Multiple Camera Collaboration Strategies for Dynamic Object Association

Shung Han Cho[1], Yunyoung Nam[2] and Sangjin Hong[1]

[1]Mobile Systems Design Laboratory, Dept. of Electrical and Computer Engineering,
Stony Brook University-SUNY, Stony Brook, NY 11794 - USA
[e-mail: {shcho, snjhong}@ece.sunysb.edu]
[2]Center of excellence for Ubiquitous System, Ajou University,
Suwon, 443-749 - South Korea
[e-mail: youngman@ajou.ac.kr]
*Corresponding author: Sangjin Hong

Abstract

In this paper, we present and compare two different multiple camera collaboration strategies to reduce false association in finding the correspondence of objects. Collaboration matrices are defined with the required minimum separation for an effective collaboration because homographic lines for objects association are ineffective with the insufficient separation. The first strategy uses the collaboration matrices to select the best pair out of many cameras having the maximum separation to efficiently collaborate on the object association. The association information in selected cameras is propagated to unselected cameras by the global information constructed from the associated targets. While the first strategy requires the long operation time to achieve the high association rate due to the limited view by the best pair, it reduces the computational cost using homographic lines. The second strategy initiates the collaboration process of objects association for all the pairing cases of cameras regardless of the separation. In each collaboration process, only crossed targets by a transformed homographic line from the other collaborating camera generate homographic lines. While the repetitive association processes improve the association performance, the transformation processes of homographic lines increase exponentially. The proposed methods are evaluated with real video sequences and compared in terms of the computational cost and the association performance. The simulation results demonstrate that the proposed methods effectively reduce the false association rate as compared with basic pair-wise collaboration.

Keywords: Object detection, object association, homographic line, multiple cameras, collaboration

---

1. Introduction

Recently, multiple cameras based surveillance system has received much attention to cover larger areas with possibly overlapped views of multiple cameras. The redundant view by multiple cameras can improve objects detection and tracking by minimizing the effect on the system caused by false or failed detection and occlusion [1][2]. One of the key requirements in multiple cameras based system is to have the consistent view of objects among different cameras as maintaining the redundant information consistently and robustly. Especially, when multiple cameras flexibly change their views for a large scale surveillance system, it is critical to minimize falsely associated objects as ensuring a high association rate. Once the inconsistent information by the false association is generated in the system, it may be propagated in time and it is difficult to be corrected. Therefore, minimizing the false association through multiple camera collaboration for the object association has become an important issue in the large scale surveillance system.

There are numerous association approaches using feature matching or the geometry of multiple cameras to find the correspondence of objects among multiple cameras [3][4][5][6][7][8][9][10][11][12][13][14][15][16]. Feature based approaches usually suffer from the unavailability of distinctive features for all objects [3][4][5][6][7][8], and more over are sensitive to detection performance. On the other hand, geometry based approaches require the accurate calibration process to construct the relationship among cameras [9][10][11]. Some methods combine both approaches to find the correspondence of objects [12][13]. However, feature based approaches have a higher chance to generate the false association due to the sensitivity to detection performance. In geometry based approaches, the boundaries of camera views on the ground plane are used as stationary homographic lines in other cameras to associate targets when they cross the corresponding boundaries of camera views [14][15]. The boundary information of camera views is predetermined in advance or determined by observing the motion of objects. However, the association process is limited by stationary objects which do not cross the boundaries of camera views. Moreover, the determination process of the boundary information makes the system difficult to promptly support the flexible movements of cameras.

In order to dynamically establish the association for objects, Kyong et al. [16] present an association method that homographic lines are locally generated on targets in each camera and they are projected to among the other cameras. Since it is not necessary to have a ground plane as a common reference plane, all the cameras do not need to see the ground plane. Homographic lines are generated when the degree of separation between them is satisfied. The required minimum separation between each pair of cameras is predetermined by incorporating the effect of targets height uncertainty and frame synchronization errors because the reference plane may not be the same as the actual height of targets. The method can be extended to support multiple cameras through pair-wise collaboration for the object association and combine the association information from each pair of collaborating cameras. While the pair-wise collaboration is effective for objects with the enough separation, the association is not well-established for objects without the enough separation and it may generate the false association. Therefore, an effective camera collaboration is necessary to reduce inconsistent and uncertain information.

In this paper, we extend the locally initiating homographic lines based association method to two different multiple camera collaboration strategies that reduce the false association.

Collaboration matrices are defined with the elements of the required minimum separation presented in [16]. The first strategy compares the collaboration matrices with the minimum separation of objects for each pair of cameras and selects the best pair out of many cameras satisfying the required minimum separation. After targets are associated in selected cameras, the association information is propagated to unselected cameras by transforming the global information constructed from the associated targets. The selection based strategy efficiently collaborates on the object association with the best pair of the cameras as reducing the false association. However, it requires the long operation time to increase the association rate due to unsatisfied separation when a large number of targets are detected. In order to shorten the operation time for the high association rate, the second strategy initiates the collaboration for all the pairing cases of cameras regardless of the separation. When each pair of cameras collaborates on object association, a homographic line is generated on each target and it is projected to the other collaborating camera. The other camera generates homographic lines on only the crossed targets by the projected homographic lines and they are re-projected to the one camera. This association process is iteratively operated for all the unassociated targets in all the pairing cases of cameras. The proposed methods are evaluated with real video sequences and they are compared with the basic pair-wise collaboration to demonstrate the effective and efficient association.

The remainder of this paper has 4 sections. In Section 2, we present the overview of homographic lines based association method and describe the association problem in terms of the false association and the computational costs. Section 3 investigates two collaboration strategies for objects association to minimize the inconsistency in the system and to improve the efficiency of using homographic lines. In Section 4, we verify the proposed methods with the real video sequences. Finally, our contribution is summarized in Section 5.

2. Background and Problem Description
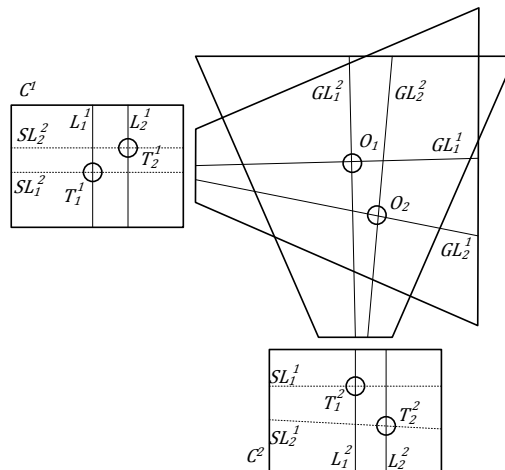
2.1 Background

Fig. 1. Illustration of the homographic lines based association method

The locally generated homographic lines based association method is used to collaboratively associate targets among multiple cameras. **Fig. 1** illustrates the homographic lines based association method where two cameras are used to associate objects [16]. A homographic line

L_i^k is generated on T_i^k, a detected target of object i, in each camera C^k. Each of these homographic lines is transformed to a global plane and projected to the other camera. GL_i^k denotes a transformed homographic line on a global plane and SL_i^k denotes a projected homographic line from GL_i^k on the other camera C^l. The association between targets is established if a transformed homographic line intersects with a corresponding target distinctively. **Table 1** illustrates the information of crossed targets for **Fig. 1**. For example, T_1^1 is crossed by a projected homographic line generated from T_1^2 and T_1^2 is crossed by a projected homographic line generated from T_1^1. Thus, targets $\{T_1^1, T_1^2\}$ and $\{T_2^1, T_2^2\}$ are found as the correspondence of targets respectively.

Table 1. Crossed targets by homographic lines for **Fig. 1**

	Crossed targets in C^1	Crossed targets in C^2
T_1^1	-	$\{T_1^2\}$
T_2^1	-	$\{T_2^2\}$
T_1^2	$\{T_1^1\}$	-
T_2^2	$\{T_2^1\}$	-

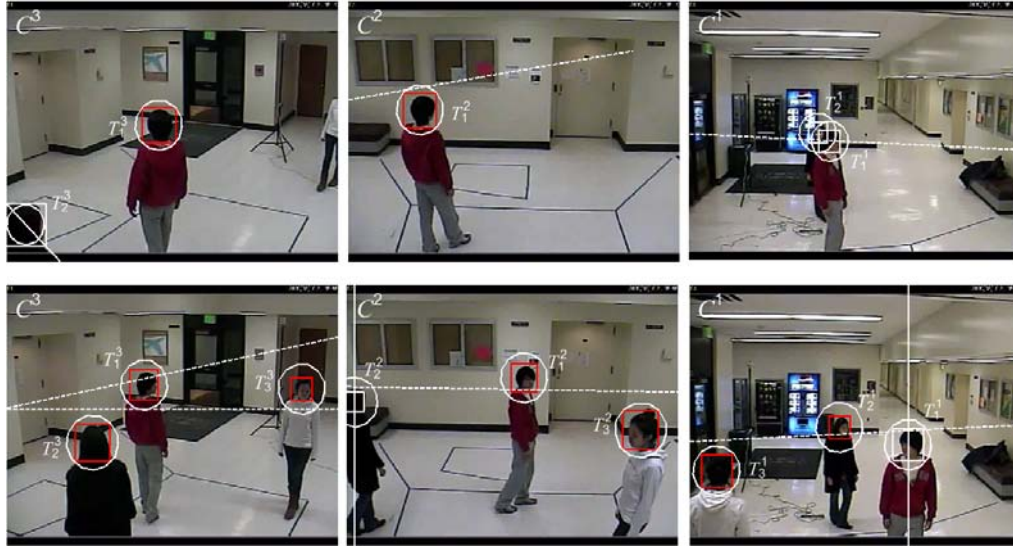## 2.2 Problem Description and Approach

Fig. 2. Illustration of an unsuccessful association of targets due to insufficient separation in a non-ideal situation, and its propagation creating the inconsistent information by a false association (the frame difference between the first frames (top) and the second frames (bottom) is 9 *frames*)

The successful objects association depends on the separation between homographic lines on the other cameras. When the sufficient separation between homographic lines is not guaranteed, homographic lines can cross multiple targets and the intersections with multiple targets can create ambiguity in determining the correspondence of objects. Moreover, detection uncertainty as well as lack of common reference may create uncertainty in deciding

the intersections. In order to guarantee the correct intersections of homographic lines with corresponding targets, a tolerance circle, $s_{\min}$, is defined for each target by considering the effect of targets height uncertainty, frame synchronization errors and detection uncertainty. In **Fig. 2**, a white circle around each target denotes its tolerance circle, a red rectangle an associated target and a white rectangle an unassociated target. Because the required separation for objects association increases, the association performance is affected by the tolerance circle of targets. Due to the insufficient separation in the first frames, homographic lines are not generated in camera C^1 and targets remain unassociated with corresponding targets in the other cameras.

When many cameras (i.e. more than two cameras) are involved in the object association, the locally generated homographic lines based association method can be extended to support them through pair-wise collaboration. However, each pair of collaborating cameras may contradict the object association due to the insufficient separation. It may generate false associations and create the inconsistent information of uncertain association in the system. They are also propagated in time and continuously degrade the association performance. Although only one object is not associated in one instant, it affects the overall performance of object association in time. For example, in **Fig. 2**, targets are unassociated during 9 frames and they create a false association between targets T_1^1 and T_2^2 in the second frames. The system keeps the inconsistent information of the falsely associated targets until they leave the surveillance region. Moreover, when multiple cameras collaborate on objects association without the sufficient separation between targets, homographic lines are unnecessarily generated on targets and projected to the other cameras without establishing objects association. The association failure due to the insufficient separation wastes the computational costs of transforming homographic lines. In a distributed camera network, minimizing data to be transferred is also very important to efficiently exchange data between the camera systems without any data loss or latency. It also increases the frame rate achieved by the system and improves the performance of algorithms. In general, each target requires transformation of a local homographic line to a global homographic line or transformation of a global homographic line to a local homographic line. The number of the intersection tests is proportional to the multiplication of the number of targets and the number of transformed homographic lines. It is assumed that a pair-wise association is utilized when more than two cameras are used. Then, the computational costs for using homographic lines are represented by the number of transformations and intersection tests, C_T and C_C, respectively

$$\begin{aligned} C_T &= 2I \times 2(K-1), \\ C_C &= I^2 \times 2(K-1), \end{aligned} \tag{1}$$

where I denotes the number of detected targets in K cameras. **Fig. 3** shows the computational costs for using homographic lines according to the number of targets and the number of cameras. The number of the transformation increases linearly to the number of targets and the number of intersection tests increases exponentially to the number of targets. It also exponentially increases the amount of exchanged data for target information on the network. As the performance of object association is improved, the number of targets to be associated is decreased at one time. Eventually, it decreases the amount of the exchanged data in the long term. Thus, a proper collaboration camera strategy is necessary to efficiently associate targets among different cameras as minimizing the inconsistent and uncertain information.
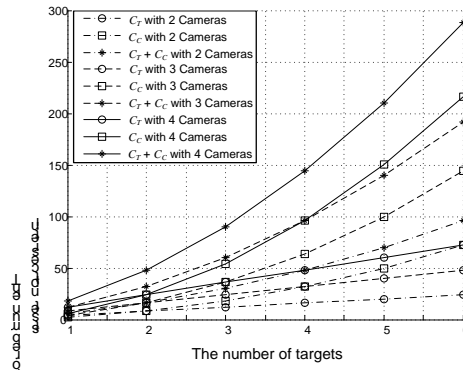
Fig. 3. Illustration of the computational costs for using homographic lines according to the number of targets and cameras

We consider two effective collaboration strategies to reduce false associations as well as to improve the efficiency of using homographic lines. Collaboration matrices are defined to indicate the feasibility of successful association between any two cameras. The elements in the collaboration matrices represent the required minimum separation obtained by incorporating the effect of targets height uncertainty, frame synchronization errors and detection uncertainty. The first strategy uses the collaboration matrices to select the best pair out of many cameras by using the degree of separation between homographic lines. After targets in the selected cameras are associated, the association information is propagated to unselected cameras by transforming the global information constructed from the associated targets.

However, when a large number of objects is detected, the selected cameras cannot cover all the targets and the required minimum separation is hardly satisfied due to overlapped and occluded targets. The second strategy initiates the collaboration process of objects association for all the pairing cases of cameras regardless of the separation. In each pair of cameras, a homographic line is generated on each target and it is projected to the camera. Then, the other collaborating camera generates homographic lines on only the intersected targets with the projected homographic lines and they are re-projected to the one camera. Since the system tests the association for each target at a time, it minimizes association ambiguity caused by homographic lines generated from all the targets. This association process is iteratively operated for all the unassociated targets in all the pairing cases of cameras.

3. Multiple Camera Collaboration

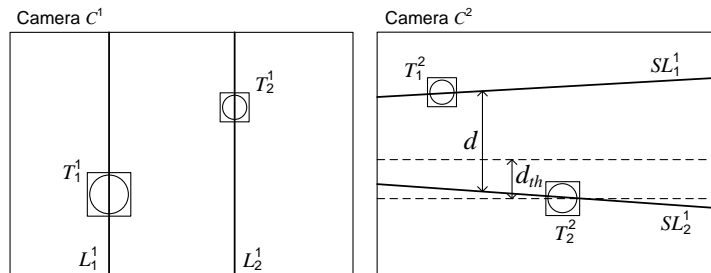### 3.1 Collaboration Matrices and Characterization

Fig. 4. The required separation of homographic lines to be effective for objects association in the other collaborating camera with the known reference plane such as a ground plane

Two collaborating cameras participate in an association process by generating a homographic line on each target in each camera. The homographic line is transformed to a global reference plane such as a ground plane and its transformed homographic line is projected to the other collaborating camera. While the system knows each target that generates its homographic line in a local camera, its corresponding target in the other collaborating camera is determined by the intersection with its projected homographic line. The association of the corresponding targets is established when a projected homographic lines intersects a corresponding target in each camera. However, the association is not established if a projected homographic line intersects multiple targets in each camera. Thus, the separation of projected homographic lines is a key parameter determining the successful association and the size of a target determines the required separation of projected homographic lines. For example, homographic lines are generated on each target in camera C^1 in **Fig. 4** and the distance between projected homographic lines is denoted by d in camera C^2. In order for projected homographic lines SL_1^1 and SL_2^1 to be effective in camera C^2, the separation of them should be greater than the twice size of targets to be associated. We denote the required minimum separation of homographic lines as a threshold and it is represented by d_{th}. Since the threshold can be used to indicate the effectiveness of projected homographic lines, the determination of an appropriate threshold is critical in homographic lines based association.

The critical issue in determining the threshold is that the separation of projected homographic lines depends on the location and the orientation of cameras. Also, if cameras are flexibly titling and panning, the threshold should incorporate the effect of tilting and panning on the separation. When all the possible variations for a camera configuration are considered, the threshold needs to be determined for each case of a camera configuration. Since it is not trivial to construct all the thresholds according to all the camera configurations, we consider only the worst effect among them on the separation. The threshold is determined for each pair of cameras by measuring the maximum size of targets and the threshold matrix with the threshold for each pair of cameras is represented by

$$\boldsymbol{D}_{th} = \begin{bmatrix} d_{th}^{1,1} & d_{th}^{1,2} & \cdots \\ d_{th}^{2,2} & d_{th}^{2,2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix},$$

where its size is K×K and $d_{th}^{k,l}$ denotes the threshold between camera C^k and C^l. A negative value indicates that the threshold is always satisfied in a local camera because a system knows which targets generate homographic lines.
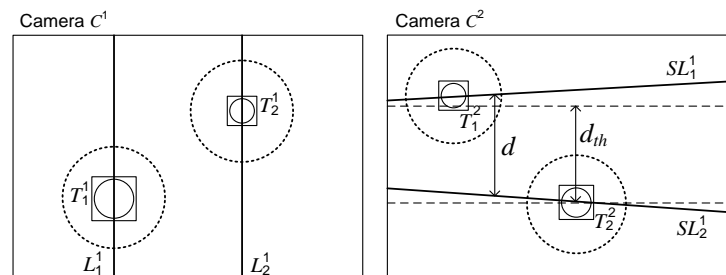

Fig. 5. The required separation of homographic lines to be effective for objects association in the other collaborating camera with the unknown heights of objects and the frame synchronization errors between cameras

There are two additional factors influencing the threshold which increases the size of a target. One is the unknown heights of objects and another is the frame synchronization errors between cameras in a real situation. The dotted circle of targets represents the tolerance circle incorporating the effect of them as shown in **Fig. 5.** Due to the effect of the additional factors, d is increased and may be smaller than d_{th} in the figure. The size of the tolerance circle is determined by measuring the amount of pixels that can be deviated from a corresponding target by the additional factors.

For the unknown heights of objects, the size of a tolerance circle is defined by

$$s_{min,i}^k = \max(r_{b,i}^k, \sigma_{h,i}^k),\tag{2}$$

where $r_{b,i}^k$ denotes the original size of a target and $\sigma_{h,i}^k$ denotes the possible number of deviated pixels from the centroid of a target by mismatched height of a target in camera C^k. $\sigma_{h,i}^k$ depends on camera configurations (locations, tilting angles and panning angles) generating homographic lines. Then, $\sigma_{h,i}^k$ is represented by

$$\sigma_{h,i}^k = \max_{l \neq k}(\sigma_{h,i}'^{k,l}),\tag{3}$$

where $\sigma_{h,i}'^{k,l}$ denotes the maximum number of deviated pixels from the centroid of a target when homographic lines are generated from camera C^l to C^k with possible camera configurations. We use the reference plane with the average height of targets to transform and project homographic lines to a different camera since the actual heights of targets are unknown. When a homographic line is transformed to the reference plane and its transformed homographic line is projected to a different camera, a projected homographic line is deviated from the point of a target with the actual height. In order to measure the amount of pixels to include the deviation by the height uncertainty, it is assumed that only the height range of targets is given to the system. The amount of pixels is measured by comparing a projected homographic line using the average height with a projected homographic line using the maximum or the minimum height because the heights of targets are unknown.
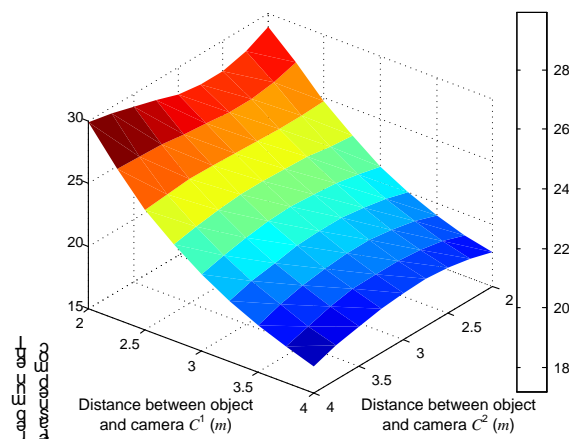

Fig. 6. Illustration of the effect of height uncertainty between cameras C^1 and C^2 according to the distance to an object

Each value of $\sigma_{h,i}'^{k,l}$ between cameras C^k and C^l is determined by finding the maximum

effect of the given amount of height uncertainty according to object locations. **Fig. 6** shows an example how many pixels are deviated from an original point between cameras C^1 placed at ($x=3m$, $y=0m$, $z=3m$) and C^2 placed at ($x=6m$, $y=3m$, $z=3m$). In order to measure the number of deviated pixels, two homographic lines are generated from camera C^2 to camera C^1 according to an object's location. One homographic lines is generated with an actual height and the other is generated with an average height different from the actual height by 0.1m. The number of deviated pixels is maximized when an object is close to a camera onto which homographic lines are transformed. The simulation to measure the amount of pixels is repeated for different camera configurations and the maximum value among them is selected. **Fig. 7** illustrates the amount of pixels with other cameras incorporating the effect of height uncertainty (0.1m ~ 0.4m) where $\sigma'^{k,l}_{h,i}$ denotes the amount of pixels. They are proportional to the amount of height uncertainty.
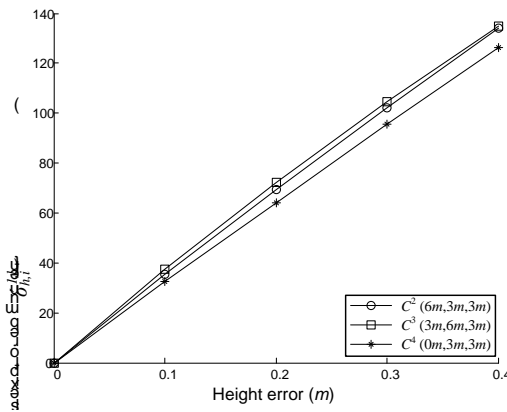

Fig. 7. Illustration of the number of pixels required to compensate for the effect of height uncertainty in camera C^1 placed at ($x=3m$, $y=0m$, $z=3m$) with other cameras (image size: 704×480)
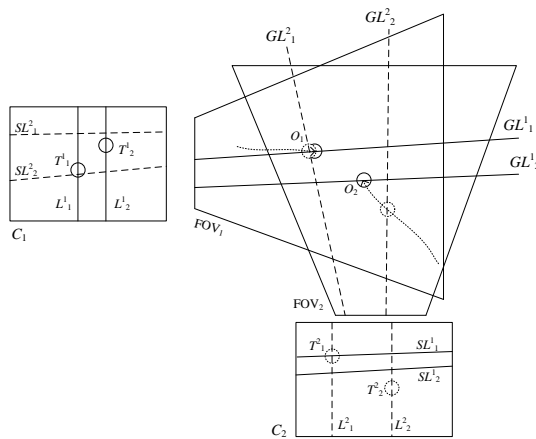

Fig. 8. Illustration of an association problem by frame synchronization errors between cameras (solid circles indicate the targets detected by camera C^1 and dotted circles the targets detected by camera C^2 with frame synchronization errors)

Another factor that influences the size of the tolerance circle is frame synchronization errors between cameras. It also causes the deviation of a projected homographic line because targets can be different locations due to differently captured time. It is assumed that the differently captured time of cameras is within at most 1 frame. **Fig. 8** shows the example of the frame synchronization errors. Solid circles and dotted circles represent two different locations where objects are detected by cameras. Object O_2 cannot be associated because SL_2^1 does not intersect with T_2^2. In order to incorporate frame synchronization errors, the radius of a tolerance circle needs to be adjusted by

$$s_{min,i}^k = \max(r_{b,i}^k, \sigma_{s,i}^k),\tag{4}$$

where $\sigma_{s,i}^k$ denotes the possible number of deviated pixels from the centroid of a target by the synchronization issue between cameras in camera C^k. The synchronization effect depends on the sampling period of a camera T^F and the direction toward which an object moves. Then, $\sigma_{s,i}^k$ is obtained by

$$\sigma_{s,i}^k = \max_{l \neq k}(\sigma_{s,i}'^{k,l}),\tag{5}$$

where $\sigma_{s,i}'^{k,l}$ the maximum number of deviated pixels from the centroid of a target by the synchronization issue between cameras C^k and C^l. Since the effect is maximized when the optical axes of two cameras are perpendicular to each other, only the paired cases of perpendicular cameras is considered.
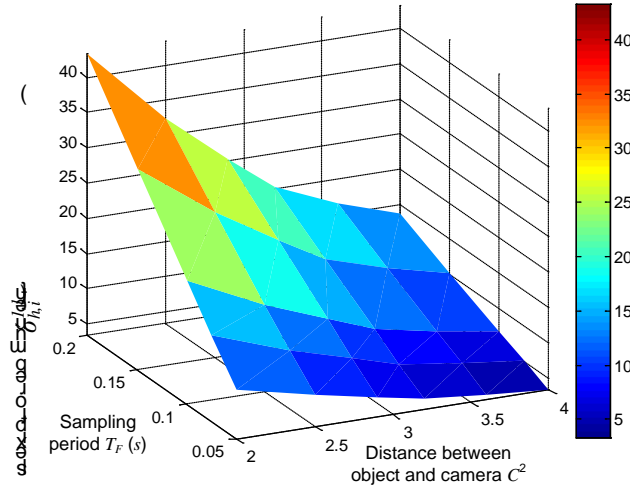

Fig. 9. The simulated effect of frame synchronization errors between cameras

Fig. 9 shows the amount of pixels to incorporate the effect of frame synchronization errors where $\sigma_{s,i}'^{k,l}$ denotes the amount of pixels. It is noted that the deviation of a projected homographic line is related to the relative speed of an object per each frame, not the absolute speed of an object. When the deviation of a projected homographic line is estimated by the simulation, the speed of an object is set to $2m/sec$. Since the frame rate varies from $0.05sec$ to $0.2sec$ in the simulation, it has the same effect of having the relative speed of an object,

$0.1m/frame$ to $0.4m/frame$. Since the frame synchronization errors are maximized with perpendicularly placed cameras, a homographic line is generated from the delayed image of C^1 to the image of C^2. The maximum pixel distance error between a projected homographic line and a corresponding target in camera C^2 is measured. As the sampling rate increases, the amount of pixels decreases.

When the effects of the additional factors are considered at the same time, they can compensate for each other. For example, when a homographic line is generated from the deviated position of a target by detection algorithm, a transformed homographic line with an average height of a target can be accidentally shifted to the position with an actual height of a target. A similar effect can also occur with synchronization issues. However, the system cannot predict the compensation effect by non-ideal parameters. Thus, the size of a tolerance circle should consider the worst effect by

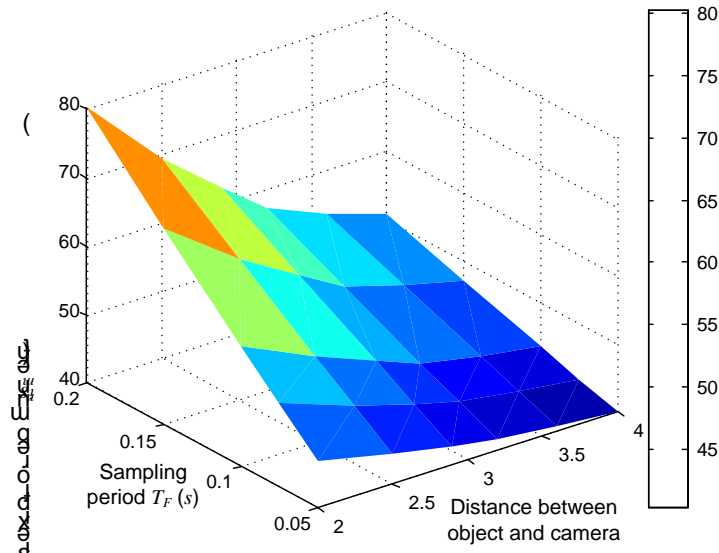$$s_{min,i}^k = \max(r_{b,i}^k, \sigma_{h,i}^k + \sigma_{s,i}^k). \tag{6}$$

Fig. 10. Illustration of the number of deviated pixels by (6)

Fig. 10 illustrates s_{min} for cameras C^1 and C^2 in terms of the number of required pixels by (6) assuming that the height uncertainty is set to be $0.1m$. If the effect of detection performance is considered, the values are expected to be increased. While expanded radii of association circles guarantee that targets are crossed by corresponding transformed homographic lines, they can degenerate association performance because homographic lines generation can be ineffective due to insufficient separations of targets. A threshold indicating effectiveness of homographic lines can be represented by radii of association circles. A threshold d_{th}^k in camera C^k is defined as

$$d_{th}^k = arg \min_{i,j}(s_{min,i}^k + s_{min,j}^k), \tag{7}$$

where i, j denote indices of neighboring targets. The smallest sum of the sizes of two neighboring tolerance circles indicates effectiveness of homographic lines in an association process in camera C^k.
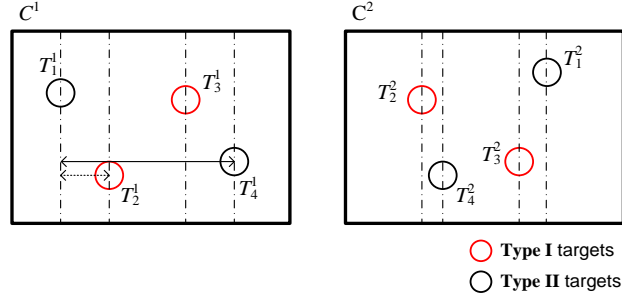
3.2 Camera Selection Based Approach

Fig. 11. Illustration of grouping targets by using the local tracking and association information

A camera selection based approach is to select the best pair out of many cameras to increase the effectiveness of projected homographic lines. Since the effectiveness of projected homographic lines depends on the threshold in the other collaborating camera, a system tests the separation of homographic lines for each pair of cameras. Each camera determines the shortest distance between neighboring homographic lines to be tested for the effectiveness in the other collaborating camera. If the transformed shortest distance has enough separation in the other collaborating camera, the separations of other homographic lines are also satisfied since the transformation of a homographic line is a linear process.

In order to reduce the dependence of the shortest distance on the separation of targets, targets are grouped into two types as shown in **Fig. 11**. Type I is a newly detected target and Type II is a locally tracked target having the association. Each of the shortest distances is determined for each type of targets. If they are not grouped, the shortest between targets in camera C^1 is the distance between targets T_1^1 and T_2^1 otherwise, the distance increases to the distance between targets T_1^1 and T_4^1 for Type I targets. G^k denotes a set of Type I targets and $\tilde{G}^k$ denotes a set of Type II targets in camera C^k. Then, sets for the targets are represented by,

$$G^1 = \{T_1^1, T_4^1\}, \tilde{G}^1 = \{T_2^1, T_3^1\},$$
$$G^2 = \{T_1^2, T_4^2\}, \tilde{G}^2 = \{T_2^2, T_3^2\}.$$

The association processes are operated on sets of equivalent types by using the homographic line based association.

$$H^{1,2}(G^1, G^2), \tilde{H}^{1,2}(\tilde{G}^1, \tilde{G}^2),$$

where function $H^{m,n}$ is the homographic line based association for Type I targets between camera C^m and C^n and function $\tilde{H}^{m,n}$ for Type II targets. H and $\tilde{H}$ are equivalent but the type of targets are different. The target grouping also decreases the number of intersection tests by projected homographic lines. When the targets are not grouped in this example, $C_T + C_C$ is 16 + 32 = 48. Otherwise, C_C decreases to 16 with grouped targets and $C_T + C_C$ becomes 32.
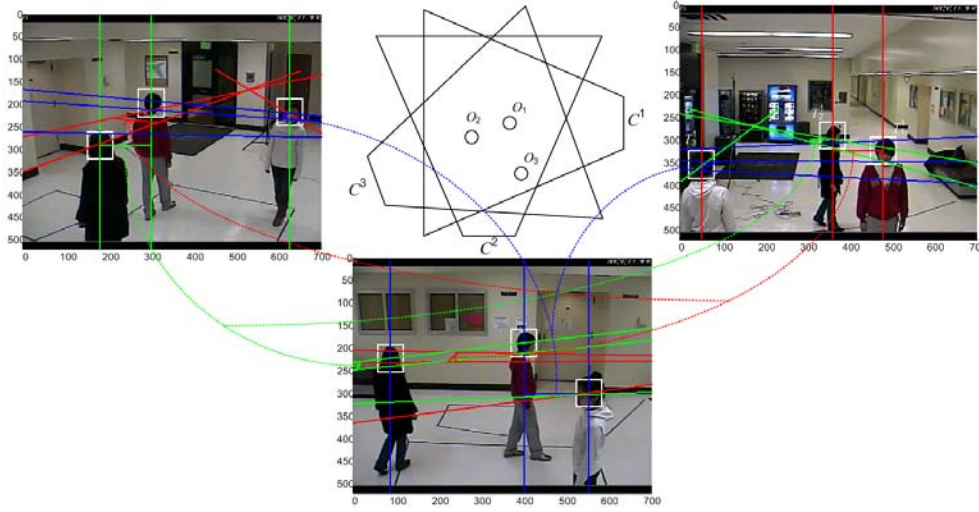
Fig. 12. Illustration of the camera selection strategy in multiple cameras (red homographic lines from C^1, blue homographic lines from C^2, and green homographic lines from C^3)

When the shortest distance between neighboring homographic lines is determined in each camera, it is represented as a starting point and an ending point to be projected to the other cameras. It is assumed that an image has the top left origin. The x-coordinates of the starting point and the ending point are the same as the x-coordinates of the targets respectively and the y-coordinates are set to be the greater of y-coordinates of targets. The reason for selecting the greater of the two y-coordinates is that a closer line to a camera has a shorter length when it is transformed and the shorter one should be tested for the separation in the other cameras. For example, the coordinates of two targets are (482, 323) and (363, 289) in camera C^1 of **Fig. 12**. The shortest distance between them consists of two points (363, 482) and   (482,   323). $d_{i,j}^k$ denotes the length of the shortest line between neighboring targets T_i^k and T_j^k, and $d_{i,j}^{k,l}$ denotes the transformed length of $d_{i,j}^k$ at camera C^l. The length of the shortest line for each pair of cameras can be represented by a matrix with the size of $K \times K$ for convenience. In this example, the distance matrix $\mathbf{D}$ is obtained by

$$\mathbf{D} = \begin{bmatrix} 119 & 32.177 & 112.8608 \\ 48.2162 & 153 & 19.7655 \\ 59.1755 & 16.9171 & 121 \end{bmatrix}.$$

The distance values on diagonals of this matrix are the pixel distance of two targets in each local camera and others are projected pixel distances in the other cameras. If only one object is detected by a camera, distance is ∞.   Since this matrix is determined by the coordinates of targets, it is updated every frame.

The threshold matrix $\mathbf{D}_{th}$ for **Fig. 12** is obtained by using the simulated data from **Fig. 7** and **Fig. 9**

$$\mathbf{D}_{th} = \begin{bmatrix} -1 & 74.46 & 87.23 \\ 76.61 & -1 & 64.27 \\ 82.70 & 74.46 & -1 \end{bmatrix},$$

where the height error is set to 0.1m and the frame rate is set to 8*frames/sec*. When D and D_{th} are compared, a pair of C^1 and C^3 is possibly selected to cooperate for objects association. If multiple choices are possible, a pair of cameras having the maximum difference between D and D_{th} can be chosen.

Targets in unselected cameras are associated by global information constructed by the associated targets in selected cameras. When targets are associated in two selected cameras, the system constructs global information such as height and position. **Fig. 13** illustrates an association information update by selected cameras. When homographic lines generated with the known height from associated targets are transformed to other unselected cameras, they should intersect at near another corresponding target. $g^{\tilde{k}}(T_i^m, T_j^n)$ denotes an intersection point of homographic lines on camera $C^{\tilde{k}}$ from associated targets T_i^m and T_j^n where $\{m,n\} \in K$. If several unassociated targets exist in unselected cameras, a target having the minimum distance to the transformed point is associated with $\{T_i^m, T_j^n\}$. Then, index $\tilde{i}$ of unassociated targets to be associated with them is determined by,

$$\arg \min_{\tilde{i}} D(g^{\tilde{k}}(T_i^m, T_j^n), T_{\tilde{i}}^{\tilde{k}}), \tag{8}$$

where $D(a,b)$ returns the distance between points a and b. This may falsely associate targets when they are occluded by each other. Hence, if an intersection point is within the tolerance circle of more than two targets, association information is not updated to prevent from false association. Even with these strategies, false association is still possible when targets are occluded each other. However, their association can be confirmed after they are separated enough for the association. For the computational costs, this process requires one global to local transformation and the intersection tests by the number of targets in (8). The first strategy by the camera selection based approach is summarized in Algorithm 1.
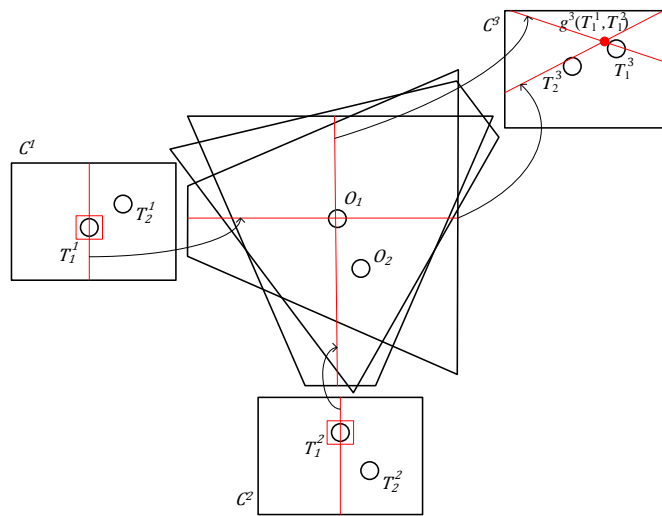

Fig. 13. Illustration of an association information update in unselected cameras by using global information.

Algorithm 1: Camera selection based approach

Input : Detected targets at each camera C^k, threshold matrix D_{th}
Output : Associated targets
repeat
   Classify targets into Type I and Type II targets by local tracking information in each camera C^k
   Construct distance matrix D with the minimum distance between targets in each camera C^k
   Select the pair of the best cameras $C_p = \{(C^m, C^n) \mid m, n \in k\}$ based on D and D_{th}

   for $C^k \in C_p$ **do**
     for all T_i^k **do**
       if T_i^k is Type I target **then**
         Generate a vertical homographic line and transform it to other selected camera
       end
     end
   end
   Find associated targets in two selected cameras with homographic lines and s_{min}
   Append association information into associated objects list **A**
   for A $\in$ **A** **do**
     for $C^k \notin C_p$ **do**
       for all T_i^k **do**
         if T_i^k is Type I target **then**
           Two homographic lines from A are transformed onto C^k and check association
with
           s_{min} by (8)
         end
       end
     end
   end
until System stops

3.3 Iteration Based Approach

The association performance of the camera selection based method is ineffective due to a large number of targets. Any elements of matrix D may not be satisfied with the threshold matrix D_{th}. A distance matrix D for **Fig. 14** is obtained by

$$D = \begin{bmatrix} 4 & 0.56 & 2.63 \\ 6.08 & 16 & 6.61 \\ 16.77 & 6.07 & 34 \end{bmatrix}.$$

The figure shows the ineffectiveness of homographic lines for associating targets when any pairs of cameras do not satisfy the threshold matrix D_{th}. Thus, the system needs to wait until thresholds are satisfied. This may cause undetermined delay for the association process.
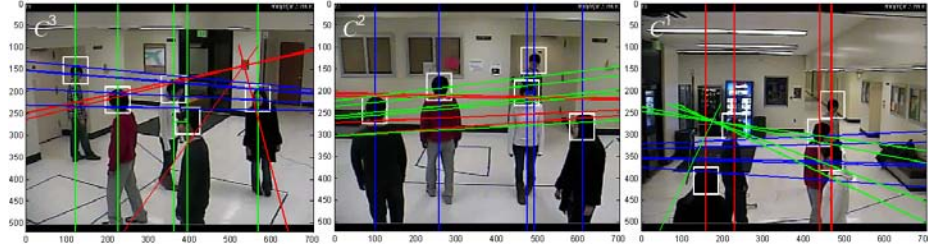
Fig. 14. Illustration of a case in which any pairs of cameras do not satisfy thresholds (red homographic lines from C^1, blue homographic lines from C^2, and green homographic lines from C^3)

Moreover, the camera selection based method does not always provide a correct decision in selecting cameras because the average height of targets is used to transform and project a horizontal line to other cameras. Also, since projected homographic lines are not parallel in other cameras, the association process is not always successful. These limitations waste projected homographic lines without establishing objects association. For example, although cameras C^2 and C^3 are selected by the satisfied threshold in **Fig. 15**, any targets are not associated due to the insufficient separation. Another limitation of the camera selection based method is that some pairs of cameras having possible objects association are disregarded.
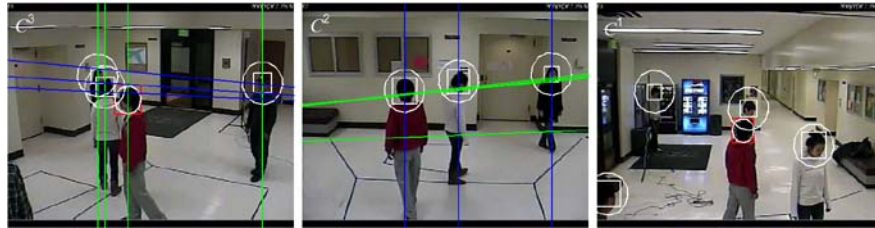

Fig. 15. A limitation of camera selection strategy when cameras C^2 and C^3 are selected for objects association (blue homographic lines from C^2, green homographic lines from C^3 and red retangulars indicate that targets are associated in the previous frame)
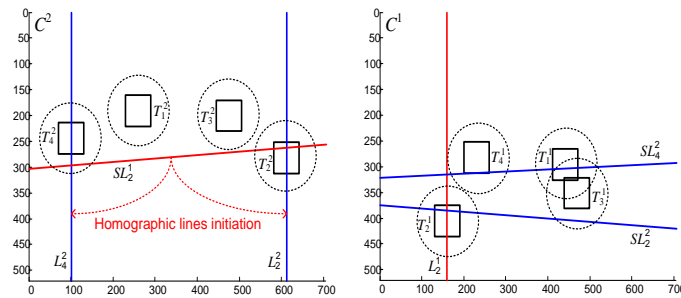

Fig. 16. The illustration of an iteration based strategy

Table 2. Crossed targets by homographic lines for **Fig. 16**

	Crossed targets in C^1	**Crossed targets in C^2**
T_2^1	-	$\{T_2^2, T_4^2\}$
T_2^2	$\{T_2^1\}$	-
T_4^2	$\{T_1^1, T_3^1, T_4^1\}$	-

In order to remedy the limitations of the camera selection based method, a system initiates the collaboration process of objects association for all the pairing cases of cameras regardless of the separation. Similarly to the camera selection based approach, the targets in each camera are grouped by the two types defined in the previous section to minimize the unnecessary association process for the different types of targets. When each pair of cameras is operated for objects association, a homographic line is generated on each target at a time and it is transformed to the other collaborating camera. A homographic line can be generated in any order of targets such as the left to the right on image. After a homographic line is transformed to the other collaborating camera, all the crossed targets by the transformed homographic line generate homographic lines. The association is established for only the targets generating homographic lines. **Fig. 16** illustrates an example of a homographic line generation on one target for two cameras. Camera C^1 generates a homographic line on only target T_2^1 first. Its homographic line is projected to camera C^2 and camera C^2 generates each homographic line on the intersected targets T_4^2 and T_2^2. Since only the two of the four targets participate in an association process for target T_2^1, it has a higher chance that the projected homographic lines have the sufficient separation in camera C^1. **Table 2** illustrates the crossed targets by homographic lines of an iteration based strategy for **Fig. 16**. As a result, targets T_2^1 and T_2^2 are associated. This process is repeated for each target in each pair of collaborating cameras. The second strategy by the iteration based approach is summarized in Algorithm 2.

Unassociated targets by the camera selected based approach in **Fig. 15** are associated by the iterative association approach in **Fig. 17**. Since three cameras are used, six iterated association cases exist (C^1 to C^2, C^1 to C^3, C^2 to C^3, and vice versa). One additional association for targets is established in **Fig. 17-(b)**. Although the iterative association approach has an advantage to check every association case, the computational costs of the transformation and the intersection test are much higher than those of the camera selection based association.

Algorithm 2: Iteration based approach

Input : Detected targets at each camera C^k
Output : Associated targets
repeat

    Classify targets into Type I and Type II targets by local tracking information in each camera C^k

    Construct all the possible pairing cases $\boldsymbol{C}_p = \{ (C^m, C^n) \mid m, n \in k \}$

    for $C_p \in \boldsymbol{C}_p$ **do**

      for $C^k \in C_p$ **do**

        for all T_i^k **do**

          if T_i^k is Type I target **then**

            Generate a vertical homographic line and transform it to the other collaborating camera

            Check association with $s_{\min}$ and newly initiated homographic lines

            if single Type I target in the collaborating camera is associated with T_i^k **then**

              Append association information into associated objects list **A**

            end

          end

         end
      end
    end
  for $A \in \mathbf{A}$ **do**
    for all C^k **do**
      for all T_i^k **do**
        if T_i^k is Type I target **then**
          Two homographic lines from A are transformed onto C^k and check association with $s_{\min}$ by (8)
        end
      end
    end
  end
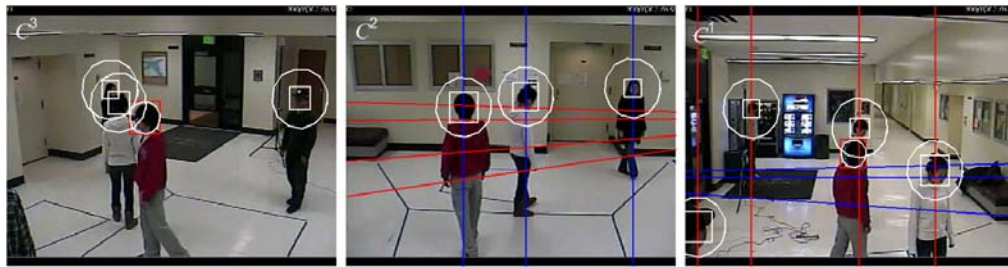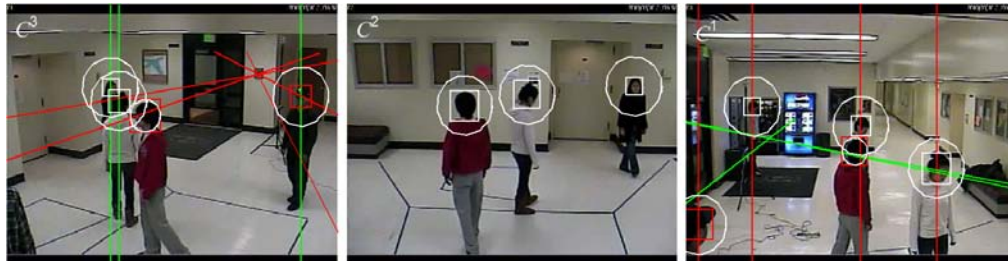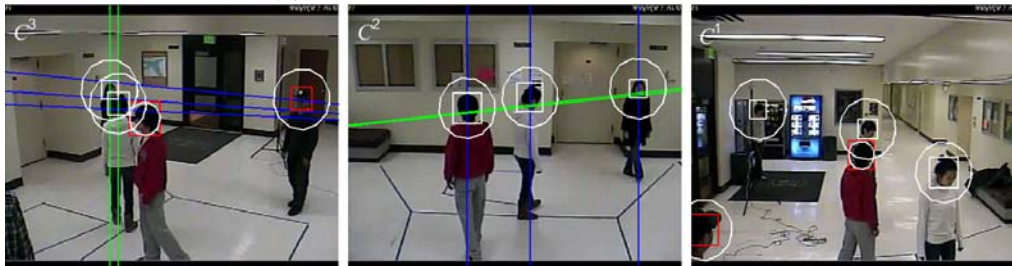**until** System stops

(a) association between cameras C^1 and C^2

(b) association between cameras C^1 and C^3

(c) association between cameras C^2 and C^3

Fig. 17. Illustration of the multiple objects association by the iterative association approach (red homographic lines from C^1, blue homographic lines from C^2, green homographic lines from C^3 and red retangulars indicate that targets are associated in the previous frame)

4. Simulation and Analysis

4.1 Simulation Setup

Fig. 18 illustrates a simulation setup with six objects and three cameras for analyzing the association performance and the computational costs of the transformation and the intersection tests. We also compare the proposed methods with basic pair-wise collaboration extended from [16] to prove the improvement on the false association rate. Because the basic pair-wise approach initiates the association process for each pair of participating cameras, three different pairs of cameras execute the association process redundantly. In order to compare the performance between the methods, we measure three different rates such as successful association rate, failed association rate, and false association rate to check the inconsistency. Camera C^1 is placed at (x = 3.65m,y = 0m,z = 2.37m) with tilting angle 82.3°and panning angle 0°, camera C^2 is placed at (x = 0m,y = 3.5m,z = 2.45m) with tilting angle 76°and panning angle 90°, and camera C^3 is placed at (x = 1.83m,y = 7.32m,z = 2.37m) with tilting angle 78°and panning angle 156°. The total number of frames is 150 and the average height of targets is 1.7m. Another important issue is how to locally track targets in each camera. When targets (i.e., faces) are occluded each other by three fourths of their size, they fail in the local tracking and also lose association information. Since Type II targets are already associated, only Type I targets are considered for the association in the simulation.
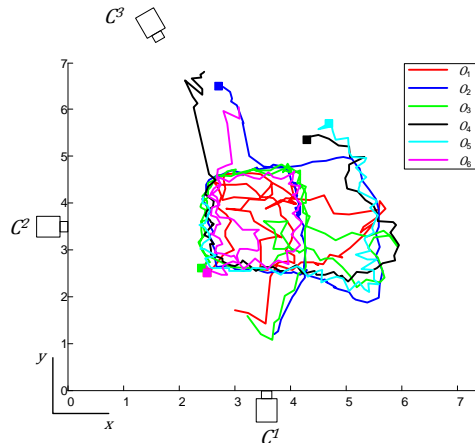

Fig. 18. Illustration of a simulation setup with six objects trajectories and three cameras

4.2 Association Performance and Complexity Comparison

The first collaboration strategy selects a pair of cameras satisfying the threshold. The values for D_{th} are the same as the values used in Section 3. The shortest distance between homographic lines in each camera is determined and its projected distance is compared with the corresponding element of D_{th} at every frame. **Fig. 19** shows the variation of each element of distance matrix D with respect to a corresponding element of threshold matrix D_{th} for the camera selection based approach. If the value of $D(m,n)/D_{th}(m,n)$ is greater than 1, the transformed minimum distance is satisfied with the threshold for a pair of corresponding cameras C^m and C^n. When one unassociated target is remained in each camera, the distance between targets is set to be the width of image instead of the infinity for the simulation. **Fig. 20**

shows the result of selected cameras at each association time with the results of **Fig. 19**. If any thresholds are not satisfied, none of cameras is selected.
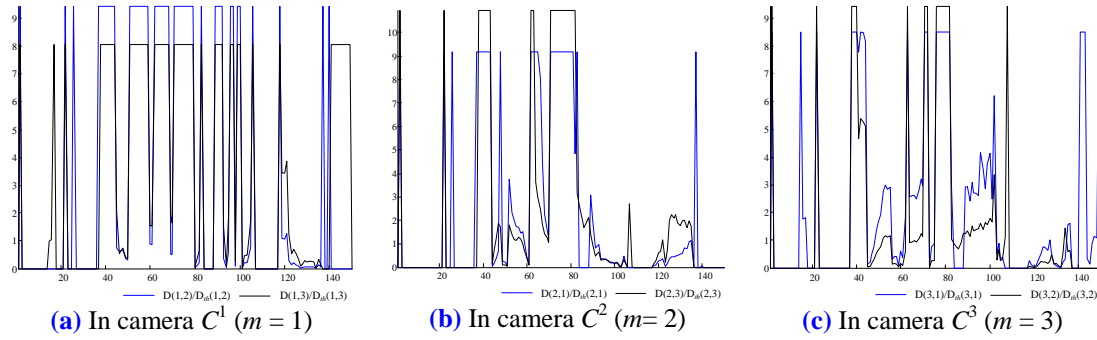

(a) In camera C^1 ($m = 1$)   **(b)** In camera C^2 ($m= 2$)   **(c)** In camera C^3 ($m = 3$)

Fig. 19. Simulation results of $D(m,n)/D_{th}(m,n)$ to show possible cameras for the cooperation at each association time
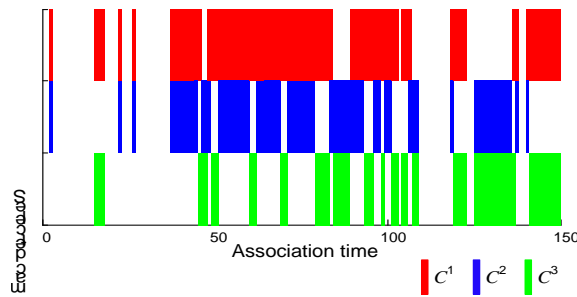

Fig. 20. Selected cameras by the proposed camera selection method at each association time (seletcted cameras are shaded by corresponding colors)

Fig. 21 and **Table 3** show the comparison of the proposed approaches with the basic pair-wise approach. In **Table 3**, we consider only a case that objects are detected by multiple cameras because the collaboration for the object association is not required for a single camera. The performance of the basic pair-wise association approach almost indicates a limit because the scheme considers all the pairing cases of multiple cameras through pair-wise collaboration. However, the effectiveness of the basic pair-wise association is affected by the redundant collaboration of the pair-wise association process among multiple cameras and it creates false associations in the system. In order to clearly show the improvement of the performance, the false associations are not included in successful associations. The average false association rate by the method in [16] is about 6.2% and they are not corrected until they leave the surveillance region. It affects the average successful association rate in time. On the other hand, the iteration based approach reduces the average false association rate by almost a zero and improves the average successful association rate. When the iteration based approach is used for multiple objects association, the number of successful association is greater than that of the camera selection based approach. This is mainly because the camera selection based approach usually selects only a pair of cameras satisfying the threshold and disregards the rest of possible association cases. On the other hand, the iteration based approach initiates the association process for each target and it leads to improve the association performance. Thus, the camera selection based approach requires the longer operation time to increase the association rate than the iteration based approach.
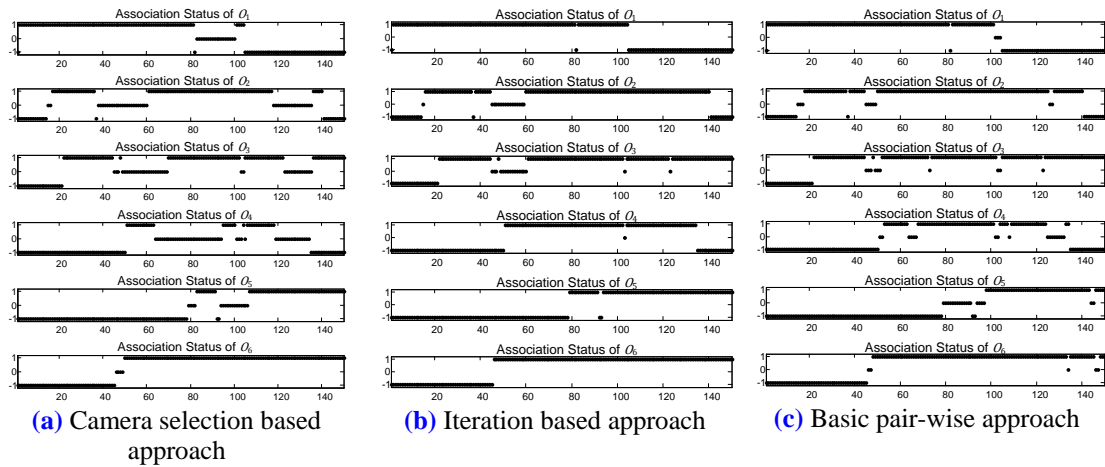
(a) Camera selection based approach   (b) Iteration based approach   (c) Basic pair-wise approach

Fig. 21. The objects association status comparison of the camera selection based approach and the iteration based approach with the basic pair-wise approach at each association time ('1' indicates successful association, '0' failed association, and '-1' that object is not shown on both cameras)

Table 3. Performance comparison for **Fig. 21**

O_i	Camera selection based approach			Iteration based approach			Basic pair-wise approach		
	Success (%)	Failure (%)	False (%)	Success (%)	Failure (%)	False (%)	Success (%)	Failure (%)	False (%)
O_1	82.3	17.7	0.0	100.0	0.0	0.0	88.3	2.9	8.8
O_2	64.9	34.3	0.8	84.8	14.4	0.8	83.3	7.9	8.8
O_3	69.0	30.2	0.8	85.2	14.8	0.0	89.1	7.8	3.1
O_4	39.3	60.7	0.0	91.6	8.4	0.0	75.0	20.2	4.8
O_5	75.8	24.2	0.0	98.5	0.0	1.5	67.0	27.2	5.8
O_6	96.1	3.9	0.0	100.0	0.0	0.0	89.6	4.7	5.7
Avg.	71.2	28.5	0.3	93.4	6.2	0.4	82.1	11.7	6.2

Fig. 22 shows the simulation result to measure the number of transformations (i.e. C_T) and the number of intersection tests (i.e. C_C) by the association approaches according to the number of objects. The results show that the proposed methods are more efficient than the basic pair-wise approach. Also, the number of transformations and intersections tests in the camera selection based approach is lower than that of the iteration based approach. This is because only two selected cameras are participating in the association process. However, in terms of the association performance, the selected cameras cannot cover occluded targets as the number of targets increases. As a result, the association performance is degraded as compared with the iteration based approach in **Fig. 23**. The iteration based approach increases the association performance with the cost of the increased number of transformations and the number of intersection tests.
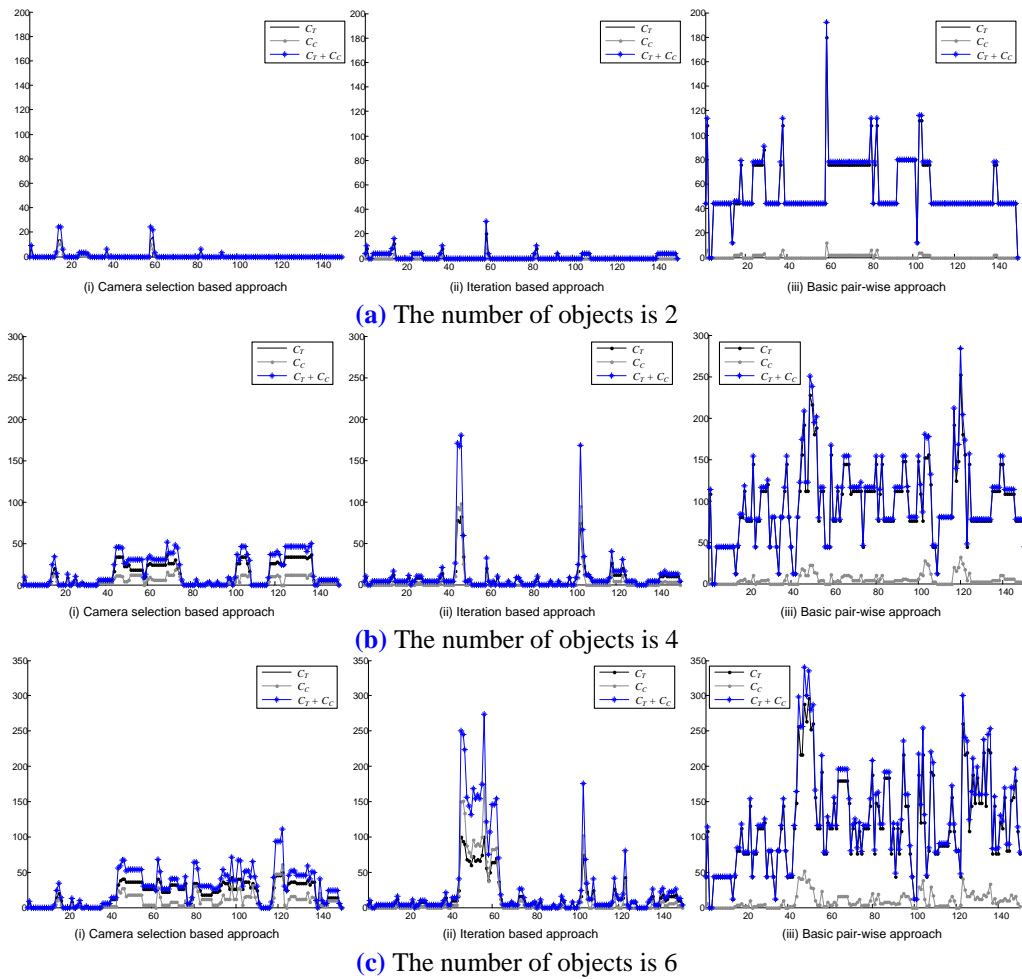
(a) The number of objects is 2

(b) The number of objects is 4

(c) The number of objects is 6

Fig. 22. The computational costs comparison of the camera selection based approach and the iteration based approach with the basic pair-wise approach according to the number of objects
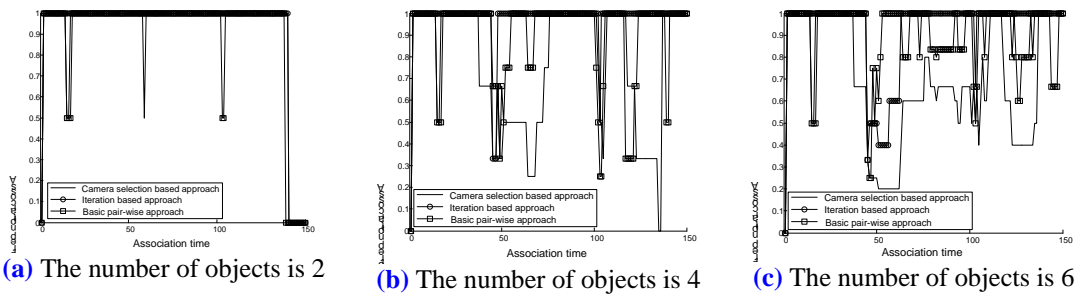

(a) The number of objects is 2

(b) The number of objects is 4

(c) The number of objects is 6

Fig. 23. The association performance comparison of camera selection based approach and iteration based approach with the basic pair-wise approach according to the number of objects

4.3 Discussion

The simulation result shows that the camera selection based approach has the lower successful association rate than the others. However, it is not critical because the result is obtained in the limited amount of time and the objects are associated when the sufficient separation is satisfied. Hence, the successful association rate can be improved as objects move

within the surveillance region in time. A more important issue is to minimize the false association so that the consistent global view of multiple cameras is maintained in the system. Once objects are falsely associated, they are propagated in time and it may continuously generate the inconsistent information through tracking. Especially, when objects are more densely populated than the simulation, there is a high possibility that homographic lines are not effective for objects association due to the insufficient separation. Then, the basic pair-wise collaboration may create false associations more and they can corrupt the consistent information in the system. Thus, the effective association collaboration scheme is important to accurately and effectively maintain association information for the insufficient separation.

5. Conclusions

We present two different strategies for multiple camera collaborations to reduce the false association for the object association. Collaboration matrices are defined with the required minimum separation for each pair of cameras and used to select a pair of cameras having the maximum separation of homographic lines in the first strategy. We have shown that the first strategy reduces the number of transformations and intersection tests using homographic lines for the object association. However, as a large number of objects are detected, it may require the long operation time to achieve the high association rate due to the unsatisfied separation. In order to remedy the limitation of the first strategy, the second strategy initiates the collaboration process of objects association for all the pairing cases of cameras regardless of the separation. The simulation result demonstrates that the association performance is improved by the repetitive association processes while the computational costs of using homographic lines increase exponentially. The comparison simulation with the basic pair-wise approach also shows that the proposed methods reduce the false association effectively.

References

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

[2] S. Velipasalar and W. Wolf, "Multiple object tracking and occlusion handling by information exchange between uncalibrated cameras," in *Proc. of IEEE International Conf. on Image Processing*, pp. 418-421, Sept. 11-15, 2005. Article (CrossRef Link)

[3] J. Orwell, P. Remagnino and G.A. Jones, "Multiple camera color tracking," in *Proc. of IEEE International Workshop on Visual Surveillance*, pp. 14-24, June 26, 1999.

[4] 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 International Workshop on Visual Surveillance*, pp. 3-10, July 1, 2000. Article (CrossRef Link)

[5] 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. on Computer Vision*, pp. 18-36, May 28-31, 2002. Article (CrossRef Link)

[6] J. Li, C.S. Chua and Y.K. Ho, "Color based multiple people tracking," in *Proc. of IEEE International Conf. on Control, Automation, Robotics and Vision*, vol. 1, pp. 309-314, Dec. 2-5, 2002. Article (CrossRef Link)

[7] Y. Caspi, D. Simakov and M. Irani, "Feature-based sequence-to-sequence matching," *International Journal of Computer Vision*, pp. 53-64, June, 2006. Article (CrossRef Link)

[8] Q. Cai and J. K. Aggarwal, "Tracking human motion using multiple cameras", in *Proc. of International Conf. on Pattern Recognition*, Vienna, Austria, vol. 3, pp. 68-72, Aug. 25-29, 1996. Article (CrossRef Link)

[9]   J. Black and T. Ellis, "Multi camera image tracking," *Image and Vision Computing*, vol. 24, pp. 1256-1267, Nov., 2006. Article (CrossRef Link)

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

[11] A. Utsumi, H. Mori, J. Ohya and M. Yachida, "Multiple human tracking using multiple cameras," in *Proc. of IEEE International Conf. on Automatic Face and Gesture Recognition*, pp. 498-503, Apr. 14-16, 1998. Article (CrossRef Link)

[12] J. Kang, I. Cohen, and G. Medioni. "Continuous tracking within and across camera streams," in *Proc. of IEEE International Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 267-272, June 18-20, 2003. Article (CrossRef Link)

[13] T. H. Chang, S. Gong and E.J. Ong, "Tracking multiple people under occlusion using multiple cameras," in *Proc. of British Machine Vision Conf.*, pp. 566-575, Sept. 11-14, 2000. Article (CrossRef Link)

[14] S. Calderara, A. Prati, R. Vezzani and R. Cucchiara, "Consistent labeling for multi-camera object tracking," in *Proc. of International Conf. on Image Analysis and Processing*, pp. 1206-1214, Sept. 6-8, 2005. Article (CrossRef Link)

[15] S. Khan and M. Shah, "Consistent labeling of tracked objects in multiple cameras with overlapping fields of view," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, Oct., 2003. Article (CrossRef Link)

[16] Y. Kyong, S. H. Cho, S. Hong and W. D. Cho, "Local initiation method for multiple object association in surveillance environment with multiple cameras," in *Proc. of IEEE International Conf. on Advanced Video and Signal based Surveillance*, Sept. 1-3, 2008. Article (CrossRef Link)

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 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. in Electrical and Computer Engineering from 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. He was a recipient the Presidential Award for Excellence in Graduate School of Information and Communication in 2004 and 2007. He earned the Best Paper Award at KDC 2006. He is selected for inclusion in the 2011 Edition of "Who's Who in America." 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.