# Locally Initiating Line-Based Object Association in Large Scale Multiple Cameras Environment

**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 April 6, 2010; revised May 25, 2010; accepted June 18, 2010; published June 30, 2010*

## Abstract

Multiple object association is an important capability in visual surveillance system with multiple cameras. In this paper, we introduce locally initiating line-based object association with the parallel projection camera model, which can be applicable to the situation without the common (ground) plane. The parallel projection camera model supports the camera movement (i.e., panning, tilting and zooming) by using the simple table based compensation for non-ideal camera parameters. We propose the threshold distance based homographic line generation algorithm. This takes account of uncertain parameters such as transformation error, height uncertainty of objects and synchronization issue between cameras. Thus, the proposed algorithm associates multiple objects on demand in the surveillance system where the camera movement dynamically changes. We verify the proposed method with actual image frames. Finally, we discuss the strategy to improve the association performance by using the temporal and spatial redundancy.

***Keywords:*** Multiple object association, parallel projection model, homographic line, multiple cameras

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), the Korean government, as a result of UCN's subproject 10C2-T3-10M.

**DOI: 10.3837/tiis.2010.06.010**

## 1. Introduction

The multiple cameras based surveillance system is one of the most interesting research areas in recent years [1][2][3][4][5][4][7][8][9][10][11][12][13]. The purpose of the system is to identify and track objects in the surveillance area. When only the single camera is used for tracking, the performance is affected by the problems from the limited view such as occlusion and overlapping. On the other hand, the multiple cameras can provide the redundant view to resolve these problems. When the view of one camera is interfered with the obstacle or the object, the other camera may observe the tracked objects. However, the multiple cameras based surveillance system also has other problems such as the object handover between cameras and the correspondence between cameras.

When multiple cameras operate collaboratively, the detected targets in multiple cameras should be associated for the consistent and reliable tracking. The various methods are investigated for multiple objects association. In feature based matching method, the features such as shape, motion or color are used to find the corresponding targets [6][7][9][10][11]. However, these methods do not guarantee the association performance where objects have similar features. Some authors use both geometric information and visual information to find correspondence between cameras [12][13]. Also, there is the approach that uses the multiple features at the same time [8]. The common problem for the feature-based methods is to utilize the probabilistic approach. This may cause the false association and the tracking becomes unreliable. When the geometry information is available, the globally defined homographic lines are used to find the corresponding targets [4][5]. However, this approach requires the camera calibration or the training process before the system operates. Moreover, the calibration of surrounding is necessary whenever the camera movement changes.

A reliable association mechanism is necessary for maintaining the consistent tracking information in the large scale multiple cameras environment. The association needs to be performed whenever the system requires to establish or confirm the object association. Also, the system should support the dynamic change of the camera configuration for broadening the effective tracking coverage. However, the conventional camera model requires the calibration whenever the camera configuration changes for accuracy. This is not appropriate for the automatic surveillance where the change of camera configuration frequently occurs. Therefore, the large scale tracking system requires the association method supporting the flexible camera configuration on demand.

In this paper, the locally initiating line-based object association method is presented so that association is established on demand. In order to avoid the calibration whenever the camera configuration changes, the parallel projection model is used for generating and transforming a homographic line [14][15]. We investigate plausible non-ideal parameters (i.e, transformation error, height uncertainty and asynchronous problem) affecting the association performance. With this analysis, the threshold distance between the detected targets is defined to indicate the effectiveness of a locally generated homographic line on the other cameras. The threshold distance based line generation algorithm is presented for multiple objects association. We verify the proposed method with actual image frames. Finally, we discuss the strategy to improve the association performance by using the temporal and spatial redundancy.

The remainder of this paper has 4 sections. In Section 2, we present overview of the application model of visual surveillance system with multiple cameras. Section 3 presents an

association method using local initiated homographic lines and investigates the non-ideal parameters affecting the association. In Section 4, we verify the proposed method with the actual image frames and discuss the strategy to improve the association performance by using the temporal and spatial redundancy. Finally, our contribution is summarized along with future work in Section 5.

## 2. Application Model and Problem Description

### 2.1 Application Model

When more than two cameras detect objects, the system obtains the redundant information. The benefit of multiple cameras is to complement each other by using the redundancy. When multiple cameras are used for the surveillance, it improves the occlusion situation arisen in the single camera and broadens the field of view of tracking. In order to consistently track objects, it is critical to find corresponding targets among cameras (i.e., multiple objects association among multiple cameras). Currently, the human operator analyzes and maintains the information. However, the efficiency and accuracy are restricted by the number of cameras. Thus, the association method supporting multiple cameras is required for the automated surveillance system.

The aim of the proposed method is to support the situation that the common ground plane may not be shown on all the cameras and association targets are the faces of objects. The conventional association methods construct fundamental matrix or homography matrix with the predetermined corresponding points or known references. In a normal situation, the ground plane is usually shown to all the cameras and the homography matrix is easily constructed on the ground plane. Once a homography matrix is constructed on the common (ground) plane, the correspondence of objects can be easily found by transforming points with the homography matrix. However, it is not always guaranteed that the common plane is shown to all the cameras in the surveillance system. Even though they share the common plane, it is difficult for the system to automatically find corresponding points for constructing the homography matrix. Thus, the association methods using the fundamental matrix or homography matrix are not appropriate for an autonomous surveillance system with dynamically moving cameras.

### 2.2 Related Work

There are two approaches to find the corresponding targets among multiple cameras. One is feature based and another is geometric based approaches. Researchers have used various features such as color, histogram, height and motion for feature based approaches. However, all of them are not the unique characteristics for objects. Although some features are used together to improve accuracy, they still rely on the probabilistic approach. Their performance is severely affected by objects having the similar features [6][10][11]. For instance, when people wear the similar uniform or the cloth which has the different color at the front and the back, it is hard to construct the color-based information and associate objects. Also, the accurate feature extraction requires the prohibitively large computation. In [16], the principal axis of a target is used for finding correspondences. Although this can be more reliable than other features, the decision may be confused when the floor is shown in one camera and not in another camera.

In the geometric based method, some authors try to align two different images by using the geometric transformation between cameras [17][18]. Although the known geometric

relationship between two different cameras facilitate an association process finding the correspondence of objects, a system requires known reference points between two different cameras to contruct the geometric relationship. This kind of the pre-construction process can restrict the camera movement in a large scale multiple camera environment. Also, the performance is only ensured when disparity between the geometric relationship of cameras is small. Another method is to use the boundary of the field of view (FOV) of cameras on the ground plane [4]. The boundary information of FOVs are projected onto the image of the other cameras. The object association is established when corresponding targets cross the same boundary on each camera. Although this method finds the correspondence of objects effectively, there are several limitations. The association is established only when objects cross over the boundaries of FOVs. This cannot support the re-association right after association failures due to miss detection or occlusion. Also, when a camera pans or tilts, the system needs the calibration or training to generate the FOV lines on the other cameras. Another method is to use the epipole line with the relationship between cameras [19]. A point on a camera is in a line with the camera's focal point. This line is shown in the other camera by using the camera geometry. If the epipole line from an object in a camera passes the object in other camera, the correspondence between detected targets is established. However, this method highly depends on the camera geometry and is sensitive to the accuracy of the camera calibration.

The fundamental problem for objects association with these approaches is to restrict the camera movement because of the training or the calibration. They usually use the perspective projection model for the transformation and alignment. Cameras should be calibrated to decrease the non-ideal error before the system starts in the application. However, the calibration process usually needs the planar and regular target as reference, which is hard to obtain in real time tracking applications [20][21][22][23][24]. That is why the calibration is performed at the initial system stage with the known reference. The camera movement is also restricted to avoid the calibration while the tracking system operates. Moreover, in an autonomous surveillance system, the proper camera movement for the object tracking is necessary to secure the effective view and to manage the resource.

### 2.3 Approach Overview

In the proposed approach, the parallel projection model based transformation technique is applied for the object association [14]. At the initial system stage, the zoom factor and the non-ideal parameter compensation table are measured. Since the zoom factor table is independent of the camera configurations (i.e., panning, tilting and zooming), the camera does not require the calibration whenever the camera changes the movement or location. The object association is locally initiated by generating a homographic line on each target. Locally generated homographic lines are exchanged among different cameras to find intersections with corresponding targets. The detailed association process with locally generated homographic lines is explained in the following section.

## 3. Multiple Objects Association

### 3.1 Local Line-based Association

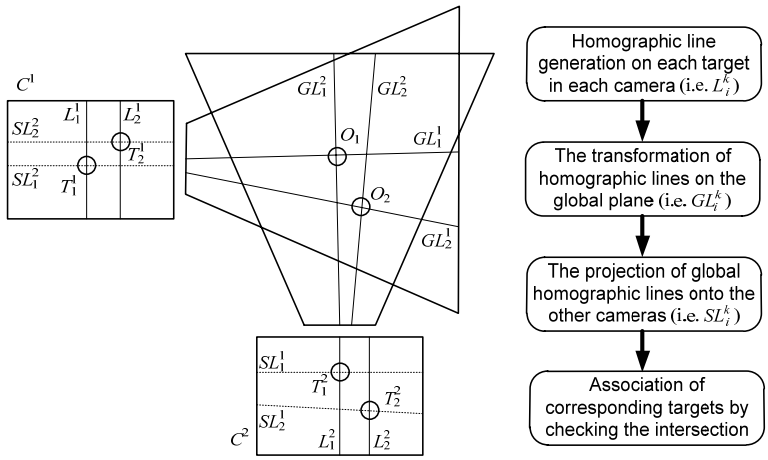



**Fig. 1**. Illustration of the homographic lines based association method with the flow chart

Object association is initiated by locally generating homographic lines on detected targets in each camera. The locally generated homographic lines are exchanged among different cameras to find intersections with targets through a global plane such as the ground plane or the reference plane. Fig. 1 illustrates the homographic lines based association method using two cameras. $L_i^k$ denotes a locally generated homographic line on target $T_i^k$, a detected target of object $i$ on camera $C^k$. $GL_i^k$ is a transformed homographic line on a global plane and $SL_i^k$ is a projected homographic line from $GL_i^k$ on the other camera $C^l$. In the figure, solid lines are locally generated homographic lines and dotted lines are projected homographic lines. The association between targets is established when their projected homographic lines intersect with corresponding targets each other. For example, $T_1^1$ intersects with a projected homographic line generated from $T_1^2$ and vice versa, and targets $\{T_1^1, T_1^2\}$ and $\{T_2^1, T_2^2\}$ are associated respectively.

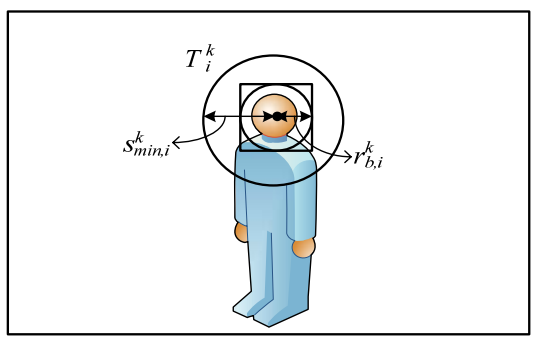



**Fig. 2**. Illustration of key parameters in locally initiating homographic lines based association method

The association establishment of locally generating homographic lines based association is determined by the correct intersection with the corresponding targets. However, a projected homographic line can be deviated from the centroid of a target because of an inaccurate global plane and synchronization issues between cameras. The accuracy problem of a global plane becomes significant when the common ground plane may not be shown to all the cameras and the faces of objects are extracted [25][26] as association targets. Because the heights of

association targets are not known to a system, the global plane with the average height of targets is used for the transformation and the projection of homographic lines it can create the deviation problem. In order to ensure the correct intersection with the corresponding target, we introduce the tolerable region around a target with the size $r^k_{\min,i}$ as shown in **Fig. 2**. Radius $s^k_{\min,i}$ denotes the adjusted radius of the tolerable region for a projected homographic line deviated by an inaccurate global plane and synchronization issues between cameras. The circle with $s^k_{\min,i}$ is called as an association circle throughout the paper. In the following section, we identify non-ideal parameters causing the deviation problem and how to determine the size of $s^k_{\min,i}$.

### 3.2 Non-ideal Parameters in Line Generation

The size of an association circle affects the association performance of the proposed method. While it ensures the correct intersection of homographic lines against the non-ideal parameters, the large size of an association circle may increase the possibility that the target intersects with homographic lines generated from irrelevant detected targets. Therefore, the appropriate size of an association circle is important in terms of the association performance. Since the size of an association circle depends on the effect of the non-ideal parameters among multiple cameras, we predetermine the size of an association circle by considering all the possible camera movements. In the following subsections, we show the effect of the non-ideal parameters to determine the size of an association circle. While camera $C^1$ is fixed at $P_1(3m, 0m, 3m)$, camera $C^2$ is considered to be three positions (i.e., $P_2(6m, 3m, 3m)$, $P_3(3m, 6m, 3m)$ and $P_4(0m, 3m, 3m)$) as varying the tilting and panning angle also.

1) Effect of the Transformation Error: When the homographic line is generated, the image coordinates are transformed to the global coordinates by using the parallel projection model as shown in **Fig. 3**. This model uses the table based error compensation to support the dynamic camera configuration instead of using the known reference points. Since the image is affected by the lens distortion or the camera configurations (panning, tilting, and zooming), this deteriorates the transformation accuracy. This can cause the problem that a homographic line does not intersect with the detected target.

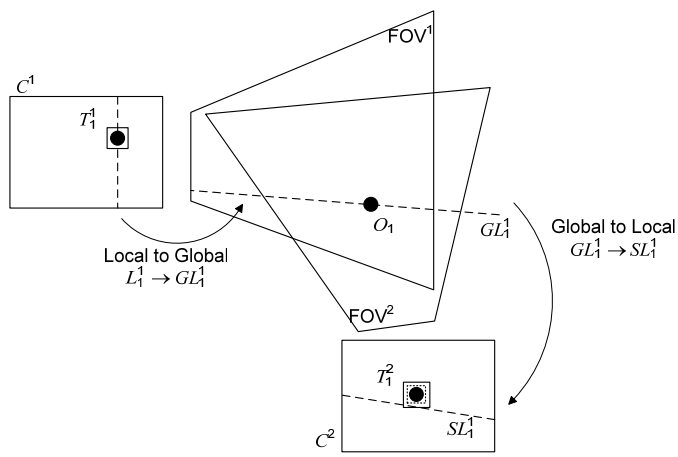
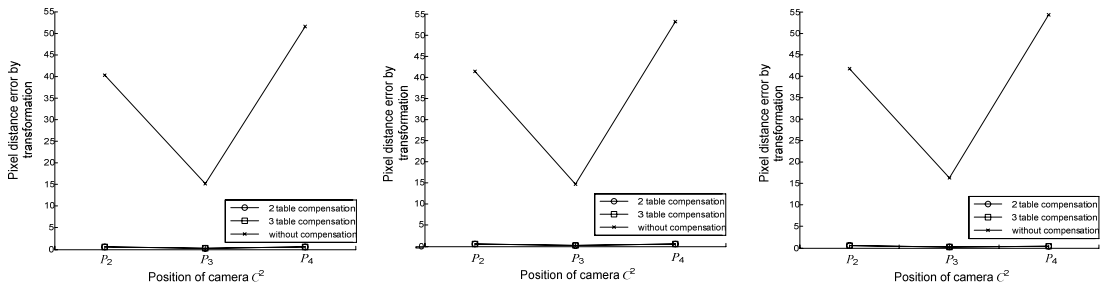



**Fig. 3**. Illustration of the transformation error caused by the camera nonlinearity

**Fig. 4** shows the maximum transformation error of the generated homographic line on camera $C^2$. For simplicity, the zoom factors of horizontal axis and vertical axis of the image are set to 1 and 1.7 respectively. Camera $C^1$ is fixed at $P_1$ and the panning angle of camera $C^2$ varies $-10°$ to $10°$ and the tilting angle varies $45°$ to $60°$ at each position. The homographic lines from camera $C^1$ are transformed onto the ground and the global lines are projected onto camera $C^2$. Among the projected lines, the maximum pixel distance from the ideally projected line is shown in y-axis. When the homographic line is transformed without the compensation for the lens non-linearity, the pixel distance error is very larger than the homographic line with the compensation. The result shows that the pixel distance error is almost 1 *pixel* on the image when the lens non-linearity is compensated and the transformation error is negligible with the compensation.




(a) Ttilting angle = 45° (b) Tilting angle =50° (c) Tilting angle = 55°

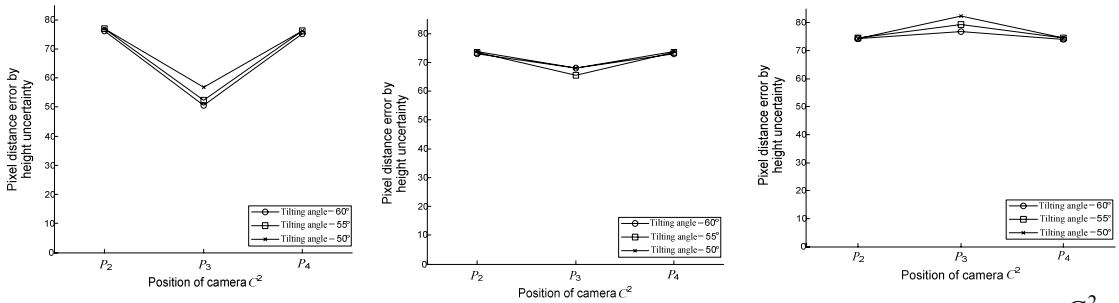
**Fig. 4**. The transformation error on $C^2$ according to the camera status (i.e., position, panning and tilting angle)

2) Effect of the Height Uncertainty: Since the faces of objects are association targets, the system does not know the height of the global plane accurately where homographic lines are generated. Thus, the system uses the average height of objects for the global plane. The difference between the actual height and the average height of objects induces homographic lines to be shifted from the corresponding targets. $\sigma_h$ denotes the amount of pixels to compensate for the shifted homographic lines due to height uncertainty.

**Fig. 5** shows how much the homographic line is shifted from the generated homographic line due to the height uncertainty. The effect of the height uncertainty is simulated similarly to the transformation error. The average height is assumed to be $1.6m$. As varying the status of camera $C^2$ at each position with the height uncertainty, $-0.1m \sim 0.1m$, the maximum pixel distance error is measured. The value of *y*-axis indicates the maximum number of pixels between the projected homographic line with the known height and the projected homographic line with the height uncertainty. Since the effect of the height uncertainty is measured by transforming homographic lines, it also considers the effect of the transformation error. The result indicates that the effect of the height uncertainty increases when the zoom factor is large on the other camera. This is because an object is viewed more closely to a camera when the zoom factor is larger.

3) Synchronization Issue: Another problem in object association is synchronization issue among cameras. Suppose $N$ cameras process $P$ frames per second. Then, the maximum synchronization error is $1/P$ second (i.e., out of synchronization by 1 frame) because they sample an image at the same rate. In **Fig. 6**, camera $C^2$ is delayed by 1 frame two cameras are placed in the perpendicular way for the maximum effect of the synchronization issue. Object

$O_2$ moves in the diagonal way to the optical axis of camera $C^2$. Since the captured views of cameras are different from each other, the projected homographic line does not intersect with the detected targets of object $O_2$.




(a) When the zoom factor of $C^2$ is 0.2 smaller than that of $C^1$

(b) When the zoom factors of two cameras are the same

(c) When the zoom factor of $C^2$ is 0.2 larger than that of $C^1$

**Fig. 5**. The effect of the height uncertainty on the other camera according to the camera status (i.e., position, panning and tilting angle, zoom factor)

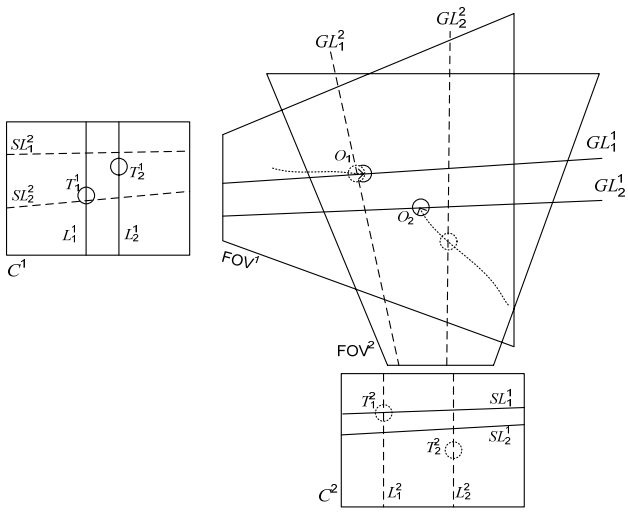



**Fig. 6**. Illustration of association problem in asynchronous image frames.

**Fig. 7** shows the effect of the synchronization problem. $v_f$ represents the object speed per the image sampling time. It is assumed that the maximum frame delay between cameras $C^1$ and $C^2$ is 1 frame as an object moves in the perpendicular way to the optical axis of $C^1$ at speed $v_f$. The homographic line is generated from the delayed image of camera $C^1$ to the image of camera $C^2$. The maximum pixel difference error between the homographic lines from the delayed image and the detected target in camera $C^2$ is measured. As the object moves fast and closely to camera $C^2$, the pixel error increases. $\sigma_s$ is defined as the maximum pixel distance error caused by the synchronization issue depending on the applications (i.e., object speed and the camera position).

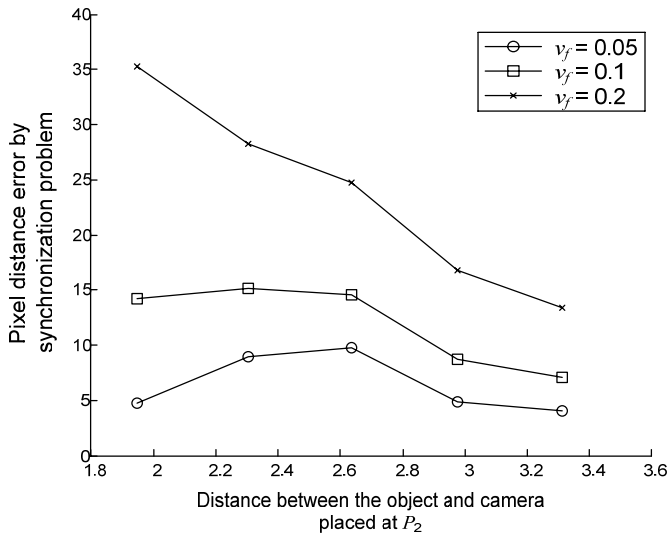



**Fig. 7**. The effect of the synchronization problem depending on the distance between the object and camera

### 3.3 Parameters Affecting Association Performance

Since the effect of transformation is included in the height uncertainty, the required minimum size of an association circle is represented as,

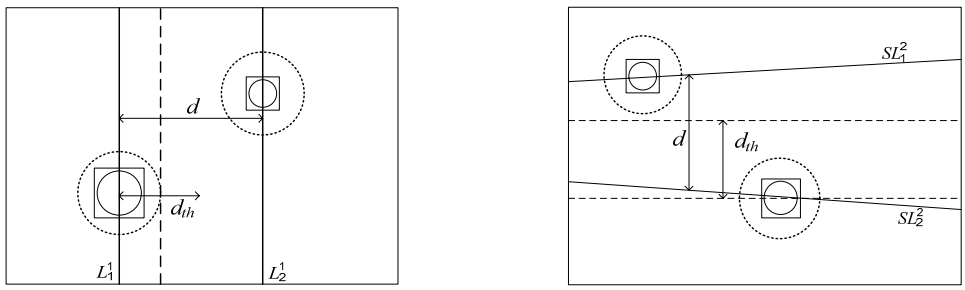
$$s_{\min} \approx \max(\sigma_h + 2\sigma_s)/2, \tag{1}$$

where $\sigma_h$ and $\sigma_s$ are measured in all the possible camera positions for the application. Based on this, the system increases the box size to $s_{\min}$ so that a projected homographic line intersects with the detected box. On the other hand, the increased size of the detected target also increases the case that an object is intersected with multiple homographic lines. The proposed algorithm is affected by this case and the increased size may decrease the association performance. Moreover, the association condition is not satisfied when a homographic line intersects with more than two same objects on both cameras. Thus, the parameter indicating the effectiveness of the homographic line should be def ined.

The threshold distance is the minimum distance between targets, which is required to generate the effective vertical homographic lines on the other cameras. **Fig. 8** illustrates the threshold distance in each camera. The system knows which vertical homographic line is generated from which object in the local camera. Hence, although generated vertical homographic lines intersect with two detected targets in the local camera, it does not affect the association process. However, when the projected homographic lines from them intersect with two detected targets in the other camera, the system cannot determine the corresponding target. The one way to expect this is to use the globally defined threshold distance, $d_{th}$, which is represented as

$$d_{th} \approx \max(\sigma_h + 2\sigma_s), \tag{2}$$

where the maximum errors of the non-ideal parameters are considered. When the homographic lines are generated in one camera, the separation of the transformed lines from them is anticipated with $d_{th}$ in the other cameras. However, we need to consider that $d_{th}$ is the reference value obtained with the predefined zoom factor.




(a) The view of camera $C^1$ (b) The view of camera $C^2$

**Fig. 8**. The view of each camera with $d_{th}$ when the homographic line is generated from camera $C_1$

**Fig. 9** illustrates the size of detected target compared with $d_{th}$ and the zoom factor of the other camera. $d_{th}$ is determined by (2) based on the simulation with the predefined zoom factor for the non-ideal parameters. The size of most detected targets is smaller than $d_{th}$ with the height uncertainty. Thus, the box size of them needs to be expanded so that the non-ideal factors can be compensated. However, $d_{th}$ decreases when the height is known because the effect of the height uncertainty is negligible. Moreover, if the other cameras use the different zoom factor from the camera where a homographic line is generated, $d_{th}$ also changes in proportion to the zoom factor.

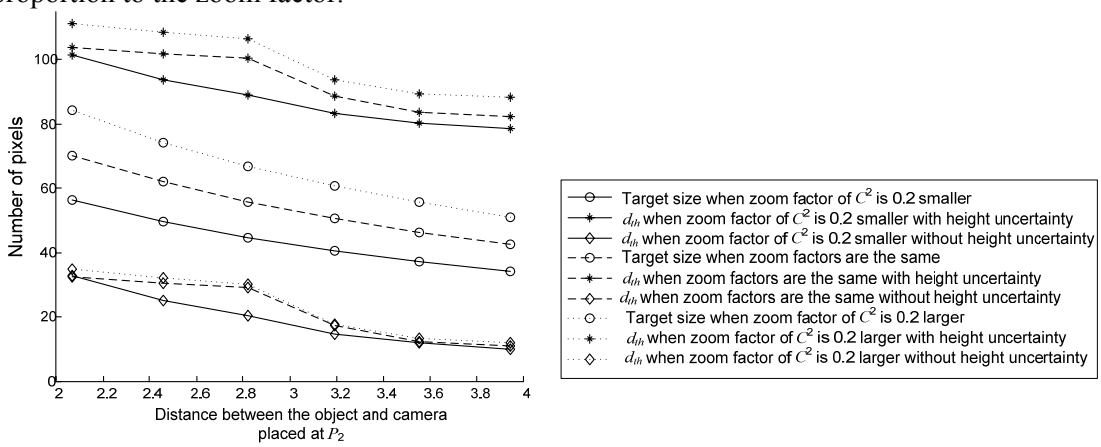



**Fig. 9**. Relationship between the size of the detected target and $d_{th}$

When $d_{th}$ is considered for the line generation, the system has the case that two objects are too close to be associated with the vertical homographic lines as shown in **Fig. 10**. If $d$ is smaller than $d_{th}$, the vertical homographic line is not effective as shown in **Fig. 11**. Hence we introduce a slant homographic line. As generating a slant homographic line, the distance

between two homographic lines satisfies with the separation $d_{th}$. The angle $\theta$ is the slope of the homographic line. In order to obtain the new distance between the homographic lines, we use the distance between a point and a line. For example, the homographic line passing the center of object $O_1$ is defined by

$$\tan(\theta)x - y - \tan(\theta)x_1 + y_1 = 0. \tag{3}$$

The distance between the line and the center of $O_2$ is calculated by

$$d' = \left| \frac{\tan(\theta)x_2 - y_2 - \tan(\theta)x_1 + y_1}{(\tan(\theta))^2 + 1} \right|. \tag{4}$$

The angle $\theta$ is chosen as finding $d' \geq d_{th}$. In order to decrease the number of comparison cases, the system can choose $\theta$ among the limited angle candidates (i.e., 0°, 45° and 135°).

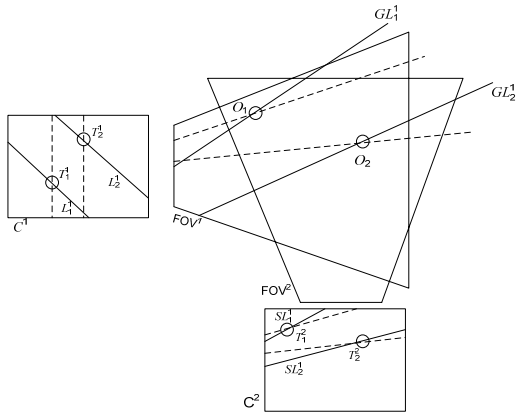
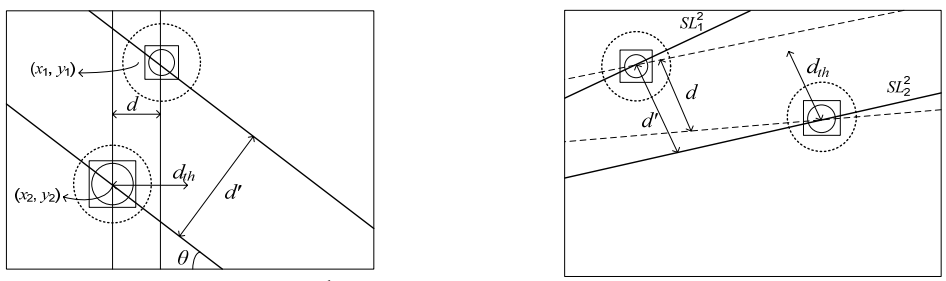



**Fig. 10**. Illustration of the case that the vertical homographic line is not effective




(a) The view of camera $C^1$ (b) The view of camera $C^2$

**Fig. 11**. The view of each camera when the slant homographic line is generated from camera $C^1$

1) Vertical Line based Association: This association method uses only the vertically generated homographic lines. Each camera generates the vertical homographic lines on the detected target considering the threshold distance. If the homographic line is not satisfied with the threshold distance, the homographic line is not used for object association. As shown in **Fig. 12**, all the cameras generate homographic lines on detected targets. However, the homographic lines generated by camera $C^2$ do not satisfy the threshold distance. The distance between $L_1^2$ and $L_2^2$ is smaller than $d_{th}$. Thus, they are not effective anymore and only cameras $C^1$ and $C^3$ associate the corresponding targets.

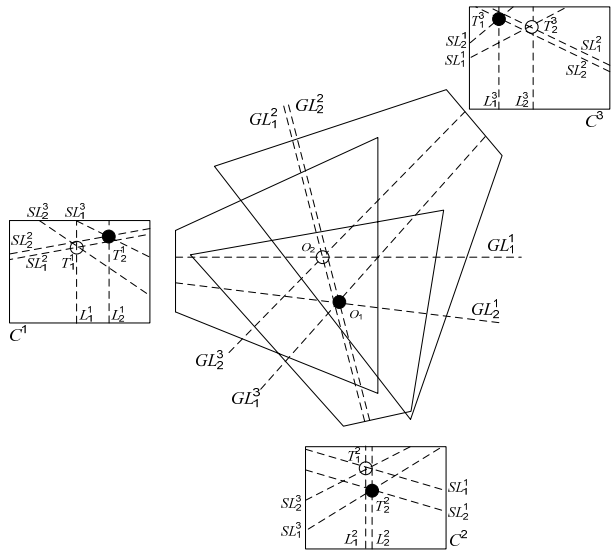


**Fig. 12**. Illustration of object association by the vertical homographic lines

2) Slant Line based Association: When slant homographic lines are used for object association, there are two methods to generate them. One is the locally selected line and another is the globally selected line. In the locally selected line method, the angle $\theta$ is chosen when the homographic line is generated. **Fig. 13** shows the example how the slant homographic line is applied to object association. In camera $C^2$ which does not participate in object association with the vertical homographic lines, a slant homographic line is generated on each detected target.

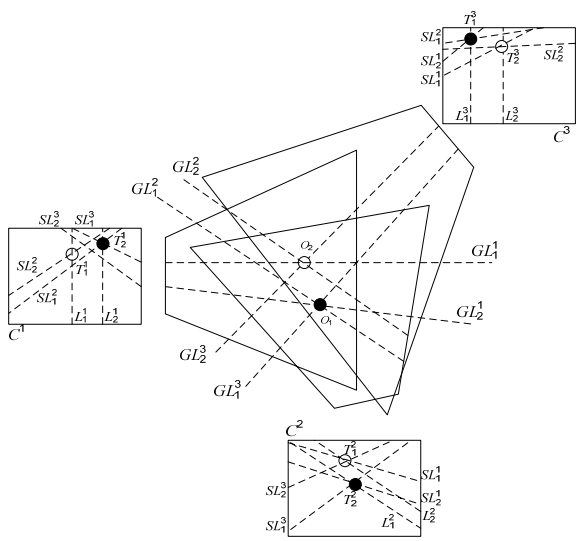


**Fig. 13**. Illustration of object association by the possible slant homographic lines

In the figure, $d$ of lines generated by $C^2$ is larger than $d_{th}$. Hence they are still effective and camera $C^2$ can also participate in object association. In the second approach, all the kinds of effective homographic lines including the vertical lines are generated and transferred to the system. The system determines which lines are used for association after testing the threshold

distance in all the cameras. Although this can choose the best homographic lines for object association, this requires the large amount of computation for comparison with the threshold distance.

Algorithm 1 explains the line generation method based on $d_{th}$. The purpose of this algorithm is to find the homographic lines with $\theta$ having the maximum separation in other cameras so that the association performance increases. First, the system determines the set of $\theta$s to be tested. As increasing the number of elements in the set, the comparison cases increase. Then, the system calculates the minimum distances when the homographic line is generated

**Algorithm 1**: the line generation algorithm with $d_{th}$

Input : $d_{th}$ ,
Coordinates of detected objects with the expanded box size from each camera
**for** $k$ = 1 *to* $K$ **do**
  Calculate *min* $d^{90^\circ}$ between detected objects at $C^k$
  **if** $d^{90^\circ} < d_{th}$ **then**
    find *max* $d^{\theta}$ satisfying $d^{\theta} \geq d_{th}$
  **end**
  **if** $d^{\theta}$ *is found* **then**
    **for** $i$ = 1 *to* $I$ **do**
      **if** *A detected target* $T_i^k$ *has at least* $d_{th}$ *of separation* **then**
        Generate homographic lines with $\theta$ on detected objects
      **end**
    **end**
  **end**
**end**

## 4. Simulation and Analysis

### 4.1 Simulation Setup

**Fig. 14** shows object trajectories and camera placements for analyzing the proposed association algorithm. Camera $C^1$ is placed at (x = 3.57$m$, y = 0.05$m$, z = 2.24$m$) with tilting angle 73°and panning angle 0°and camera $C^2$ is placed at (x = 0.05$m$, y = 2.97$m$, z = 2.34$m$) with tilting angle 68° and panning angle 0°. The total number of frames is 45 and object $O_3$ is shown to both cameras after frame 23. The height of object $O_1$ is 1.75$m$, the height of object $O_2$ is 1.87$m$, and the height of object $O_3$ is 1.72$m$. The average height of the objects is 1.78$m$. Since a target height is considered to be the center of an object face in height estimation, the target height is shorter than the real height of an object. Thus, z = 1.7$m$ is used as an initial global plane where homographic lines are transformed. If an object height is estimated, an object has its own global plane of an estimated object height. 60 *pixels* are used for the initial value of $\sigma_h + \sigma_s$ .

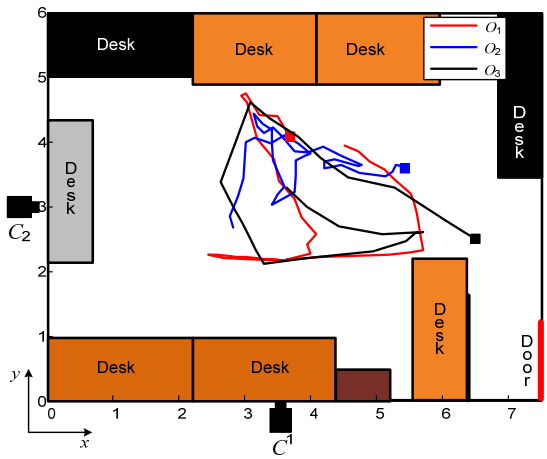



**Fig. 14**. Illustration of our simulation setup showing objects trajectories and camera placements (Each square denotes the starting position of each object) with each $\theta$. Finally, the system selects $\theta$ which has the maximum distance among the minimum distances. Then, the system tests if each object has at least $d_{th}$ of the separation with other objects. If this is satisfied, the system generates the homographic line with $\theta$.

## 4.2 Performance with Constant Radius of Association Circle

**Fig. 15** shows snapshots for objects association with a constant radius. The system generates homographic lines on images only when the distance between targets is satisfied with the threshold. At frame 5, targets are not associated because transformed homographic lines intersect with multiple targets. This indicates that the threshold does not always lead to the correct decision for generating homographic lines. However, if the distance between targets is smaller than the threshold, the generated homographic lines are even less effective for objects association. After frame 5, targets between two images are associated by locally initiating homographic lines. As shown in the figure, an adjusted region size $s_{\min}$ of each target is constant.

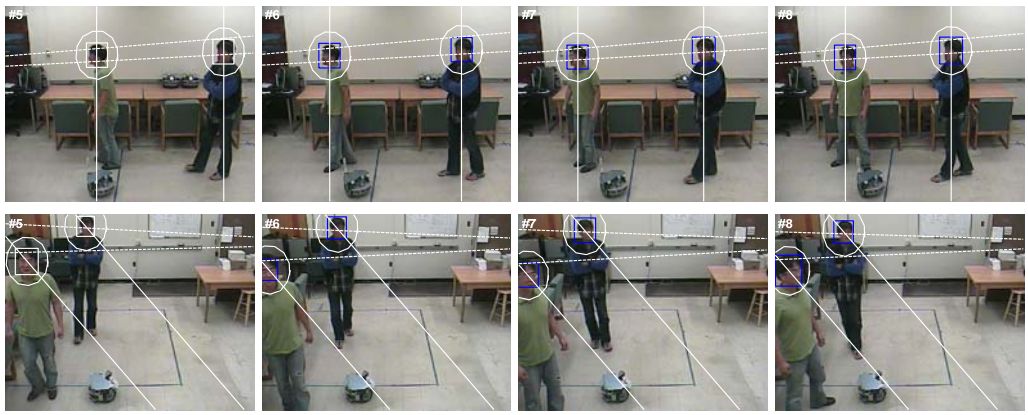


**Fig. 15**. Image snapshots of object association with constant threshold based association algorithm at frame 5 ~ 8 (The upper images are captured by camera $C^1$ and the lower images are captured by camera $C^2$. Blue rectangles of targets indicate associated targets; white solid lines, locally initiated

homographic lines; and white dotted lines, projected homographic lines from locally initiated homographic lines).

Moreover, the simulation shows that projected homographic lines deviate from the centroid of targets because of the non-ideal parameters. The association status of each object is shown in **Fig. 16**.

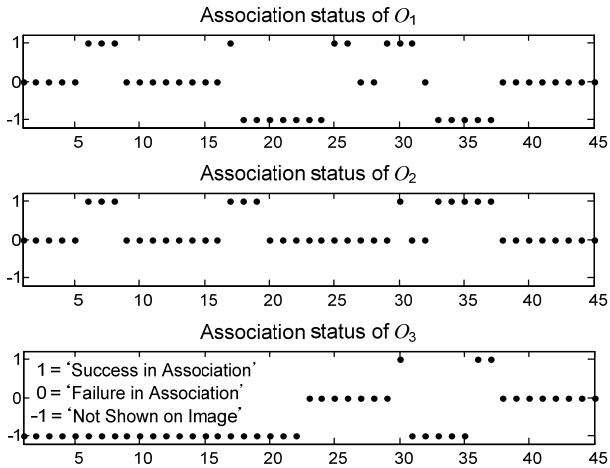



**Fig. 16**. Simulation result of the association status of each object with the constant radius based association algorithm

### 4.3 Association Performance Improvement Using Redundancies

This section discusses the strategy to improve the association performance by minimizing the adjusted radius of an association circle after targets are associated between cameras. The first non-ideal parameter to be compensated for is the height uncertainty of targets. The effect of the height uncertainty is maximized when there is a significant discrepancy between actual height and average height of a target. If we can estimate a target height within a certain range, it is not necessary to consider significant uncertainty. A target can be localized in 3-D through finding the intersection of projected lines from targets in cameras to a ground [16]. However, the intersection of the projected lines may not exist because targets are detected in different views. Hence, we utilize the shortest line between two projected lines.

A corresponding point $p_i$ in each camera is projected onto the global ground plane (z = 0) using $p_i = Cp_i'$ where $C$ is a known camera matrix and $p_i'$ denotes a transformed point on a ground plane. Ideally, constructed lines between corresponding points and transformed points should intersect with each other if the centroid of targets is the same position of an object in 3-D. However, this is not always guaranteed because the centroid of targets can be deviated by detection algorithms. Thus, we obtain a line which has the shortest distance between the constructed lines. Then, an object position is estimated using the middle point between two points that form the shortest line between constructed two lines.

The average height of targets (i.e., centroid of detected faces of objects) is set to be $1.7m$. The target height of object $O_1$ is $1.65m$, the target height of object $O_2$ is $1.73m$, and the target height of object $O_3$ is $1.6m$. **Fig. 17** shows estimated heights of three objects assuming that their corresponding targets are known. Estimated heights can be deviated from an actual target

height because targets are detected in the different views of cameras. However, estimated heights converge on actual target heights.

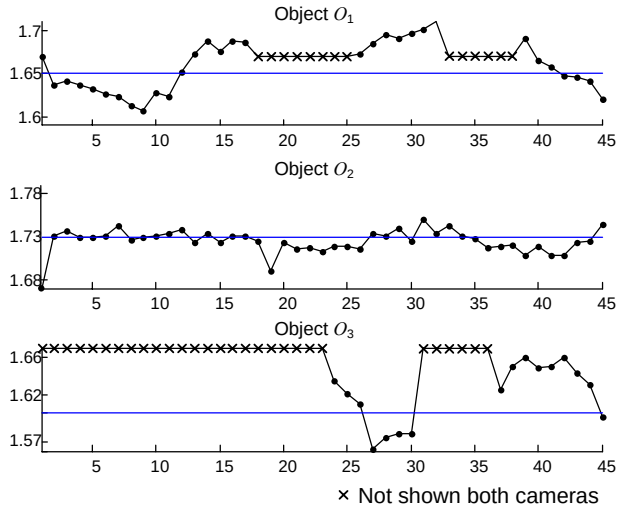
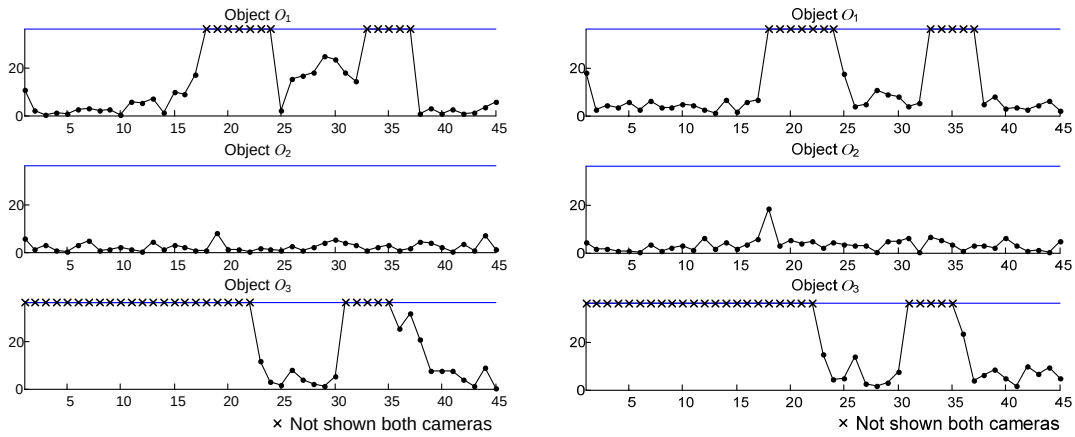



**Fig. 17**. Estimated target heights of objects by the shortest distance between two lines after corresponding targets are found (a blue solid line is the average height of targets)

**Fig. 18** shows how many pixels homographic lines are deviated from corresponding targets when estimated heights are utilized for the generation of homographic lines. The initial value of height uncertainty $\sigma_h$ is set to be 37 *pixels* assuming that the degree of height uncertainty is $0.1m$. When an object is not shown on both cameras, the initial value is used for the number of deviated pixels. The result indicates that estimated heights alleviate the deviation of homographic lines.




(a) Targets in camera $C^1$ (b) Targets in camera $C^2$

**Fig. 18**. The number of deviated pixels from the centroid of targets when estimated heights are utilized to generate and transform homographic lines

Since synchronization issues are caused by the physical network delay, it is difficult to correct the effect of synchronization issues. However, $\sigma_s$ can be optimally selected by using

information such as an object speed and a distance between an object and a camera. In order to estimate $\sigma_s$ for the next frame, the global position of an object needs to be predicted. The next position is estimated based on the previous position and velocity. The velocity of an object is estimated based on the positions for at least two consecutive frames. The velocity of object $O_i$ at time $n$ (i.e., $v_{x,i}(n), v_{y,i}(n)$) is obtained by

$$\begin{aligned} v_{x,i}(n) &= x_i(n) - x_i(n-1), \\ v_{y,i}(n) &= y_i(n) - y_i(n-1), \end{aligned} \tag{5}$$

where $(x_x(n), y_i(n))$ is a global position at time $n$. This assumes a constant velocity object model. Since the sampling rate is usually higher than object velocity, the predicted position error by the incorrect object model is not significant. The object speed $v_f^i$ of object $O_i$ per a frame is also obtained by

$$v_f^i = \sqrt{v_{x,i}^2 + v_{y,i}^2}. \tag{6}$$

**Fig. 19** shows the estimated speeds of objects per a frame when the frame rate is 8*frames*/*sec* (the maximum time difference between cameras is 0.125*sec*). The estimated speed is used to determine the optimal value for $\sigma_s$ with the pre-constructed table representing the influence of the synchronization effect.

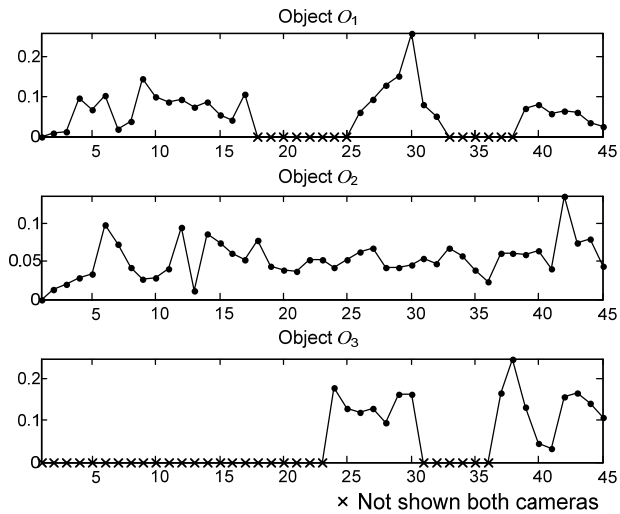


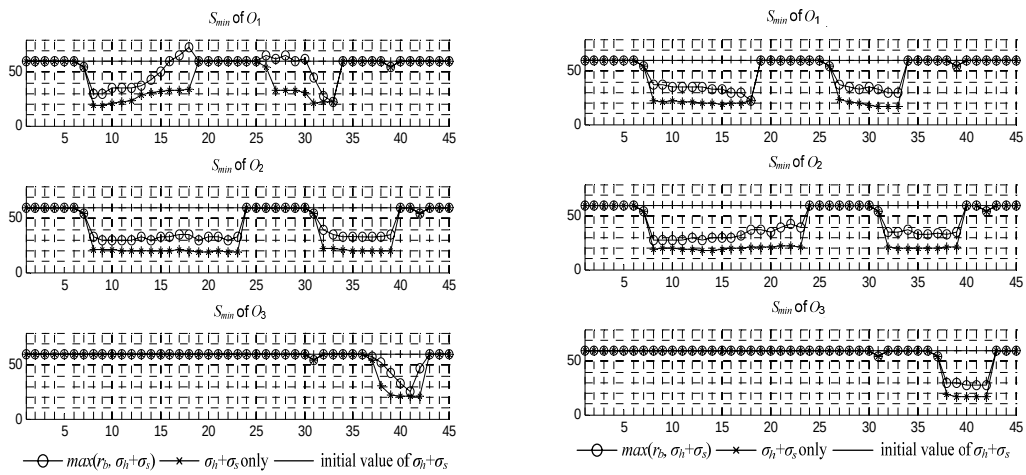

**Fig. 19**. The estimated speeds of objects per a frame assuming that the frame rate is 8*frames*/*sec* (the maximum time difference between cameras is 0.125*sec*)

**Table 1** shows the example of constructed pixel error table by using the simulated data from **Fig. 7**. $v_f / 2$ corresponds to the sampling period of **Fig. 7** since the synchronization effect is measured with the average object speed 2*m*/*s*. $D_i^{k,l}$ denotes a distance between the position of target $T_i^k$ and camera $C^l$. The table can be constructed in more detail for more subdivided selection.

**Table 1**. Constructed pixel error table for synchronization effect according to **Fig. 7**

| $v_f / 2$ | $D_i^{k,l} < 3$ | $D_i^{k,l} \geq 3$ |
|---|---|---|
| 0.05 | 10 | 5 |
| 0.1 | 22 | 12 |
| 0.2 | 45 | 25 |

**Fig. 20** shows the variation of an adjusted radius $s_{min}$ as objects move around the surveillance region. The maximum object speed is slower than 0.2*m*/*s* in this simulation. Thus, 22 *pixels* are used when the distance between an object and a camera is closer than 3*m* and 12 *pixels* otherwise according to **Table 1**. The line "$\sigma_h + \sigma_s$ only" is the smallest $s_{min}$ that can be achieved by estimating the object height and compensating for the synchronization issue. However, it is not necessary to decrease $s_{min}$ smaller than the detected box size. Thus, the system uses the line "max($r_b, \sigma_h + \sigma_s$)" for the association process. The line "initial value of $\sigma_h + \sigma_s$" is a radius that the constant radius based method uses for the entire time.




(a) $s_{\min}$ of each target in $C_1$ (b) $s_{\min}$ of each target in $C_2$

**Fig. 20**. The variation of $s_{\min}$ of each object as a function of Time

**Fig. 21** shows the association status of each object when the adjusted radius based algorithm is used. The figure shows a prominent association improvement as compared with **Fig. 16**. A new object (i.e., object $O_3$) or a failed object is also associated or recovered as time elapses.

**Fig. 22** compares the average association performance between the constant radius based association algorithm and the adjusted radius based association algorithm. An ideal situation indicates that heights of targets are known and cameras are synchronized. Although the proposed association algorithm outperforms the constant radius based association algorithm, the result indicates that homographic lines based association is not always the best solution. A system utilizes rather local tracking information to maintain association information when homographic lines cannot be guaranteed to distinctively intersect with targets.

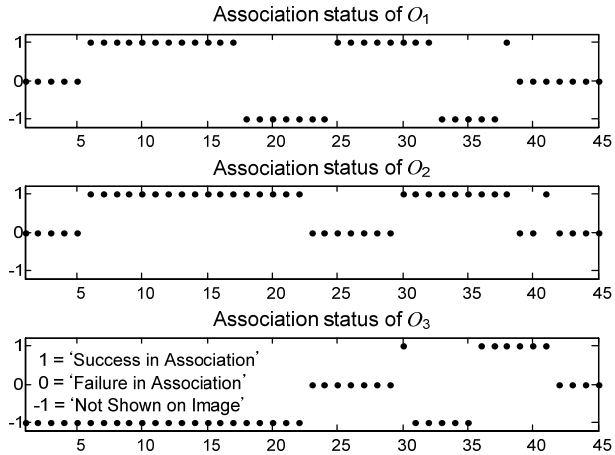
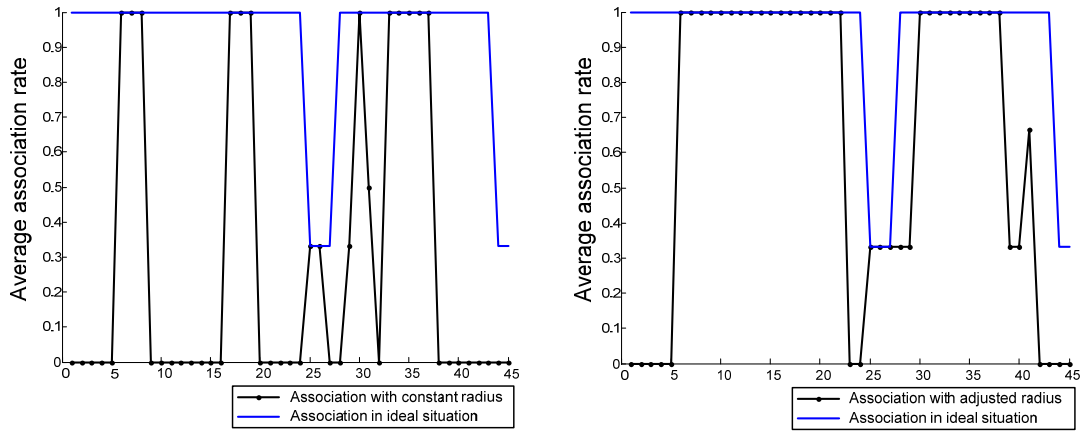



**Fig. 21**. Improvement over **Fig. 16** by adjusted radius based association algorithm




(a) Based on constant radius (b) Based on adjusted radius

**Fig. 22**. Simulation result of average association performance with constant radius based association and adjusted radius based association

## 5. Conclusions

In this paper, we showed a novel association method with local initiated homographic line for surveillance application with multiple cameras. This method is based on the parallel projection model to support the camera movement. We investigate the plausible parameters to affect the association performance. This information is used to define the threshold distance to indicate the effectiveness of the locally generated homographic line. We also discuss the strategy to improve the association performance using the temporal and spatial redundancies.

Future extensions to this work will be to maintain association information when objects are densely populated. In the densely populated environment, there is a very high possibility that homographic lines are not effective due to closely detected targets. Then, association information can only be maintained with local tracking, and this may cause false or failed association. Abandoned unassociated targets seriously delay association establishment. Thus, association maintenance scheme with previously given association information is required to accurately and effectively maintain association information.

## References


[1] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, vol. 34, no. 3, pp. 334-352, Aug. 2004.

[2] A. Utsumi, H. Mori. J. Ohya, and M. Yachida, "Multiple-view-based tracking of multiple humans," in *Proc. of Int'l Conf. Pattern Recognition*, pp.197-601, 1998.

[3] S. L. Dockstader and A. M. Tekalp, "Multiple camera tracking of interacting and occluded human motion," *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1441-1455, Oct. 2001.

[4] S. Khan and M. Shah, "Consistent labeling of tracked objects in multiple cameras with overlapping fields of view," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1355-1360, Oct. 2003.

[5] S. Calderara, A. Prati, R. Vezzani, and R. Cucchiara, "Consistent labeling for multi-camera object tracking," in *Proc. of Image Analysis and Processing (ICIAP)*, pp. 1206-1214, 2005.

[6] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer, "Multi-camera multi-person tracking for easy living," in *Proc. of IEEE Int'l Workshop Visual Surveillance*, Dublin, Ireland, pp. 3-10, Jul. 2000.

[7] J. Li, C.S. Chua, and Y.K. Ho, "Color based multiple people tracking," in *Proc. of IEEE Int'l Conf. on Control, Automation, Robotics and Vision*, vol. 1, pp. 309-314, 2002.

[8] Q. Cai and J. K. Aggarwal, "Tracking human motion using multiple cameras," in *Proc. of Int'l Conf. on Pattern Recognition ,Vienna, Austria*, pp. 68-72, 1996.

[9] Y. Caspi, D. Simakov and M. Irani, "Feature-based sequence-to-sequence matching," *International Journal of Computer Vision*, pp. 53-64, 2006.

[10] J. Orwell, P. Remagnino and G.A. Jones, "Multiple camera color tracking," *IEEE Int'l Workshop Visual Surveillance*, pp. 14-24, Jun. 1999.

[11] A. Mittal and L.S. Davis, "M2Tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo," in *Proc. of European Conf. Computer Vision*, pp. 18-36, May 2002.

[12] J. Kang, I. Cohen, and G. Medioni. "Continuous tracking within and across camera streams," in *Proc. of IEEE Int'l Conference on Computer Vision and Pattern Recognition*, vo.1, pp.267-272, 2003.

[13] S. Chang and T.-H. Gong. "Tracking multiple people with a multi-camera system," in *Proc. of IEEE Workshop on Multi-Object Tracking*, pp. 19-26, 2001.

[14] S. H. Cho, S. Hong. W. Cho, "Homographic line generation and transformation technique for dynamic object association," in *Proc. of Int'l Workshop on Machine Learning for Signal Processing*, pp. 273-278, Oct. 2008.

[15] K. S. Park, J. Lee, M. Stanaćević, S. Hong and W. D. Cho, "Iterative object localization algorithm using visual images with a reference coordinate," *EURASIP Journal on Image and Video Processing*, vol. 2008, Article ID 256896, July 2008.

[16] W. Hu, M. Hu, X. Zhou, T. Tan, J. Lou and S. Maybank, "Principal axis-based correspondence between multiple cameras for people tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 663-671, Apr. 2006.

[17] H. Tsutsui, J. Miura and Y. Shirai, "Optical flow-based person tracking by multiple cameras," in *Proc. of IEEE Conf. Multisensor Fusion and Integration in Intelligent Systems*, pp. 91-96, Aug. 2001.

[18] A. Utsumi, H. Mori, J. Ohya and M. Yachida, "Multiple human tracking using multiple cameras," in *Proc. of IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 498-503, Apr. 1998.

[19] J. Black and T. Ellis, "Multi camera image tracking," *Image and Vision Computing*, vol. 24, pp. 1256-1267, 2006.

[20] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, Nov. 2000.

[21] L. Lucchese, "Geometric calibration of digital cameras through multi-view rectification," *Image and vision computing*, vol. 23, no. 5, pp. 517-539, May 2005.

[22] J. Wang, F. Shi, J. Zhang and Y. Liu, "A new calibration model of camera lens distortion," *Pattern Recognition*, vol. 41, no. 2, pp. 607- 615, Feb. 2008.
[23] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correlation," in *Proc. of Conf. on Computer Vision and Pattern Recognition*, pp. 1106-1112, 1997.
[24] R. Hassanpour, Volkan Atalay, "Camera auto-calibration using a sequence of 2D images with small rotations," *Pattern Recognition Letters*, vol. 25 pp. 989-997, 2004.
[25] S. Z. Li and Z.Q. Zhang, "FloatBoost learning and statistical face detection," *IEEE Trans. on Patteran Analysis and Machine Intelligence*, vol. 26. no. 9, pp. 1112-1123, Sep. 2004.
[26] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, May 2004.

**Shung Han Cho** received B.E. (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 received M.S. in Electrical and Computer Engineering at Stony Brook University with Award of Honor in recognition of outstanding achievement and dedication in 2008. He is currently pursuing his Ph.D. degree in the department of Electrical and Computer Engineering 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 and evaluation.

**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 - State University of New York. 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 low power VLSI design of multimedia wireless communications and digital signal processing systems, reconfigurable SoC design and optimization, VLSI signal processing, and low-complexity digital circuits. Prof. Hong served on numerous Technical Program Committees for IEEE conferences. Prof. Hong is a Senior Member of IEEE.

**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.