between speeds of a target and a camera. The effectiveness of the proposed method is validated by simulations and real time experiments.

Keywords: Object tracking, object scanning, object detection, object dynamics, zooming

This research is supported by the Ubiquitous Computing and Network (UCN) project, Knowledge and Economy Frontier R&D Program of the Ministry of Knowledge Economy(MKE) in Korea as a result of UCN's subproject 11C3-T3-50S.

DOI: 10.3837/tiis.2011.06.005

1. Introduction

Tracking multiple objects with active cameras has received much attention in the community of surveillance system. Active camera systems can maximize the coverage of the surveillance system by a wide-angle view as well as ensure the high resolution of objects by pan-tilt-zoom (PTZ) cameras. This can be easily extended to post processing applications such as target classification and identification. An active tracking system is required to track many objects at least once while they dynamically move around a surveillance area. The system also needs to accurately predict the zooming location to maximize the zooming level for the close-up. However, the fairness of the target selection is not guaranteed because the system cannot track targets with a wide-angle view during the zooming process. Moreover, the prediction accuracy is degraded by unknown object dynamics and is aggravated by slow camera movements. The inaccurate prediction restricts the zooming level to avoid the tracking failure. Therefore, it is not trivial to achieve object tracking in high resolution as effectively covering many objects in real time.

Numerous active camera systems with PTZ cameras have been developed in the literature. Some approaches are presented to track the only single target with a single PTZ camera as keeping it at the center of image [1][2]. Costello et al. proposed several scheduling policies for a single PTZ camera to zoom in multiple targets but does not consider the decision to determine the zooming location and level for a selected target [3]. Also, many researchers have been focused on the camera scheduling problem with multiple cameras to track multiple targets [4][5][6]. Costello et al. [4] proposed a distributed scheduling algorithm for identifying each person in the scene by a network of PTZ cameras. Qureshi et al. [5] proposed a virtual environment simulator to test various camera sensor network frameworks. Some methods used a master camera with at least one slave PTZ camera to track multiple targets with a wide-angle view all the time [7][8][9][10][11]. The master camera with a wide-angle view coordinates the target selection and the slave PTZ camera captures close-up targets but it requires dedicated relationships among multiple cameras [12][13][14][15].

Active tracking systems require a finite amount of time to zoom in each target because of target tracking and camera movement. While a camera zooms in a target, it cannot track other targets. When a single PTZ camera is used, it is important to minimize the time to track each target and to maximize the zooming level. The zooming location is predicted by estimating the target trajectory. The incorrect zooming location may cause losing the target on an image. However, predicting of a zooming location is not a trivial problem due to unknown object dynamics and slow camera movement. Tordoff and Murray [17] also addressed the problem of the delay in the feedback loops comprising of image capture delay, platform response lag and zoom lens response lag. In order to find the optimal zooming location and level, Kalman Filter was used [16][17][18]. Although the appropriate zoom level is probabilistically found, the zooming level is maximized only when object speed is slow. Therefore, an efficient and accurate active tracking system is required to maximize the zooming level for fast moving objects and/or slow camera movement.

In this paper, we propose an active tracking system with sector based scanning for a single camera. An image is evenly divided into the set of sectors to improve the processing time for target detection. The system sequentially scans the divided sectors to select multiple target effectively. If any change is detected in a sector, the system zooms in the sector for slowly moving targets and/or fast camera movements. To maximize the zooming level of the selected

target against unknown object dynamics, the proposed method tracks and zooms the target by multiple steps with a piecewise linearized object model instead of zooming the target by one step with the inaccurate prediction. We introduce a motion box to ensure the space for the next trajectory estimation. The required zooming level of a motion box at each zooming step is maximized within the expected variation of object velocity and detection. At each step, the target location is predicted with fewer frames to shorten the estimation time so that the accuracy of the target location for unknown object dynamics is improved. Also, the prediction of the target location for a shorter time in each step enables the system to cope with fast moving targets and/or slow camera movements. The time it takes to track each object is minimized to increase the number of tracked targets. Also, the iterative calculation of the zooming location with the camera movement time compensates for the delay in the feedback loops. Finally, the proposed method is verified with real time experiments to demonstrate the effectiveness of the proposed method.

The remainder of this paper has 4 sections. In Section 2, we present the overview of application model and problem description. Section 3 explains the adaptive object zooming to maximize the zooming level. We elaborate the decision strategy for the zooming steps and levels with the camera characteristics. In Section 4, we demonstrate the effectiveness of the proposed method with a real time experiment and discuss the limitation of the proposed method. Finally, our contribution is summarized along with future work in Section 5.

2. Application Model and Problem Description

2.1 Application Model and Motivation

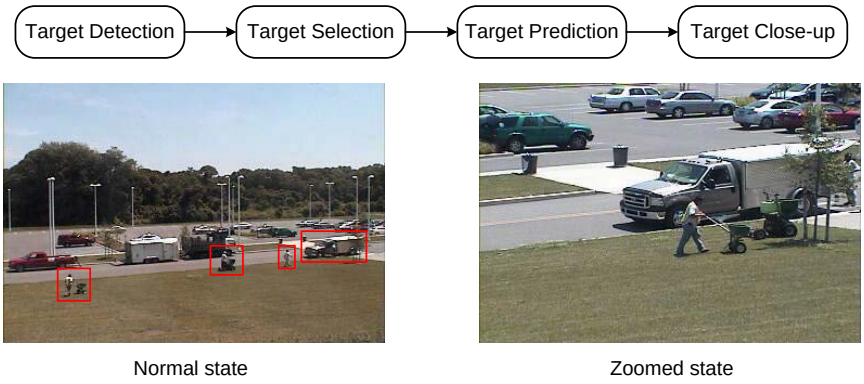


Fig. 1. Illustration of the general process flow for an active tracking system

Active tracking systems obtain the high resolution images for targets in the surveillance region and it facilitates the post processing for target classification and identification. Active tracking systems usually consist of four steps as shown in **Fig. 1**. An active tracking system observes the surveillance region and detects targets at the normal state. The system searches the entire image to select a target for the close-up. The selected target dynamically moves around the surveillance region during the zooming process. Also, a camera requires the finite amount of time to zoom in a specified location. Hence, once the target is selected for the close-up, the system tracks the target trajectory and estimates the target velocity in order to predict the target location for the zooming process. Then, a PTZ camera zooms in the predicted target location. After the zoomed image of the target is captured by the system, the system returns to the

normal state for the next target. Such system should select multiple targets at least once to capture them as many as possible. Also, the system needs to accurately predict the zooming location so that the selected target is zoomed in at the maximum zooming level.

2.2 Problem Description and Approach

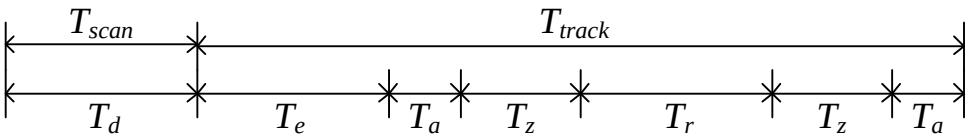


Fig. 2. The detailed time parameters for active target tracking.

The system uses the accurate detection algorithms such as background subtraction method, optical flow method, and statistical learning method [20][21][22] in order to zoom in dynamic objects. **Fig. 2** illustrates the detailed time parameters for sector based zooming for dynamic objects. T_d denotes computation time for target detection at the normal state and T_e denotes computation time for target estimation in the normal state. T_a and T_z denote the translation and zooming time of a camera respectively and they are usually proportional to the amount of the camera movement. In order to maximize the viewing time of a surveillance region, a camera is translated first during the transition to the zoomed state and a camera is zoomed out first during the transition to the normal state. T_r denotes the recording time at the zoomed state and depends on the requirements of applications.

To support a large number of targets with only a single camera, the time it takes to track each target must be short such that all the targets in the surveillance region can be tracked before any of the targets leave the region. The time it takes to track each target consists of times for detection/selection, trajectory estimation, and zooming. The system needs to detect the entire image to select a target. The detection requires a finite number of frames and the long detection time aggravates the selection of multiple targets. The processing time for each frame should be minimized to improve the overall time for the tracking. Also, due to a finite time of the camera movement, a target trajectory must be computed to compensate the time for the camera movement. The compensation for the camera movement time is very critical to accurately track each target. The accurate estimation of the expected location with the trajectory estimation is necessary to maximize the possible zooming level. Otherwise, the actual target may not be in the center of an image.

In zooming a selected target, the system estimates the target velocity to predict the target location for the zooming process. Although Kalman Filter or the probabilistic reasoning provides the accurate estimation with known object dynamics, they usually require the long estimation time and it affects the selection of multiple targets. Moreover, because their techniques are probabilistic, the accuracy of the target location for unknown object dynamics cannot be guaranteed. The inaccurate location restricts the possible zooming level. Also, as the camera movement is slower for the zooming process, the system needs to predict the target location for a longer time. Then, the predicted location can be deviated from the actual location because objects can move unexpectedly during a long movement time of a camera. It increases a chance to lose a target at the zoomed state.

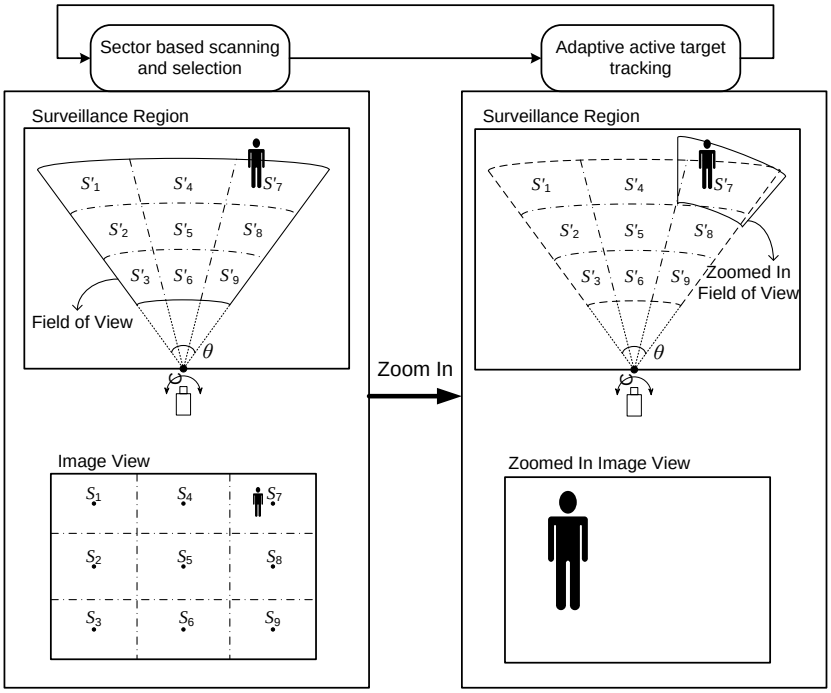


Fig. 3. Illustration of the proposed approach that consists of two different stages of scanning and tracking operations

The proposed approach uses two sub-processes as shown in **Fig. 3**. One is scanning and another is tracking. An image is evenly divided into multiple rectangular sectors to improve the computation time for the target detection. It also has an effect of reducing the search space to select targets. For simplicity, we divide the image of a camera into multiple equally divided rectangular sectors. *n* denotes the number of divided regions along the width of the image, and *m* denotes the number of divided regions along the height of the image. The total number of divided regions is $n \times m$. The size of each rectangular region corresponds to the desired zooming level by

$$z = \min\left(\frac{w}{w'}, \frac{h}{h'}\right), \tag{1}$$

where *z* denotes the maximum zooming level which can be operated by a camera and $w' \times h'$ denotes the size of each sector. The size of a sector is normally greater than at least the size of a target to minimize repetitive selection for the same target. The system scans each sector sequentially to improve the fairness of the target selection. If any change is detected in a sector, the system zooms in the sector for slowly moving targets and/or fast camera movements. When multiple targets exist in one sector, the system randomly selects one target among them for the maximized zooming. The system tracks the remaining targets in the other sectors because targets move around the surveillance region. Even though the remaining targets are stationary, they are tracked eventually when the system scans the same sector again. The maximum number of targets that the proposed method can handle depends on the target speed and the camera movement speed. If the camera movement speed is fast enough that all targets remain in the surveillance region during the one iteration of sector scanning, the proposed method tracks at least one target in each sector. To maximize the zooming level of the selected

target against unknown object dynamics, the proposed method tracks and zooms the target by multiple steps with a piecewise linearized object model instead of zooming the target by one step with the inaccurate prediction. We introduce a motion box to ensure the space for the next trajectory estimation. The required zooming level of a motion box at each zooming step is maximized within the expected variation of object velocity and detection. At each step, the target location is predicted with fewer frames to shorten the estimation time so that the accuracy of the target location for unknown object dynamics is improved. Also, the prediction of the target location for a shorter time in each step enables the system to cope with fast moving targets and/or slow camera movements. The number of zooming steps is adaptively determined according to target speed. The one extreme case is that target speed is relatively faster than the camera movement speed. Because the inaccuracy of the predicted zooming location increases during the slow camera movement, it restricts the zooming level to compromise the inaccuracy of the predicted location at each step and requires multiple steps. On the other hand, when target speed is relatively slower than the camera movement speed, the fast camera movement shortens the amount of prediction time and reduces the inaccuracy of the predicted location. The proposed method achieves the maximum zooming level by a few steps.

3. Adaptive Active Object Tracking

3.1 Active Tracking with Known Dynamics

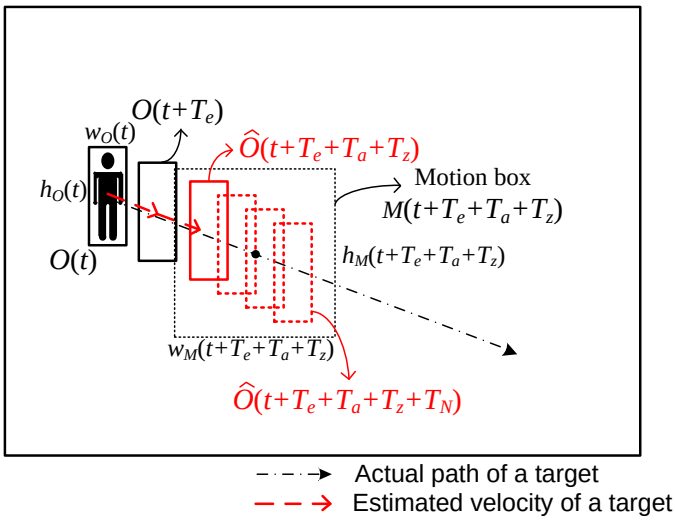


Fig. 4. The determination of the zooming level by the simple linearized trajectory model.

Target $O(t)$ is selected by the proposed sector based scanning and selection strategy at time *t*. The simple zooming method is to zoom it in by a single step with the maximum zooming level. The zooming location is predicted by the estimated velocity during T_e and the zooming level is determined by considering the size of the target. The system uses the simple linear model for the prediction to shorten the estimation time. The zooming level may also consider a space to observe the target at the zoomed state. To include the space for observing the target, we introduce a motion box as shown in **Fig. 4**. For example, if *N* frames are required

at zoomed state, the motion box ensures the space for $\hat{O}(t+T_e+T_t+T_z)$ and $\hat{O}(t+T_e+T_t+T_z+T_N)$. T_N denotes the time to process N frames for the target detection and tracking. The center of a motion box is determined with the two predicted locations by

$$\begin{aligned} x_M(t+T_e+T_a+T_z) &= (\hat{x}_O(t+T_e+T_a+T_z)+x_O(t+T_e+T_a+T_z+T_N))/2, \\ y_M(t+T_e+T_a+T_z) &= (\hat{y}_O(t+T_e+T_a+T_z)+y_O(t+T_e+T_a+T_z+T_N))/2, \end{aligned} \tag{2}$$

where $x_M(t)$ and $y_M(t)$ denote the center coordinate of motion box $M(t)$ and, $\hat{x}_O(t)$ and $\hat{y}_O(t)$ denote the center coordinate of target $\hat{O}(t)$. Since the size of a motion box needs to keep the ratio of an image, the possible width and height of a motion box are calculated first. They are represented by

$$\begin{aligned} h_M(t+T_e+T_a+T_z) &= 2\,|\,\hat{y}(t+T_e+T_a+T_z)-y(t+T_e+T_a+T_z+T_N)+h_O(t+T_e)/2\,|, \\ w_M(t+T_e+T_a+T_z) &= 2\,|\,\hat{x}(t+T_e+T_a+T_z)-x(t+T_e+T_a+T_z+T_N)+w_O(t+T_e)/2\,| \end{aligned} \tag{3}$$

where $w_O(t)$ and $h_O(t)$ denote the width and height of target $O(t)$ respectively and, $w_M(t)$ and $h_M(t)$ denote the possible width and height of a motion box $M(t)$. The zooming level is decided by making the larger size of a motion box as follows,

$$z' = \min\left(\frac{w}{w_M(t+T_e+T_a+T_z)}, \frac{h}{h_M(t+T_e+T_a+T_z)}\right). \tag{4}$$

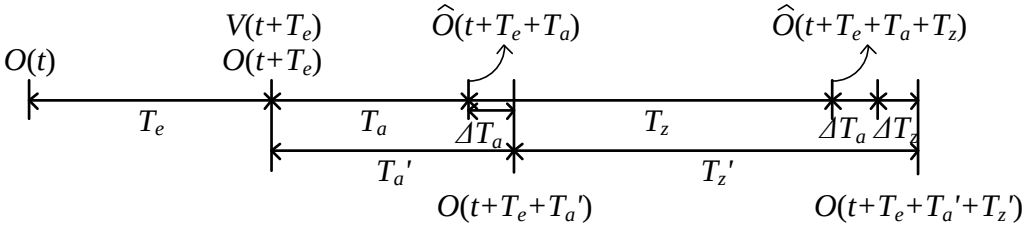
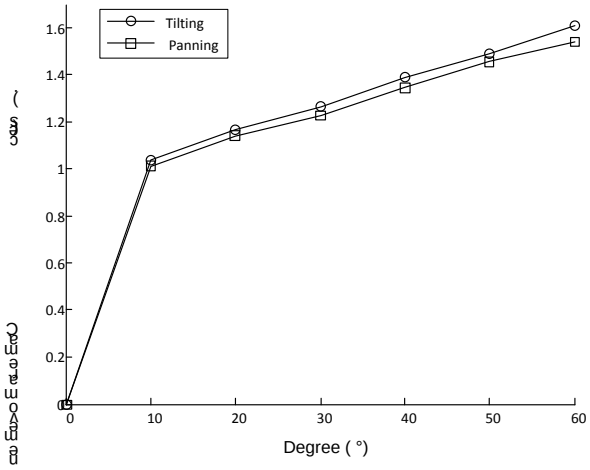


Fig. 5. The timing diagram to show the effect of object movement during the camera movement.

Although the zooming location is correctly predicted with known dynamics, it can be deviated by object movement during the camera movement as shown in **Fig. 5**. For example, the system estimates the velocity of target $O(t)$ during T_e. $V(t+T_e)$ denotes the estimated velocity of target $O(t+T_e)$. Then, the system calculates the predicted location of the target after the translation and zooming time of camera T_a and T_z. However, the actual movement time for the translation and zooming in the predicted location is different from T_a and T_z. The actual times for the translation and the zooming are denoted by T_a' and T_z' respectively. They are calculated by considering both the predicted location and the camera movement speed. Because of the time difference between T_a+T_z and $T_a'+T_z'$, denoted by $\Delta T_a+\Delta T_z$, the zoomed target can be deviated from the predicted location.

In order to cope with the prediction error due to $\Delta T_a+\Delta T_z$, we compensate for the prediction error with the camera movement speed. The time to move (i.e., zoom and translate) a camera depends on the next location on the image and the specifications of a camera. Once the next location to be translated and zoomed is determined by the system, the camera

movement time is calculated by the movement speed in the specifications of a camera. However, the actual movement time should also consider the time to transfer and execute commands to a camera. For instance, **Fig. 6** shows the measured movement time for panning, tilting and zooming of AXIS 214 PTZ network camera [19].



(a) The measured panning and tilting time

Zoom to / Zoom from	2X	3X	4X	5X	6X	7X	8X	9X	10X	11X	12X	13X	14X	15X	16X	17X	18X
1X	0.9669	1.2874	1.5065	1.6394	1.7321	1.7924	1.8722	1.9099	1.9416	2.0177	2.1197	2.2118	2.3195	2.4004	2.5009	2.68285	2.7558
2X		0.69	0.9	1.0281	1.16	1.2332	1.2824	1.3419	1.4028	1.4185	1.515	1.5941	1.7243	1.834	1.8862	2.0346	2.0442
3X			0.54	0.6953	0.8244	0.8582	0.9444	1.0284	1.0752	1.121	1.1704	1.2704	1.3794	1.468	1.5038	1.7204	1.7758
4X				0.4792	0.5701	0.6645	0.7838	0.8266	0.8616	0.8928	0.9794	1.0494	1.1342	1.2768	1.389	1.4812	1.5873
5X					0.4598	0.5719	0.6152	0.654	0.711	0.7398	0.9748	0.9748	1.0404	1.1034	1.2366	1.384	1.4431
6X						0.4717	0.5273	0.5478	0.6084	0.7024	0.7374	0.8076	0.9772	1.0284	1.088	1.2514	1.2587
7X							0.4704	0.4483	0.5776	0.596	0.6287	0.7698	0.824	0.99	1.0342	1.1406	1.2444
8X								0.45	0.4593	0.5387	0.5924	0.7312	0.7758	0.9372	0.923	1.099	1.1626
9X									0.46	0.4629	0.5865	0.6303	0.7574	0.8118	0.9402	1.055	1.1091
10X										0.4531	0.4195	0.5834	0.7092	0.7618	0.8704	0.9678	1.0406
11X											0.37	0.4976	0.5648	0.6296	0.7964	0.8868	1.0122
12X												0.4399	0.5444	0.69	0.7172	0.8536	0.9778
13X													0.45	0.5072	0.6634	0.7678	0.8966
14X														0.47	0.6016	0.681	0.771
15X															0.4528	0.5255	0.6391
16X																0.46	0.5151
17X																	0.4

(b) The measured zooming time

Fig. 6. The measured movement times for panning, tilting and zooming of AXIS 214 PTZ camera.

The predicted location is iteratively determined with the predetermined zooming level to consider the target movement during T_z. Once the next size of a motion box is determined by (4) with $T_a = 0$ and $T_z = 0$, T_z is calculated with the given specifications of a camera to zoom in an image by the new size of the motion box. The predicted location of a motion box is recalculated by considering the target movement during T_z. Then, T_a is updated with the recalculated location of a motion box. The method to determine the location and the size of a motion box is summarized in Algorithm 1.

Algorithm 1: The compensation of $\Delta T_a + \Delta T_z$

Input : $O(t+T_e)$, $V(t+T_e)$, N, the movement time of a camera

Output: z', $M(t+T_e+T_t+T_z)$

1. Determine z' with $O(t+T_e)$, $V(t+T_e)$ and N under the assumptions $T_a=0$ and $T_z=0$ by (2), (3) and (4);
2. Calculate T_z with z' according to a camera movement time;
3. Estimate $\hat{O}(t+T_e+T_z)$ and $\hat{O}(t+T_e+T_z+T_N)$;
4. Determine $M(t+T_e+T_z)$ with $\hat{O}(t+T_e+T_z)$ and $\hat{O}(t+T_e+T_z+T_N)$ by (2);
5. Calculate T_a with $M(t+T_e+T_z)$ according to a camera movement time;
6. Estimate $\hat{O}(t+T_e+T_a+T_z)$ and $\hat{O}(t+T_e+T_a+T_z+T_N)$;
7. Determine $M(t+T_e+T_a+T_z)$ with $\hat{O}(t+T_e+T_a+T_z)$ and $\hat{O}(t+T_e+T_a+T_z+T_N)$ by (2);
8. Update T_a with $M(t+T_e+T_a+T_z)$ according to a camera movement time;
9. Repeat 6 to 8 until the amount of $\Delta T_a+\Delta T_z$ is negligible;

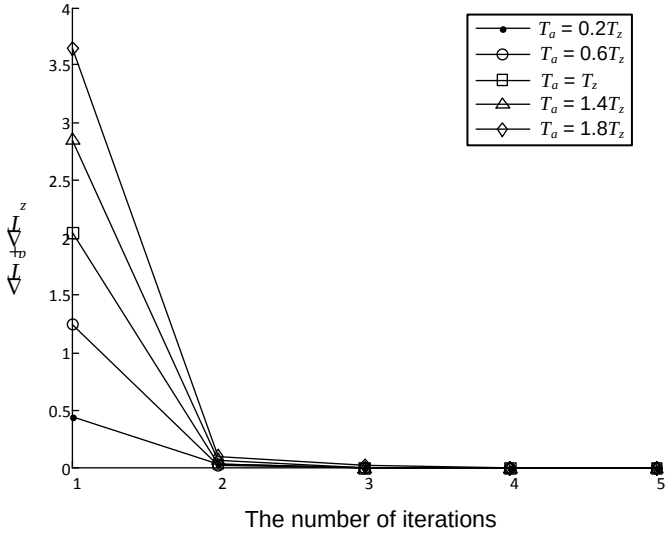


Fig. 7. The amount of $\Delta T_a+\Delta T_z$ by the iteration based determination of a motion box according to T_a when $T_z=2$.

Fig. 7 shows the amount of $\Delta T_a+\Delta T_z$ by the iteration based determination of a motion box according to the number of iterations. The data of **Fig. 6** is used for the simulation. T_z is set to 2 seconds and T_a is changed from $0.2T_a$ to $1.8T_z$ with the step of 0.4 seconds. The frame rate is set to 3 *frame/sec* and N is set to 5. The simulation results demonstrate that the amount of $\Delta T_a+\Delta T_z$ converges on zero with at most three iterations.

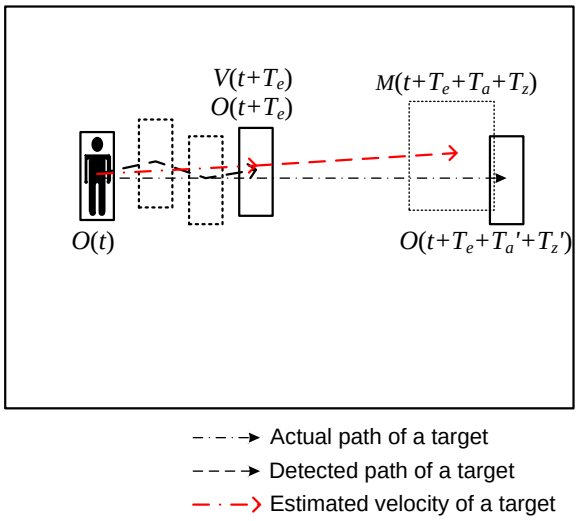


Fig. 8. The possible failure of the single step based active tracking due to the inaccuracy of the estimation and the prediction.

However, the performance of the single step zooming is affected by target velocity and detection variation as shown **Fig. 8**. The predicted location is calculated with the estimated velocity of a target, $V(t+T_e)$. The actual velocity of a target can vary while a camera translates to and zooms in the predicted location. It causes that the prediction location of a motion box $M(t+T_e+T_a+T_z)$ may not include the actual location of a target $O(t+T_e+T_a'+T_z')$. Detection variation during the velocity estimation of a target also affects the accuracy of the predicted location. When the system estimates the velocity of a target, the trajectory history of a target may not be the same as the actual trajectory of a target due to detection variation. It generates the error in the estimated velocity of a target and the inaccurately estimated velocity can cause the predicted location to be largely deviated from the actual location. As a result, the system may fail in zooming in the target with the desired zooming level. Therefore, the system should incorporate the effect of object velocity and detection variation in determining a motion box.

3.2 Effects of Unknown Factors

Object velocity variation creates the estimation error for a moving target and causes the estimated location of a motion box to be deviated from the actual location of a target. Also, the amount of the deviation is usually proportional to the zooming time while a target is not stationary. This leads to increase the size of a motion box to incorporate the deviation effect, and then the zooming level is restricted.

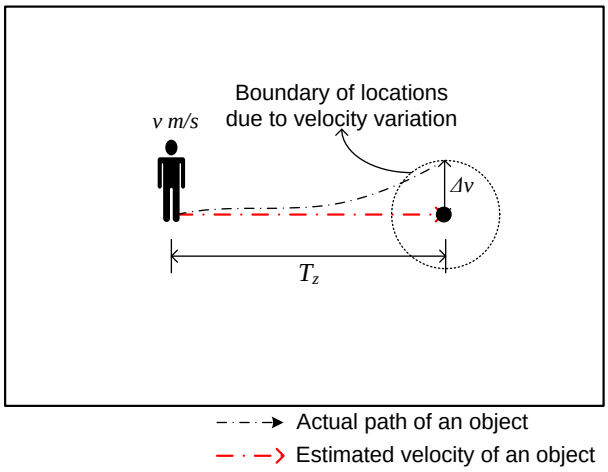
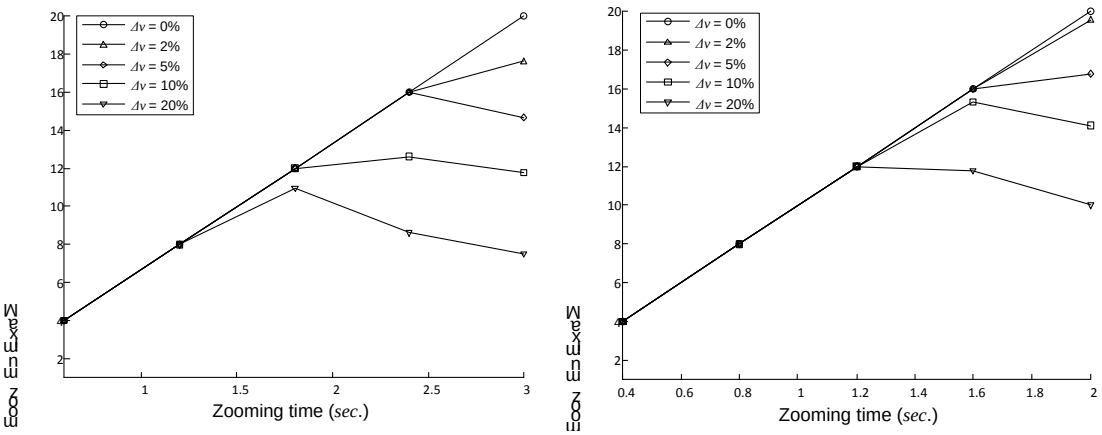
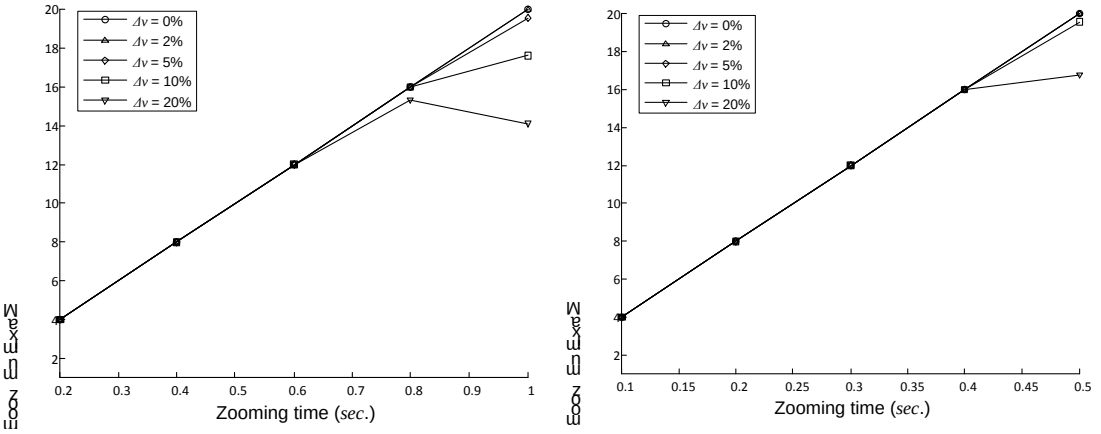


Fig. 9. The illustration of object velocity variation.

Fig. 9 illustrates the effect of object velocity variation with the zooming time. Δv denotes the variation percentage of a target velocity. In order to measure the effect of object velocity variation on the zooming level, the maximum zooming level to incorporate the effect of Δv is measured among the boundary of locations due to velocity variation. A camera is placed at (45*m*, 40*m*, 40*m*) with a top-down perspective observing an object moving from the left to the right at 2*m/s* on the ground plane in the simulation with MATLAB. The initial position of an object is (45*m*, 40*m*). The trajectory of the object is projected onto the image plane with the size of 352×240 by the perspective projection model. The focal length and lens dimension of a camera are set by the specifications of AXIS 214 PTZ camera. The frame rate is set to 5 *frame/sec* considering the detection and tracking time, and the required frames for the velocity estimation are set to $N = 3$.



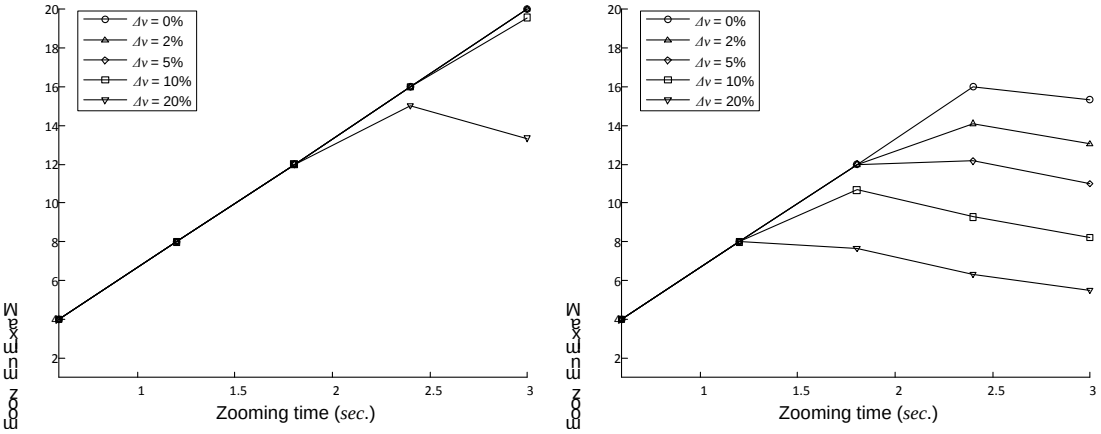
(a) $T_{z_m} = 3sec$ **(b)** $T_{z_m} = 2sec$



(c) $T_{z_m} = 1sec$ **(d)** $T_{z_m} = 0.5sec$

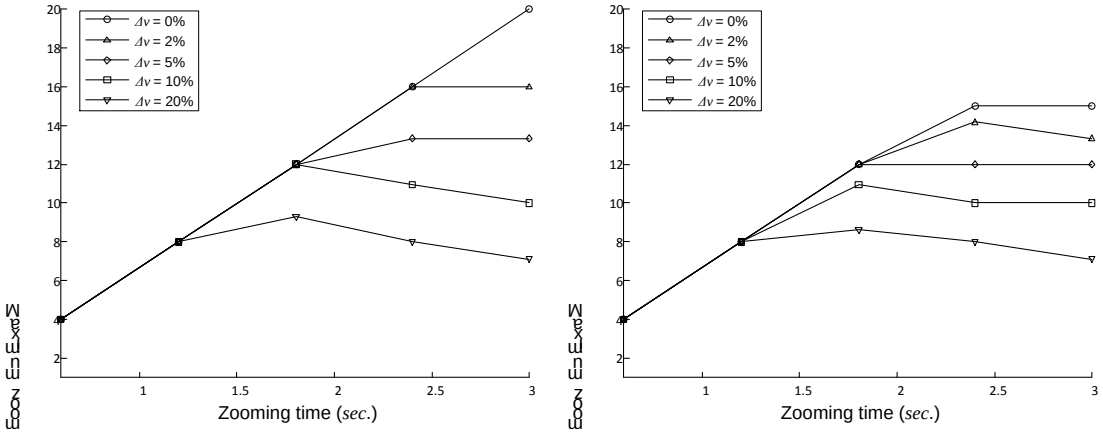
Fig. 10. The variation of the zooming level according to the variation of object velocity variation and the time of maximum zooming level.

The effect of object velocity variation according to the time of maximum zooming level is illustrated in **Fig. 10**. T_{z_m} denotes the time of the maximum zooming level and the line of $\Delta v = 0\%$ represents the original zooming level of a motion box without object velocity variation. It is assumed that the zooming time is linear to zooming level in this analysis for simplicity. However, as the degree of object velocity variation increases, a required zooming level decreases more than the original zooming level. For example of T_{z_m} = 3 seconds, the required zooming level is lower than the original zooming level of a motion box. Then, it may fail in tracking a target because a target can be out of a motion box with the original zooming level. Thus, a system decreases the original zooming level of a motion box as finding the equivalent zooming level of a required zooming level. When object velocity variation is 2%, the maximum zooming level is set to the zooming time of 2.5 *seconds* in the conservative way to prevent the failure of tracking. Also, the simulation results demonstrate that a camera with fast maximum zooming is less sensitive to object velocity variation.



(a) Object velocity $v = 1m/s$ **(b)** Object velocity $v = 3m/s$

Fig. 11. The effect of object velocity variation with T_{z_m} = 3 seconds according to an object velocity when a camera has a top-down perspective.



(a) Tilting angle 30^o at (45*m*, 20*m*, 34.64*m*) **(b)** Tilting angle 45^o at (45*m*, 11.72*m*, 28.28*m*)

Fig. 12. The effect of object velocity variation with $T_{z_m} = 3$ seconds according to a camera perspective when an object moves at $v = 2m/s$.

The effect of object velocity variation also varies according to an object velocity and a camera perspective. **Fig. 11** illustrates the zooming level according to an object velocity. As an object moves faster, the zooming level is more restricted by the required frames for the next velocity estimation. Also, **Fig. 12** illustrates the zooming level according to a camera perspective. As a camera perspective is more tilted, the zooming ratio is more restricted by the size of a target on the image.

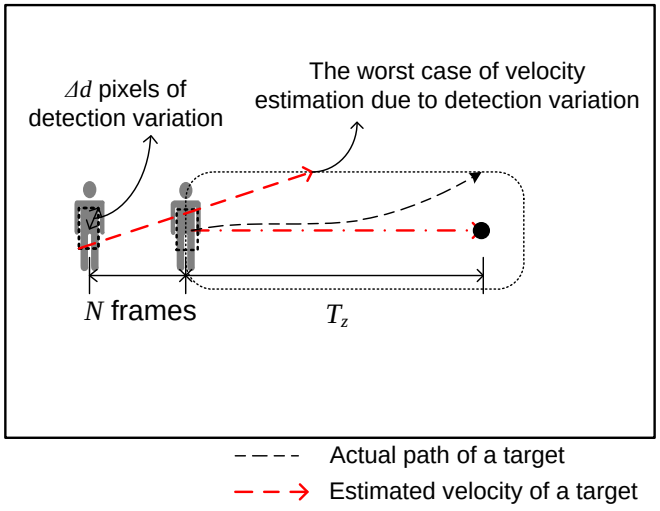
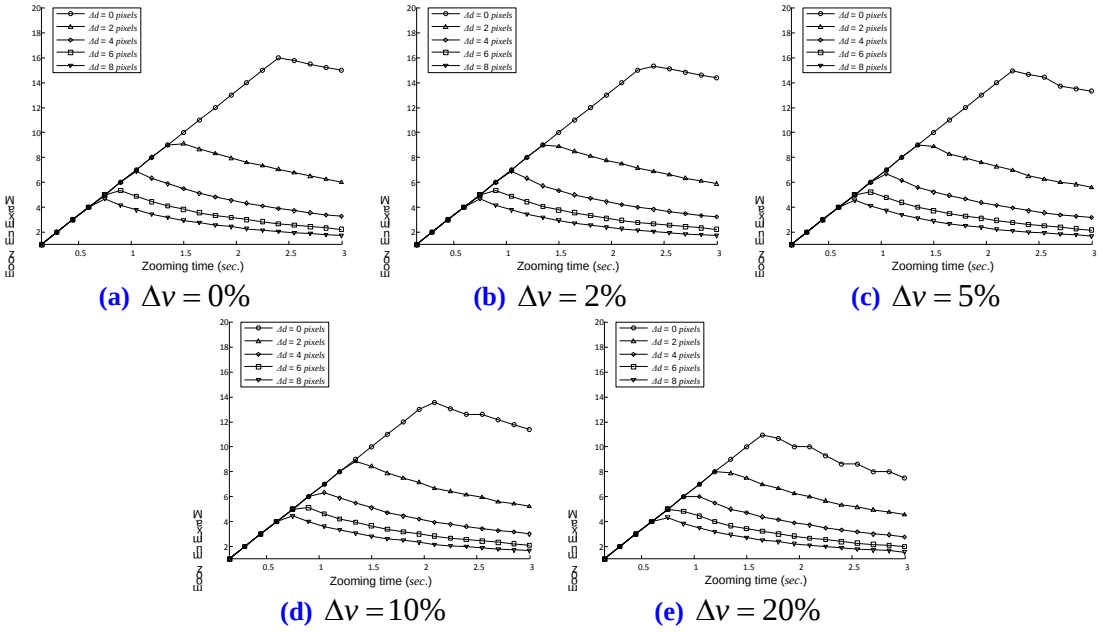


Fig. 13. The illustration of the effect of detection variation on zooming level.

Detection variation during *N* frames also creates the estimation error for an object velocity and it can affect the determination of the zooming ratio as shown in **Fig. 13**. The amount of detection variation is denoted by Δd and it represents the number of pixels from the centroid of a target. The effect of the detection variation can be negligible due to the maximum zooming level determined by considering the object velocity variation. However, the detection variation becomes significant when the maximum zooming level with the object velocity

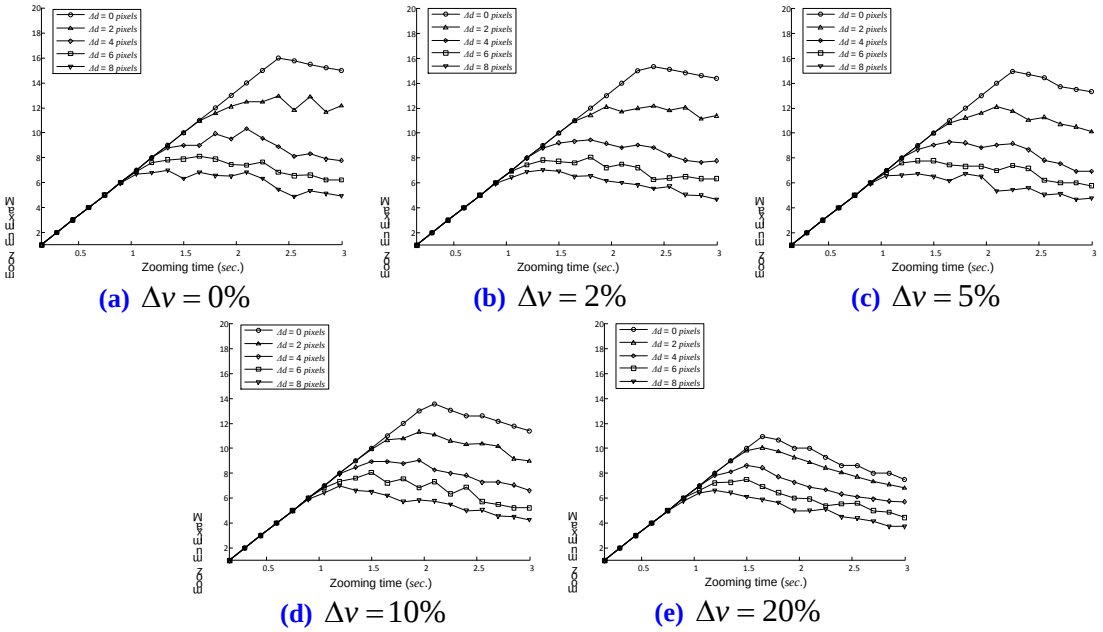
variation cannot incorporate the effect of the detection variation. The worst case of velocity estimation occurs when a target is detected at the boundary of a detection variation in the diagonal way.



(a) $\Delta v = 0\%$ **(b)** $\Delta v = 2\%$ **(c)** $\Delta v = 5\%$

(d) $\Delta v = 10\%$ **(e)** $\Delta v = 20\%$

Fig. 14. The effect of detection variation with T_{z_m} = 3 seconds according to object velocity variation when an object moves at $v = 2m/s$.



(a) $\Delta v = 0\%$ **(b)** $\Delta v = 2\%$ **(c)** $\Delta v = 5\%$

(d) $\Delta v = 10\%$ **(e)** $\Delta v = 20\%$

Fig. 15. The average effect of detection variation with T_{z_m} = 3 seconds according to object velocity variation when an object moves at $v = 2m/s$.

Fig. 14 illustrates the effect of detection variation with the worst case of velocity estimation according to object velocity variation. When $\Delta d = 0$, object velocity variation is a dominant factor on determining the zooming level. Otherwise, the effect of object velocity variation on the zooming ratio is incorporated into the effect of detection variation and the effect of detection variation is a dominant factor on determining the zooming level. **Fig. 15** shows the average effect of detection variation with the uniformly generated detection variation within Δd. The zooming level is slightly improved because the worst case of velocity estimation is hardly occurred during *N* frames.

3.3 Active Tracking with Adaptive Zooming Steps

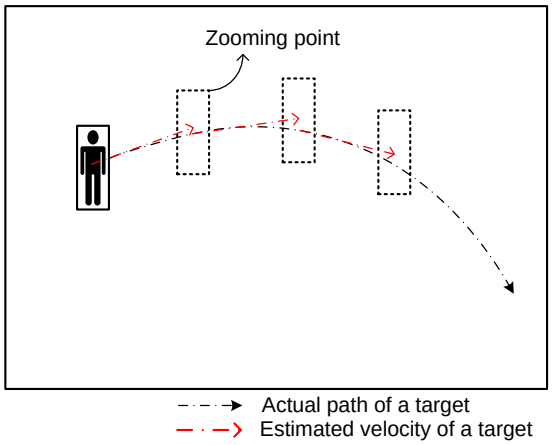


Fig. 16. The illustration of a piece-wise linearized object model.

When the unknown factors of object velocity and detection variation are considered, the long prediction time by one step zooming compromises the zooming level. In order to maximize the zooming level against the unknown factors, a piece-wise linearized object model is used as shown in **Fig. 16**. A piece-wise linearized object model also prevents the failure of active tracking against the object velocity and detection variation. The basic assumption is that an object normally moves at the constant velocity in a surveillance region. It simplifies the model estimation of a target without prior model information and the estimation error of a target is minimized. This model has the benefit of shortening the estimation time of an object model because the complex pattern of an object model can be approximated to the linear segments. Since most of objects usually move straight in a surveillance region or along with paths, it is not critical for active tracking to fail due to the inaccurately linearized model. As an object trajectory is finely divided, the velocity estimation error is minimized. Also, the faster the zooming time of a camera, the more the estimation error during the zooming time is reduced.

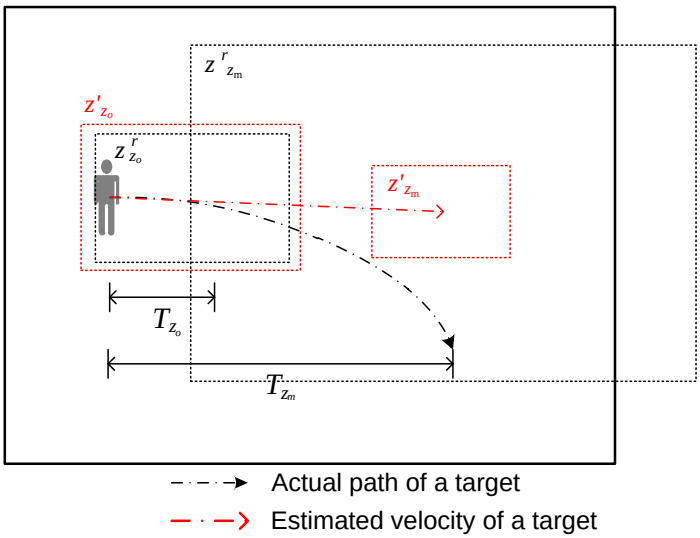
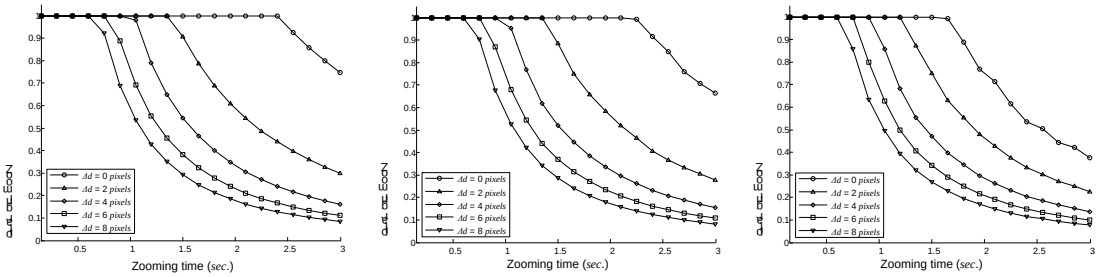


Fig. 17. Illustration of step size determination.

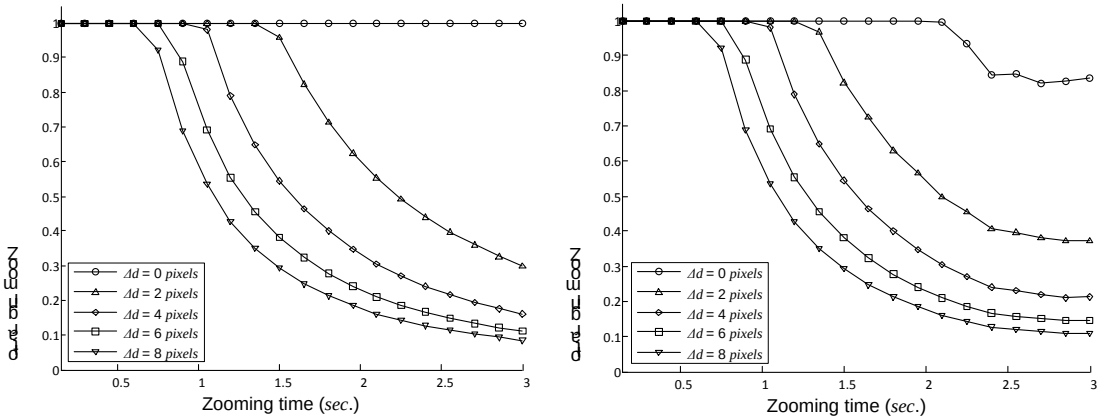
The system can zoom in a targeted object by the fixed number of multiple steps. However, repeated estimation processes for the velocity of a targeted object affect the performance of the multiple steps based active tracking. Especially, when a targeted object moves slowly, the multiple steps may not be necessary. The one extreme case is that target speed is relatively faster than the camera movement speed. Because the inaccuracy of the predicted zooming location increases during the slow camera movement, it restricts the zooming level to compromise the inaccuracy of the predicted location at each step and requires multiple steps. On the other hand, when target speed is relatively slower than the camera movement speed, the fast camera movement shortens the amount of prediction time and reduces the inaccuracy of the predicted location. The proposed method achieves the maximum zooming level by a few steps.

The number of zooming steps is determined by considering the effect of the object velocity and detection variation to adaptively cope with unknown object dynamics. The zooming level of a motion box is calculated by (4). However, the required zooming level to incorporate the effect of them is affected by the zooming time (i.e., T_z) as shown in **Fig. 14**. As a system sets the higher zooming level initially, the effect of them also increases due to the increased zooming time and the possible zooming level becomes smaller than the initial zooming level as illustrated in **Fig. 17**. z'_{z_m} denotes the calculated zooming level of a motion box with zooming time T_{z_m} and $z^{r}_{z_m}$ denotes the required zooming level to incorporate the detection variation and the velocity variation. In order to track a target successfully, z'_{z_m} is revised in the conservative way to incorporate the effect of them. $z^{r}_{z_o}$ denotes the adaptively determined zooming level with zooming time T_{z_o} . Therefore, the system adjusts the size of a motion box at each step by considering the object velocity and detection variation.



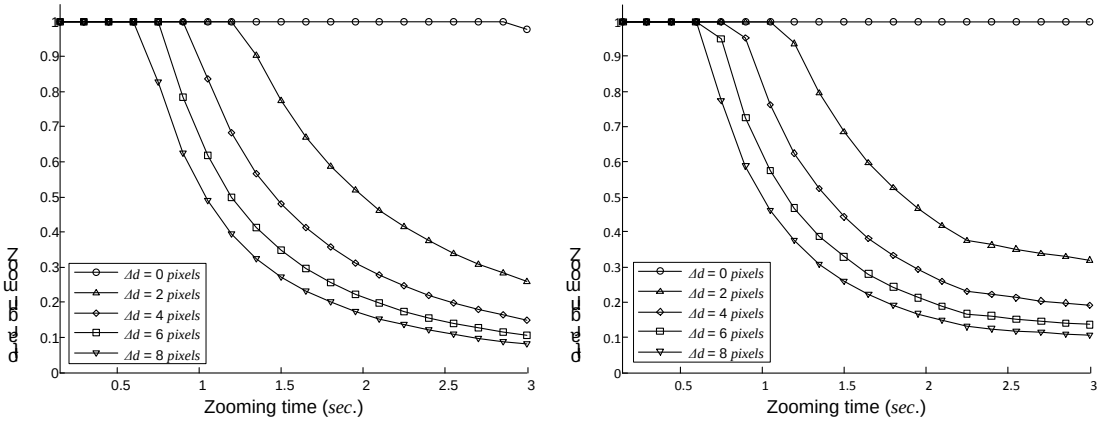
(a) $\Delta v = 0\%$ **(b)** $\Delta v = 5\%$ **(c)** $\Delta v = 10\%$

Fig. 18. The required zooming ratio of the required zooming level to the original zooming level to compensate for detection variation according to object velocity variation when T_{z_m} = 3 seconds and $v = 2m/s$.



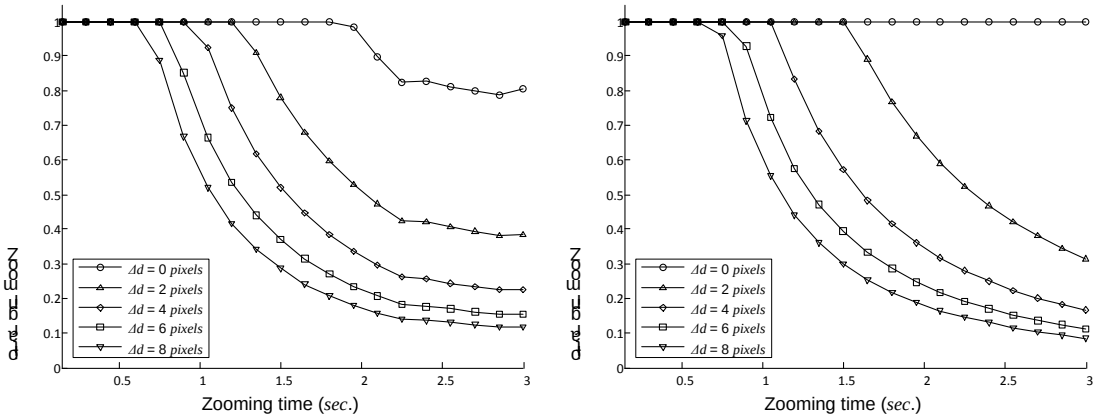
(a) $v = 1m/s$ **(b)** $v = 3m/s$

Fig. 19. The required zooming ratio of the required zooming level to the original zooming level to compensate for detection variation according to object velocity when T_{z_m} = 3 seconds and $\Delta v = 0\%$.



(a) Tilting angle 30^o at (45*m*, 20*m*, 34.64*m*) **(b)** Tilting angle 45^o at (45*m*, 11.72*m*, 28.28*m*)

Fig. 20. The required zooming ratio of the required zooming level to the original zooming level to compensate for detection variation according to camera perspective when T_{z_m} = 3 seconds, $\Delta v = 0\%$ and $v = 2m/s$.



(a) Camera at (45*m*, 40*m*, 30*m*) **(b)** Camera at (45*m*, 40*m*, 50*m*)

Fig. 21. The required zooming ratio of the required zooming level to the motion box to compensate for detection variation according to the distance between an object and a camera when T_{z_m} = 3 seconds, $\Delta v = 0\%$ and $v = 2m/s$.

Fig. 18 shows the required zooming ratio of the required zooming level to the original zooming level of the motion box to compensate for object velocity and detection variation. The same simulation setup to the previous section is used. The original zooming level is obtained by the simulation when $\Delta v = 0$ and $\Delta d = 0$. The ratio is affected by Δv when $\Delta d = 0$. Otherwise, the simulation results show the similar ratios for three different Δv when $\Delta d \neq 0$. It indicates that the ratio is highly dependent on detection variation. Also, **Fig. 19**, **Fig. 20** and **Fig. 21** show the simulation results for the ratio according to object velocity, camera perspective and the distance between an object and a camera respectively. They demonstrate that the patterns of the ratios are similar to each other and they are more dependent on detection variation.

A system determines $z^r_{z_o}$ according to the expected detection variation and object velocity variation. $z^r_{z_o}$ is obtained by multiplying the required zooming ratio to $z^r_{z_m}$ and the corresponding zooming time for this step is T_{z_o}. When the ratio is 1, the required zooming level is equal to the zooming level of a motion box. The termination of adaptive zooming is triggered by the desired zooming level or the exit of a targeted object from an image. z_d denotes the desired zooming level given to the system. If the new zooming level is equal to or less than z_d, the center coordinate and the size of a motion box are determined by considering only the next predicted location of a targeted object without the consideration of the next velocity estimation. Another termination condition is to use the difference between the previous zooming level and the new zooming level. In practice, the zooming process for the small change in zooming level is unnecessary. When the zooming process terminates, the system returns to the sector based scanning and selection to track other objects.

4. Integration and Evaluation

4.1 System Integration

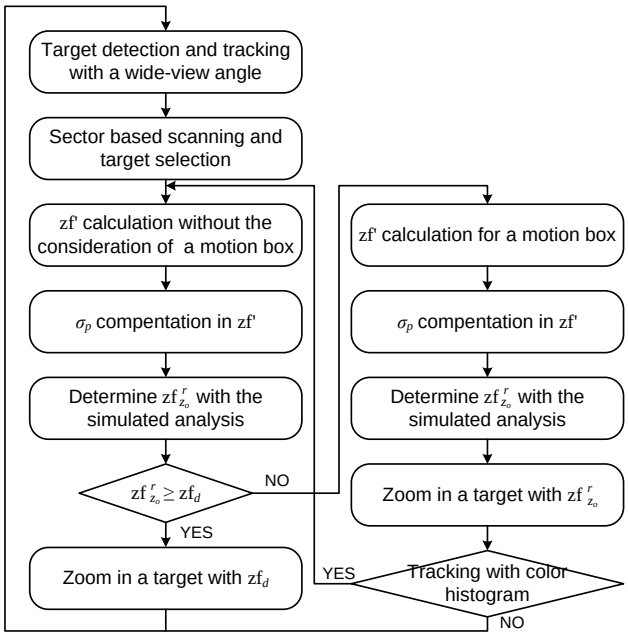


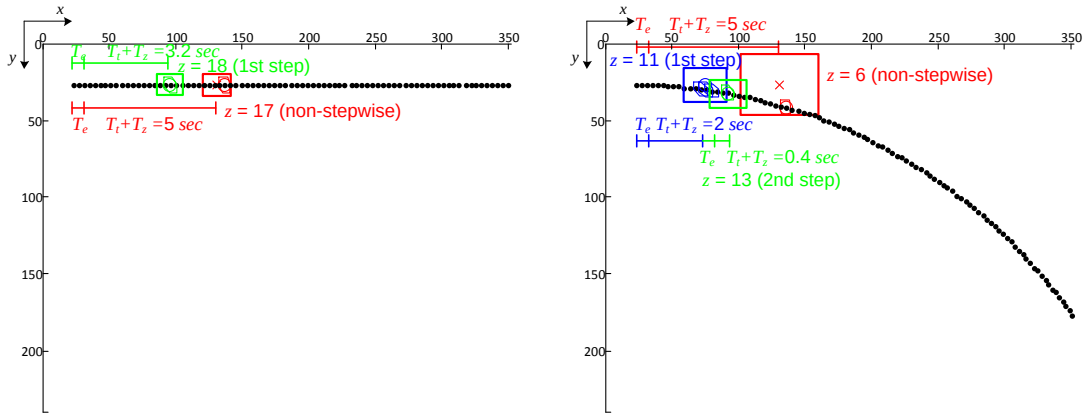
Fig. 22. The illustration of the flow chart of the proposed active tracking system.

The processing flow of the proposed active tracking system is illustrated in **Fig. 22**. The initial information is set in the system, such as the camera information for the normal state, the sector information, the desired maximum zooming level, and the expected velocity and detection variation. The system begins the background modeling for target detection. Once the background information is stabilized, the system detects and tracks targets. After one target is selected by the sector based scanning, the initial zooming level and location without considering a motion box are calculated by estimated target velocity on an image. Then, their values are recalculated to compensate for the effect of prediction error, velocity variation, and detection variation. If the revised zooming level is higher than or equal to the desired zooming level, the system zooms in the selected target at the desired zooming level and returns to the normal state. Otherwise, the zooming level and location for a motion box are calculated. Their values are also recalculated to compensate for the effect of prediction error, velocity variation, and detection variation. After the system zooms in the selected target with the calculated zooming level, it tracks the selected target by using color histogram. Then, the system repeats the zooming process for the selected target until the calculated zooming level is satisfied with the desired zooming level or it fails in tracking the selected target with color histogram. Once the calculated zooming level reaches the desired zooming level, the system also returns to the normal state.

4.2 Algorithm Evaluation

In order to highlight the advantage of the proposed method, the determined zooming level with a finite amount of zooming time is compared with the simple non-stepwise zooming by the simulation. The frame rate is set to 5 *frames/sec* and object velocity is estimated with 3 *frames* in the simulation. The zooming level with the simple non-stepwise zooming is determined by calculating the minimum box including the target when the allowed tracking time is fully utilized. A camera is placed at ($x = 0m$, $y = 40m$, $z = 40m$) with a top-down perspective for the worst case of object velocity variation. The maximum zooming level is set to 20 in 3 *sec* and

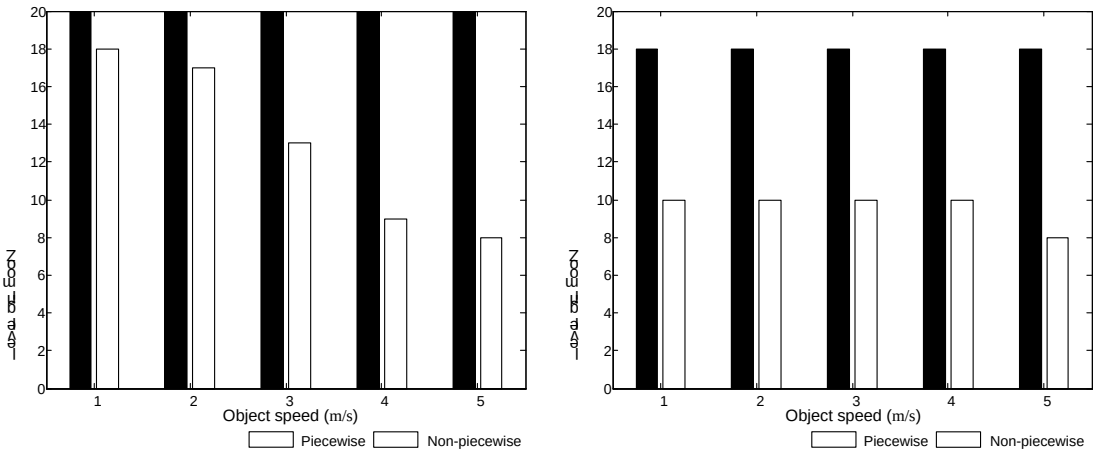
the time characteristic for zooming is assumed linear. **Fig. 23** shows the simulation results on an image plane according to object velocity variation when detection variation is zero. The x-axis and y-axis on the image plane follow the conventional image origin (i.e. top-left origin). It demonstrates the proposed method achieves higher zooming level in shorter time than the simple non-stepwise zooming. Also, the amount of zooming time is not proportional to the zooming level. Although sophisticated estimation techniques are used with given zooming time, the zooming level is restricted by object velocity variation.



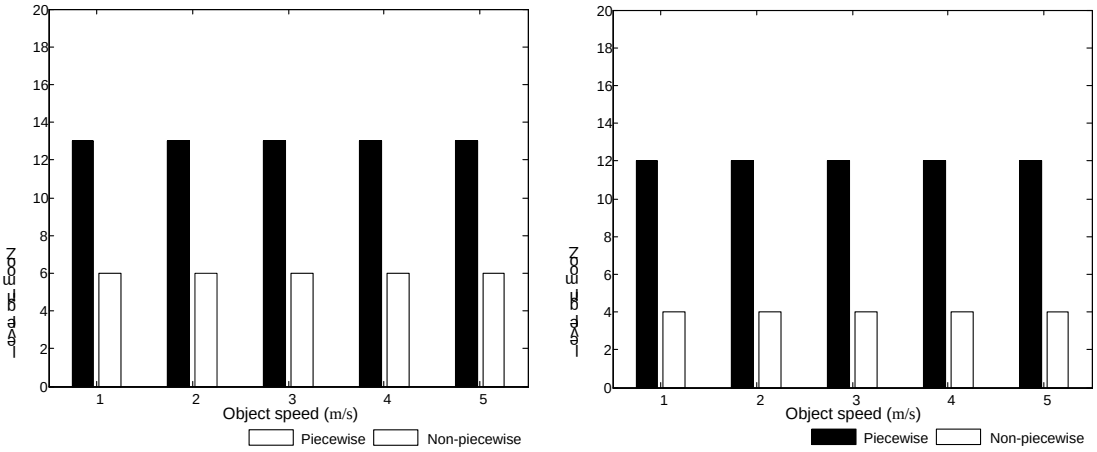
(a) $\Delta v = 0\%$ **(b)** $\Delta v = 10\%$

Fig. 23. The performance comparison in terms of zooming level when allowed tracking time is 5 sec and object speed is 2 m/s with $\Delta d = 0\, pixels$.

Fig. 24 shows the comparison with the simple non-stepwise zooming according to object speed and velocity variation when the detection variation is zero. The maximum zooming level for each velocity variation is obtained from **Fig. 14**. The simulation results show that the proposed method achieves the maximum zooming level regardless of object speed. It demonstrates that the proposed method tracks the target with the adaptive zooming steps even for fast moving targets. On the other hand, the simple non-stepwise zooming is severely affected by object speed and velocity variation.



(a) $\Delta v = 0\%$ **(b)** $\Delta v = 5\%$



(c) $\Delta v = 10\%$ **(d)** $\Delta v = 15\%$

Fig. 24. The performance comparison of stepwise and non-stepwise approach according to object speed and velocity variation when allowed tracking time is 5 sec with $\Delta d = 0\, pixels$.

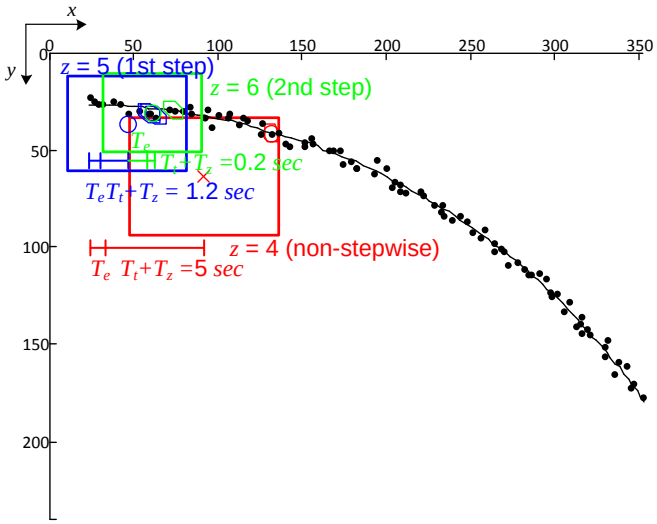
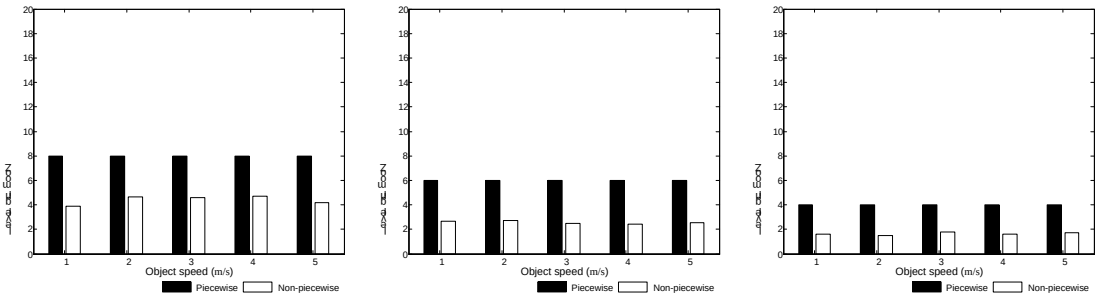


Fig. 25. The performance comparison in terms of zooming level when allowed tracking time is 5 sec and object speed is 2 m/s with $\Delta d = 4\, pixels$.



(a) $\Delta d = 2\, pixels$ **(b)** $\Delta d = 4\, pixels$ **(c)** $\Delta d = 8\, pixels$

Fig. 26. The performance comparison of stepwise and non-stepwise approach according to object speed and detection variation when allowed tracking time is 5 sec with $\Delta v = 10\%$.

Fig. 25 and **Fig. 26** show the comparison with the simple non-stepwise zooming when the detection variation is incorporated. In **Fig. 25**, the black solid line indicates the original target trajectory without the detection variation and the black dots indicate the detected location with the detection variation. The amount of the detection variation is randomly generated within the specified variation. While the proposed method achieves the maximum zooming for each detection variation, the simple non-stepwise zooming is severely affected by the effect of detection variation.

4.3 Experiment

A camera (i.e. C_1) is installed to observe the surveillance region about 50m away from objects in order to verify the proposed method. An extra camera (i.e. C_2) continuously observes the entire view of the surveillance region to show the trajectories of objects for the verification. The required zooming ratio of a motion box is obtained by simulating the effect of detection and velocity variation as described in the previous section. For the simulation parameters, T_{z_m} is set to 3 seconds and the frame rate is set to 5 *frame/sec* considering the specification of the used an AXIS 214 PTZ camera and a PC Pentium(R) D CPU 2.80 GHz with 3.5GB of RAM. Also, Δv is set to 5% and Δd is set to 4 *pixels* considering the path of the surveillance region and the detection performance. **Table 1** shows the required zooming level of a motion box by considering the effect of the specified detection and velocity variation. The table indicates the maximum zooming level of a motion box is limited to 10 due to the effect of them. If the calculated size of a motion box is greater than 10, it is set to 10 to avoid the failure of the active tracking. Thus, the number of sectors on an image is set to 3×3 with the image size of 352×240.

Table 1. The required zooming level of a motion box by the simulation when Δv is set to 5% and Δd is set to 4 *pixels*.

Zooming level by (2) and (4)	Required zooming ratio	Required zooming level
1	1	1
2	1	2
3	1	3
4	1	4
5	1	5
6	1	6
7	1	7
8	1	8
9	0.95	8
10	0.8	8
11	0.67	7
12	0.6	7

Fig. 27. Simulation results of the proposed method (the first, third, and fifth row of frames are captured by camera C_2 and the others by camera C_1).

Fig. 27 shows the simulation results of the proposed method for the active tracking. Frame numbers in the figure are counted in terms of camera C_2. Each blue rectangle in images by camera C_2 indicates a corresponding target which is selected for the active tracking in camera C_1. At frame #1509, two targets are initially detected and one target is selected by the scanning strategy. Once the velocity of the selected target is estimated, the zooming level of a motion box is calculated by (2) and (4). The initial zooming level of the motion box is 7 and it is the same as the required zooming level by **Table 1**. Because the location and size of the motion box are calculated by considering the next step zooming, the selected target is zoomed in and captured at the boundary of the image as shown in frame #1552. After the selected target is tracked by using the color histogram and the velocity of the selected target is estimated in the zoomed image, the next zooming level of a motion box is calculated. The calculated zooming level is 14 and is higher than the maximum zooming level, the next motion box does not need to include the space for the next step zooming. Thus, the zooming level is bounded to the maximum zooming level of 10 and the required zooming level is set to 8 by

Table 1 and it is zoomed in again at the maximum zooming level of 10. Frame #1584 shows the zoomed target with the maximum zooming level. For frames #1838 and #1920, the calculated zooming levels for the selected targets are higher than the maximum zooming level and they are zoomed in at the maximum zooming level at once as shown in frames #1863 and #1991 respectively. Frames #2183, #2220 and #2228 show another case of two-step zooming. Due to the size of a selected target, the first zooming level is determined to be 4. After the selected target is tracked by the color histogram, the second zooming level is determined to be 8. Frames #3381, #3396 and #3404 show the incorrect prediction of the location because the selected target changes the direction suddenly. However, the selected target is included in the zoomed image because the changed direction does not deviate from the expected velocity variation. The last two frames #3827 and #3883 also show the successfully zoomed target at the maximum zooming level at once.

5. Conclusions

The paper presents the single PTZ camera based active tracking system with the sector based scanning. The time it takes to track each object is minimized to increase the support for multiple targets. The system tracks the target trajectory with the piecewise linearized object model against unknown object dynamics. The fast trajectory estimation copes with fast moving targets and/or slow camera movements. The required zooming level at each zooming step is adaptively determined by incorporating the effect of the object velocity and detection variation. The analysis including the effect of the both variations indicates the detection variation is dominant factor in determining the zooming ratio and it is more important to minimize the detection variation when estimating the zooming location. The real-time experiments prove the effectiveness of the proposed method for the expected object velocity and detection variation.

For future work, the sector based scanning will be improved by learning the pattern of object trajectories. Since the proposed method gives the equal priority for each sector, it may scan unnecessary sectors where targets do not often appear. The pattern of object trajectories can be used to set the different priorities for sectors. In addition, it can be used to cope with unexpected object velocity variation at sectors having the corners of paths. Moreover, the proposed method can be extended to the active tracking system with multiple PTZ cameras to increase the coverage of the surveillance system and the number of tracked targets.

References

[1] R. T. Collins, O. Amidi and T. Kanade, “An active camera system for acquiring multi-view video,” in *Proc. of Int’l Conf. on Image Processing*, pp.517-520, 2002. Article (CrossRef Link).

[2] P. Kumar, A. Dick and T. S. Sheng, “Real time target tracking with pan tilt zoom camera,” *Digital Image Computing: Techniques and Applications*, pp.492-497, 2009. Article (CrossRef Link).

[3] C. J. Costello, C. P. Diehl, A. Banerjee and H. Fisher, “Scheduling an active camera to observe people,” in *VSSN '04: Proc. of the ACM 2nd Int’l Workshop on Video Surveillance & Sensor Networks*, pp.39-45, 2004. Article (CrossRef Link).

[4] C. J. Costello and I.-J. Wang, “Surveillance camera coordination through distributed scheduling,” in *Proc. of IEEE Conf. on Decision and Control*, pp.1485-1490, 2005. Article (CrossRef Link).

[5] F. Z. Qureshi and D. Terzopoulos, “Smart camera networks in virtual reality,” in *Proc. of the IEEE*, vo.96, no.10, pp.1640-1656, 2008. Article (CrossRef Link).

[6] Y. Xu and D. Song, "Systems and algorithms for autonomous and scalable crowd surveillance using robotic PTZ cameras assisted by a wide-angle camera," *Autonomous Robots*, vo.29, no.1, pp.53-66, 2010. Article (CrossRef Link).
[7] A. Cretual, F. Chaumette and P. Bouthemy, "Complex object tracking by visual servoing based on 2d image motion," in *Proc. of Int. Conf. on Pattern Recognition*, vo.2, pp.1251-1254, 1998. Article (CrossRef Link).
[8] K. Daniilidis, C. Krauss, M. Hansen and G. Sommer, "Real-time tracking of moving objects with an active camera," *Real-Time Imaging*, vo.4, no.1, pp.3-20, 1998. Article (CrossRef Link).
[9] R. T. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, "Algorithms for cooperative multi-sensor surveillance," in *Proc. of the IEEE*, vo.89, pp.1456-1477, 2001. Article (CrossRef Link).
[10] X. Zhou, R. T. Collins, T. Kanade and P. Metes, "A master-slave system to acquire biometric imagery of humans at distance," in *Proc. of ACM SIGMM Int'l Workshop on Video Surveillance*, pp.113-120, 2003. Article (CrossRef Link).
[11] M. Lalonde, S. Foucher, L. Gagnon, E. Pronovost, M. Derenne and A. Janelle, "A system to automatically track humans and vehicles with a PTZ camera," *SPIE*, vo. 6575, 2007. Article (CrossRef Link).
[12] J. Davis and X. Chen, "Calibrating pan-tilt cameras in wide-area surveillance networks," in *Proc. of IEEE Int'l Conf. on Computer Vision*, pp.144-149, 2003. Article (CrossRef Link).
[13] U. M. Erdem and S. Sclaroof, "Look there! predicting where to look for motion in an active camera network," in *Proc. of IEEE Conf. on Advanced Video and Signal Based Surveillance*, pp.105-110, 2005. Article (CrossRef Link).
[14] F. Z. Qureshi and D. Terzopoulos, "Surveillance in virtual reality: system design and multi-camera control," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp.1-8, 2007. Article (CrossRef Link).
[15] D. Rother, K. A. Patwardhan and G. Sapiro, "What can casual walkers tell us about a 3D scene?," in *Proc. of IEEE Int'l Conf. on Computer Vision*, pp.1-7, 2007. Article (CrossRef Link).
[16] J. Denzler, M. Zobel and H. Niemann, "Information theoretic focal length selection for real-time active 3D object tracking," in *Proc. of IEEE Int'l Conf. on Computer Vision*, pp.400-407, 2003. Article (CrossRef Link).
[17] B. Tordoff and D. Murray, "A method of reactive zoom control from uncertainty in tracking," *Computer Vision and Image Understanding*, vo.105, no.2, pp.131-144, 2007. Article (CrossRef Link).
[18] E. Sommerlade and I. Reid, "Information-theoretic active scene exploration," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2008. Article (CrossRef Link).
[19] AXIS 214 PTZ Network Camera, http://www.axis.com/products/cam_214/
[20] N. Dalal, B. Triggs and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *Proc. of European Conf. on Computer Vision*, vo.3952, pp.428-441, 2006. Article (CrossRef Link).
[21] A. Monnet, A. Mittal, N. Paragios and V. Ramesh, "Background modeling and subtraction of dynamic scenes," in *Proc. of IEEE Int'l Conf. on Computer Vision*, pp.1305-1312, 2003. Article (CrossRef Link).
[22] C. Stauffer and W. Grimson, "Learning patterns of activity using real time tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vo. 22, no. 8, pp. 7647-757, 2000. Article (CrossRef Link).

Shung Han Cho received B.E. degree (Summa Cum Laude) with specialization in Telecommunications from both the department of Electronics Engineering at Ajou University, Korea and the department of Electrical and Computer Engineering at Stony Brook University - SUNY, NY in 2006. He was a recipient of Award for Academic Excellence in Electrical Engineering by College of Engineering and Applied Sciences at Stony Brook University. He received M.S. degree with Award of Honor in Recognition of Outstanding Achievement and Dedication and Ph.D. degree in Electrical and Computer Engineering from Stony Brook University in 2008 and 2010 respectively. He is currently a post-doctoral researcher at Stony Brook University. He was a recipient for International Academic Exchange Program supported by Korea Research Foundation (KRF) in 2005. He was a member of Sensor Consortium for Security and Medical Sensor Systems sponsored by NSF Partnerships for Innovation from 2005 to 2006. His research interests include collaborative heterogeneous signal processing, distributed digital image processing and communication, networked robot navigation and communication, heterogeneous system modeling.

Yunyoung Nam received B.S, M.S. and Ph.D. degree in computer engineering from Ajou University, Korea in 2001, 2003, and 2007 respectively. He was a research engineer in the Center of Excellence in Ubiquitous System from 2007 to 2009. He was a post-doctoral researcher at Stony Brook University in 2009, New York. He is currently a research professor in Ajou University in Korea. He also spent time as a visiting scholar at Center of Excellence for Wireless & Information Technology (CEWIT), Stony Brook University - State University of New York Stony Brook, New York. His research interests include multimedia database, ubiquitous computing, image processing, pattern recognition, context-awareness, conflict resolution, wearable computing, and intelligent video surveillance.

Sangjin Hong received the B.S and M.S degrees in EECS from the University of California, Berkeley. He received his Ph.D in EECS from the University of Michigan, Ann Arbor. He is currently with the department of Electrical and Computer Engineering at Stony Brook University. Before joining Stony Brook University, he has worked at Ford Aerospace Corp. Computer Systems Division as a systems engineer. He also worked at Samsung Electronics in Korea as a technical consultant. His current research interests are in the areas of multimedia wireless communications and digital signal processing systems, reconfigurable VLSI Systems and optimization. Prof. Hong is a Senior Member of IEEE and a member of EURASIP journal editorial board. Prof. Hong served on numerous Technical Program Committees for IEEE conferences.

Weduke Cho received the B.S. in 1981 from Sogang University in Seoul, South Korea, and his M.S. and Ph.D. from Korea Advanced Institute of Science and Technology (KAIST) in 1983 and 1987. He had many actual industrial experiences of large scaled national projects for LG Electronics(CDMA system, Speech Vocoder), KAITECH (HDTV System), and KETI (Smart DTV, Home Server, Internet Phone System, etc) during from 1987 to 2002. Currently he is a professor of department of Electronics Engineering College of Information Technology at Ajou University in Korea, Project Manager of "Ubiquitous computing and networking (UCN) project (www.ucn.re.kr)", and president of ubiquitous convergence research institute (UCRi). His research interests included Smart Convergence Service System and Device Design for "Life-care" and "public-safety" applications on Ubiquitous Computing Environment of Smart Space, and System Architecture Design. Specifically, He is developing a human life-style pattern sensing system with life-log framework, smart bed, actively tracking system for moving target image on CCTV system.