Journal of Universal Computer Science, vol. 15, no. 13 (2009), 2463-2484
submitted: 31/10/08, accepted: 13/6/09, appeared: 1/7/09 © *J.UCS*

Human Tracking based on Multiple View Homography

Dong-Wook Seo
(University of Ulsan, Korea
seodonguk@gmail.com)

Hyun-Uk Chae
(University of Ulsan, Korea
hwchae@islab.ulsan.ac.kr)

Byeong-Woo Kim
(University of Ulsan, Korea
bywokim@ulsan.ac.kr)

Won-Ho Choi
(University of Ulsan, Korea
whchoi@mail.ulsan.ac.kr)

Kang-Hyun Jo
(University of Ulsan, Korea
jkh2009@islab.ulsan.ac.kr)

Abstract: We propose a method for detection and tracking for objects under multiple cameras system. To track objects, one need to establish correspondence objects among multiple views. We apply the principal axis of objects and the homography constraint to match objects across multiple cameras. The principal axis belongs to the silhouette of objects that is extracted by the background subtraction. We use the multiple background model to the background subtraction. In an image sequence, many changes happen with respect to pixel intensity. This cannot be characterized by the single background model so that is necessary to use the multiple background model. Also, we use the median background model reducing some noises. The silhouette is detected by difference with background models and current image which includes moving objects. For calculating homography, we use landmarks on the ground plane in 3D space. The homography means the relation between two correspondence between two coinciding points from different views. The intersection of principal axes and ground plane in 3D space are the same point shown in each view. The intersection occurs when a principal axis in an image crosses to the transformed ground plane from another image. We construct the correspondence which means the relationship between intersection in current image and transformed intersection from the other image by homography constraint. Those correspondences confirm within a short distance measuring in the top viewed plane. Thus, we track a person by these corresponding points on the ground plane.

Key Words: Multiple cameras, Multiple background model, Median background, Homography constraint, Human tracking

Category: H.2, H.3.7, H.5.4

1 Introduction

The systems using multiple cameras are of interest in the computer vision community. These systems apply a various application such as visual surveillance, gait analysis, and gesture recognition system. The multiple cameras can be covering a wide environment. Using multiple cameras, the field of view in the single camera is expanded and information from multiple views is extremely useful to handle many issues. To track objects successfully in multiple perspective images, we need to establish correspondence between objects under multiple cameras system. We need the robustly detected moving objects for those objects individually. Nevertheless it is a simple approach to detect and track moving objects using the single camera; it is the difficult to overcome when objects are continuously existed on the optical axis. The multiple cameras are useful to handle many issues from occlusion in crowded environments. Using multiple cameras, it is possible to expand the field of view. Moreover, it is useful to track or analyze the multiple objects even though some of them are occluded. One of the important points is how to define the correspondence between objects in each view. This paper describes how to correspond between the multiple views. The correspondence of multiple cameras is to find correspondence among simultaneously exited objects in different view and track objects in time sequence.

There are numerous algorithms to detect and track object in multiple cameras system. They used a various method to match objects among multiple views. Most of them used the color distribution as the main cue to correspond objects across multiple views. Generally, the Gaussian color model [Mittal and Davis 2002] and the color histogram [Orwell et al. 1999] are used to overcome the problems of correspondence between multiple views. Also, Morioka et al. [Morioka et al. 2006] use the global color model to match objects under multiple cameras. However the color-based approach gives unexpected results from illumination variance or similarities of color. For instance, if two people put on clothes of similar color are occluded together, we may get an incorrectly result. There also is to get similar result in the case of one person has both front and back of clothes which is different color.

Some researcher used the geometrical constraint to solve a limitation of color-based method. They use the feature points to correspond object across multiple vies. A central point [Utsumi et al. 1998] is used to match objects across views. A central point of objects is taken as a feature point and correspondence is established by estimating the 3D of central point in the world coordinate. Also, Chang and Gong [Chang and Gong 2001] applied a Bayesian networks to match object among multiple views. Also, they establish across views correspondence using combination of epipolar geometry, landmark modality, apparent height, and color. Khan et al. [Khan et al. 2001] used field of view (FOV) constraint to correspond and track object under multiple cameras. They use spatial relationship between camera fields of view and apply this information to correspond between different views. Furthermore, the feature points of object extracted from different views do not always correspond to the same physical the 3D point. Also, the feature point-based method is easily influenced by noise.

In this paper, we focus on the problem of correspondence objects among multiple cameras. We apply the principal axis of objects and the homography constraint to correspond objects among multiple views. In multiple cameras, we detect objects by background subtraction. We apply the multiple background model [Kim and Jo 2007]. In an image sequence, the several variations for each pixel are successfully contained by multiple background model. After detecting object, we compute the principal axis of object using the principal axis moments [Hu 1962]. We apply the homography constraint to match objects across multiple views. The homography defined in 2D space as a mapping between a point on the ground plane seen from one camera, to the same point seen from another camera. We compute the homography using correspondence between some landmarks on the ground plane. Using this information, we establish correspondence objects among multiple cameras. In the real 3D space, we have the only one point that is the intersection of principal axis of object and the ground plane. The intersection is the invariant point in different views. Therefore we estimate the invariant point for each view by a principal axis of human in a view and a transformed principal axis to the same view using the homography. The intersection occurs when the principal axis of object in an image crosses to the transformed ground plane from other image. However, the positions of intersection are depends on field of view of each camera. Therefore we construct the correspondence that means the relationship between the intersection in current image and the transformed intersection from other image by homography. Those correspondences should confirm within a short distance measuring in the top viewed plane. Thus, we track a person by these corresponding points on the ground plane.

This paper is organized as follows. Section 2 presents a process of generating the multiple background model. In section 3, the process of motion detection by background subtraction is explained. The correspondence of objects across multiple views and experimental results are described in section 4 and 5, respectively. Finally, we concluded this paper in section 6.

2 Multiple Background Model

A background is a part where do not vary too much or frequently change in an image sequence. This environment cannot be characterized by one background model. To overcome this problem, we use the multiple background model. Fig. 1 describes the three typical distributions for intensity in a pixel belonged in a background. Those distributions of intensity are taken during 100 frames for each pixel with gray scale.

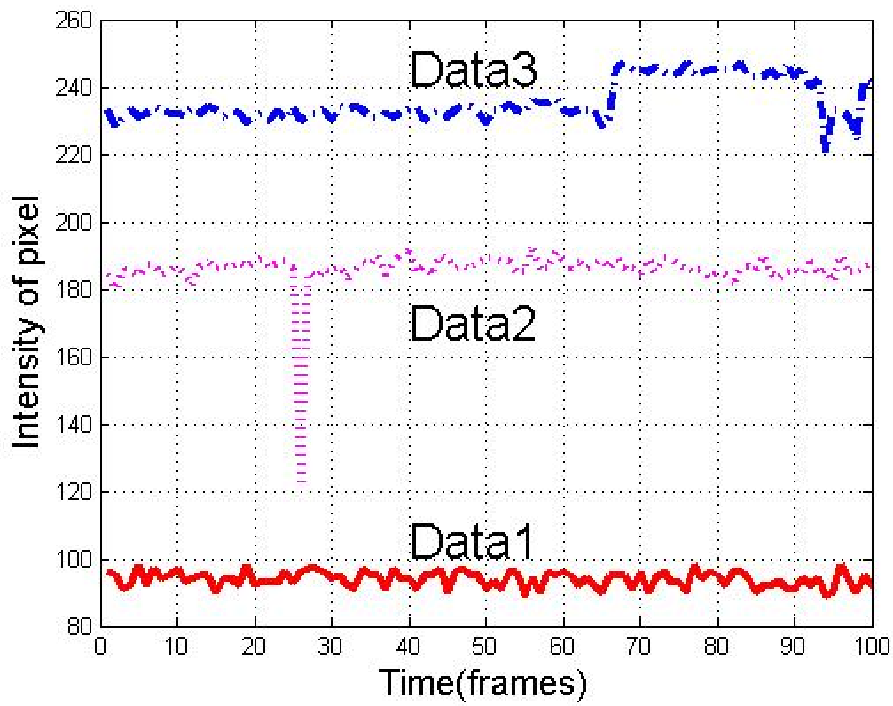


Figure 1: The three types of intensity history in 100 frames

The intensity of pixel is frequently changed with small variance that has the one background model in Data 1. Among 100 frames, Data 1 has a stable variation dislike with Data 2 and Data 3. In Data 2, we can see that an intensity of pixel change greatly in short intervals, e.g., the intervals from 20^{th} to 50^{th} frame. It means moving objects passing through in the pixel. We can see a similar distribution of pixel intensity with Data 1 in the two parts which are 1^{st} to 65^{th} and 66^{th} to 90^{th} frame on the Data3. Those two distributions are different but all of those have a change of pixel intensity over given time that means objects stayed. Above three distributions (Data 1-3) are different each other, however all of those are presented as a background model. To generate background which include those distribution altogether, we should use the multiple background model.

2.1 Background Reconstruction

We introduce a process for constructing the multiple background model (including several steps) in this section. The process has three steps which are online clustering, removing clusters and generating the multiple background model as shown in Fig. 2.

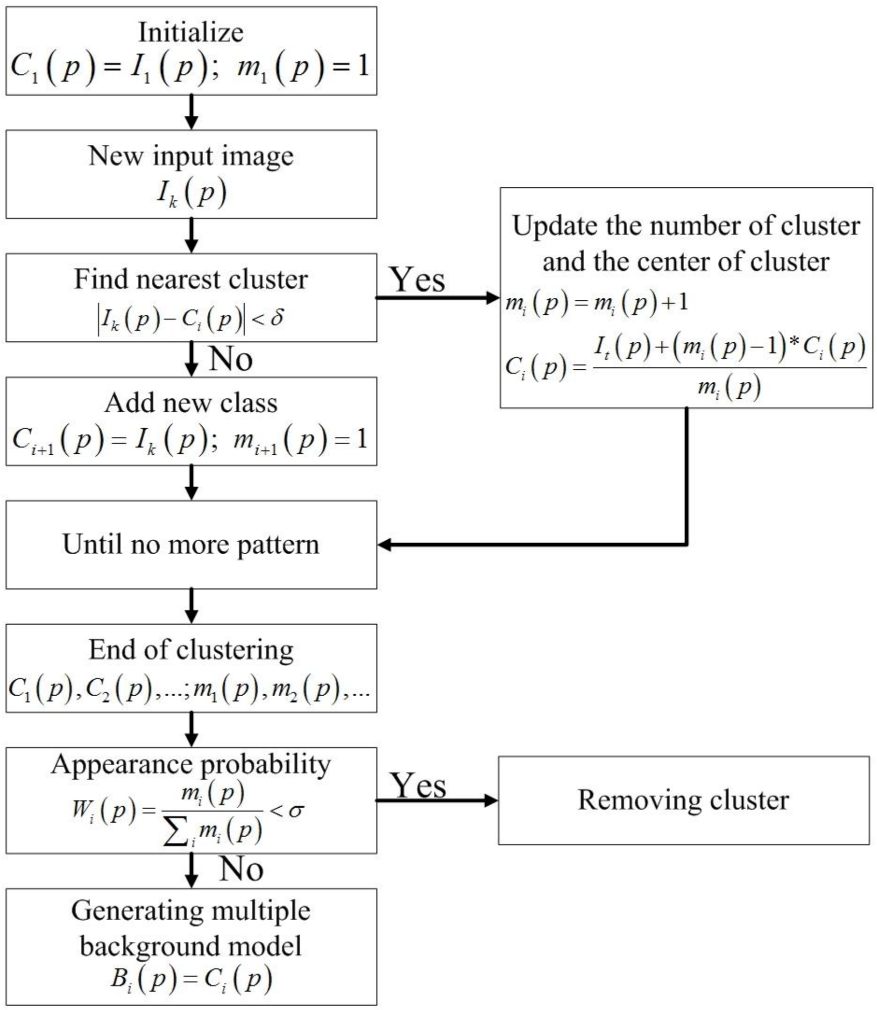


Figure 2: The block diagram of generating the multiple background model

We summarize the notations in Table 1 for explain the process of generating a background model in Fig. 2. The process need to initialize variables which are mean value of i^{th} cluster C_i and the number of i^{th} cluster m_i. We give the intensity value in 1^{st} frame as initial mean value.

Table 1: The notations for a process of generating the multiple background model

I_k	Image sequence, $k = 1, 2, 3, ..., N$
p	Index of pixel
C_i	Mean value of i^{th} cluster
m_i	The number of i^{th} cluster
δ	Threshold value for separation of each pixel among image sequence
W_i	Weight value of i^{th} cluster
σ	Threshold value for removing clusters with small weight
$N(p)$	Total number of cluster after online clustering
$n(p)$	The number of generated multiple background model
B_i	i^{th} background model
k	Index of frame
i	Index of cluster number

The first step is an online clustering based on pixel intensity. We classify a pixel value by the online clustering. All pixels in the selected N frames among whole image sequence are described in Eq. (1) and belong to a cluster which has its mean value and the used number of pixels.

$$I_k\left(p\right), k = 1, 2, 3, ..., N \tag{1}$$

The intensity value in each pixel from an input image is compared with mean value of each cluster by threshold value δ. The C_i and m_i is updated only when the difference is smaller than δ, otherwise we add new cluster. This process is iterated until no more pattern. Then, we obtain $N(p)$ clusters and their components: mean value and the referred number of pixels.

The second step is a removing cluster with a smaller weight value. Even though multi-background clusters are successfully estimated, a few clusters must be included the moving object if moving objects are existed in a time sequence. Therefore we calculate weight value of each cluster as the cue to solve expectative error which is occurred by generated clusters from a moving object. Those clusters should to have the weight value is smaller than regarded other clusters as backgrounds.

$$W_i = \frac{m_i(p)}{\sum_{i=1}^{N(p)} m_i(p)}, \qquad i = 1, 2, 3, \ldots, N(p) \tag{2}$$

The final step is a process of generating the multiple background model. We find weight value for each cluster to reduce error clusters. The clusters with small weights among $N(p)$ clusters was eliminated by comparing between W_i and σ, i.e., we obtain $n(p)$ clusters from $N(p)$ which has higher weight value than threshold σ. Using obtained cluster $n(p)$, we decide the multiple background model with their weight value and numbers. Those information will be use continuously to update background model in each frame.

$$\begin{aligned} &B_i(p) = C_i(p) \\ &i = 1, 2, 3, \ldots, n(p) \quad (n(p) \leq N(p)) \end{aligned} \tag{3}$$

$$W_i = \frac{m_i(p)}{\sum_{i=1}^{n(p)} m_i(p)}, \qquad i = 1, 2, 3, \ldots, n(p) \tag{4}$$

A background is normally and temporally changed. Therefore background updating is a important problem for not only the detection of moving object but also the understanding of environment in time sequence. One of simple classification of a background is long-term and short-term background discussed by [Elias et al. 2003]. In their updating strategy, they use temporal median filter to generate initial background and update background using stable information in long time (long-term) and temporal changes (short-term). However we use multiple background model for each pixel, we already have the distribution of stable information and temporal changes for each pixel. Therefore we update background based on clusters of each pixel with the strategy described in Elías et al.

3 Motion Segmentation

We apply a background subtraction to detect moving objects in multiple cameras. We use the multiple background model for a background subtraction. Fig. 3 shows a process of motion segmentation. We can see the background model MBM$_i$ in Fig. 3. Also, we apply the median background MB_k to reduce some noises. I_k is current image. We obtain two images. Once to get the difference between the background model and the current image D_b, and another to get the difference between the current image and the median background D_k.

$$D_b(x,y) = |I_k(x,y) - \mathrm{MBM}_i(x,y)| \qquad i = 1,2,3\ldots$$
$$D_k(x,y) = |I_k(x,y) - MB(x,y)| \tag{5}$$

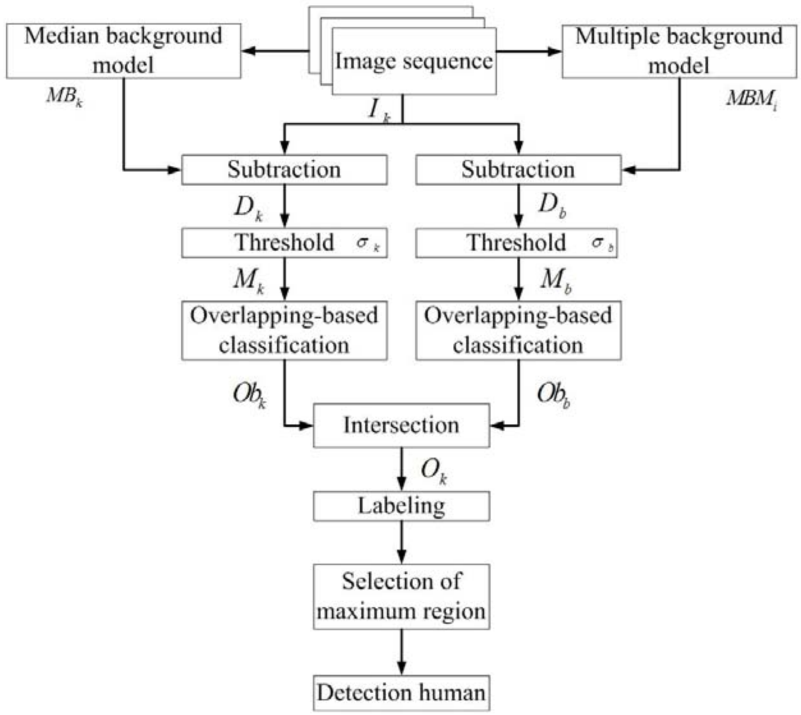


Figure 3: The block diagram of the motion segmentation process

These difference images show the probability of each pixel to be considered as moving object. We have to define a couple of threshold σ_b and σ_k to keep higher values and to refuse smaller ones. The result images will have 1's in pixels with motion, and 0's in pixels with background. Both thresholds can be selected manually.

$$M_b(x,y) = \begin{cases} 1 & D_b(x,y) > \sigma_b \\ 0 & otherwise \end{cases}$$
$$M_k\,(x,y) = \begin{cases} 1 & D_k(x,y) > \sigma_k \\ 0 & otherwise \end{cases} \tag{6}$$

M_b and M_k are comprised of moving and static regions. The static regions correspond to pixels with the same gray level between consecutive frames but different in relation to the background image. The Overlapping-based classification procedure distinguishes between both types. [Wang et al. 2004] So, in uses the M_b and M_k images and return two images. This process is shown in (9) and (10). Ob_b and Ob_k are the images comprised of regions belonging to moving objects. ε_1 and ε_2 are threshold value to separate moving and static regions. Using this threshold value, we obtain two images Ob_b and Ob_k.

$$Ob_b(x,y) = \begin{cases} 1 & if \quad M_b(x,y) \cup M_k(x,y) = 1 \quad then \quad D_b(x,y) > \varepsilon_1 \\ 0 & otherwise \end{cases}$$
$$Ob_k\,(x,y) = \begin{cases} 1 & if \quad M_b(x,y) \cup M_k(x,y) = 1 \quad then \quad D_k(x,y) > \varepsilon_2 \\ 0 & otherwise \end{cases} \tag{7}$$

We select only the regions of Ob_b that intersect with a region of Ob_k to detect moving objects. O_k is comprised moving regions. Finally, We label an intersection image O_k. We apply labeled image to select the maximum region which is a moving object.

$$O_k(x,y) = \begin{cases} 1 & Ob_b(x,y) \cap Ob_k(x,y) \\ 0 & otherwise \end{cases} \tag{8}$$

4 Correspondence among Multiple Views

4.1 Principal Axis

We apply the principal axis of detected region and the homography to correspond objects among multiple views. We use the principal axis moments to obtain a principal axis of detected region. For a gray level image with pixel intensities $I(x,y)$, *pq*-th order moments of image are defined as

$$M_{pq} = \sum_x \sum_y x^p y^q I(x,y) \tag{9}$$

For getting the principal axis of a detected region, we compute the central moments of detected region. The central moments μ_{pq} are defined as

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q I(x,y) \tag{10}$$

where $\bar{x} = \frac{M_{10}}{M_{00}}$ and $\bar{y} = \frac{M_{01}}{M_{00}}$ are the components of the centroid. The central moments of order up to 3 are defined as

$$\begin{aligned}
&\mu_{00} = M_{00},\\
&\mu_{01} = 0,\\
&\mu_{10} = 0,\\
&\mu_{11} = M_{11} - \bar{x}M_{01} = M_{11} - \bar{y}M_{10},\\
&\mu_{20} = M_{20} - \bar{x}M_{10},\\
&\mu_{02} = M_{02} - \bar{x}M_{01},\\
&\mu_{21} = M_{21} - 2\bar{x}M_{11} - \bar{y}M_{20} + 2\bar{x}^2 M_{01},\\
&\mu_{12} = M_{12} - 2\bar{y}M_{11} - \bar{x}M_{02} + 2\bar{y}^2 M_{10},\\
&\mu_{30} = M_{30} - 3\bar{x}M_{20} + 2\bar{x}^2 M_{10},\\
&\mu_{03} = M_{03} - 3\bar{y}M_{02} + 2\bar{y}^2 M_{01}.
\end{aligned} \tag{11}$$

We obtain the image orientation to use the second order central moments.

$$\begin{aligned}
&\mu 20'_{20} = \mu_{20}/\mu_{00} = M_{20}/M_{00} - \bar{x}^2\\
&\mu 20'_{20} = \mu_{02}/\mu_{00} = M_{02}/M_{00} - \bar{y}^2\\
&\mu 20'_{20} = \mu_{11}/\mu_{00} = M_{11}/M_{00} - \bar{x}\bar{y}
\end{aligned} \tag{12}$$

The covariance matrix of the image $I(x, y)$ is defined as

$$cov[I(x,y)] = \begin{bmatrix} \mu'_{20} & \mu'_{11} \\ \mu'_{11} & \mu'_{02} \end{bmatrix}. \tag{13}$$

The eigenvectors correspond to the major and minor axes of the image intensity, so the orientation can thus be extracted from the angle of the eigenvector associated with the largest eigenvalue.

$$\theta = \frac{1}{2} arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right) \tag{14}$$

where θ is an angle between basis coordinate and principal axis of a detected region. Fig. 4 shows example of the principal axis. The red line is the principal axis of detected region. The green point is the centroid of a detected region.

Figure 4: The principal axis of a detected region

4.2 Homography

We apply the homography to correspond objects between multiple views. The homography is projective transformations that map points from one plane to another plane. We consider the problem of determining the homography that map points on the ground plane as seen from one camera, to the same point on the ground plane as seen from another camera. The homography is defined as 3 by 3 homogeneous matrix **H**.

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \tag{15}$$

A point (x_i, y_j) in one image and (x'_i, y'_j) in another image are a pair of correspondence points on the ground plane.

$$\begin{bmatrix} x'_i \\ y'_j \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{16}$$

The homography matrix is computed using correspondence between several landmarks on the ground plane.

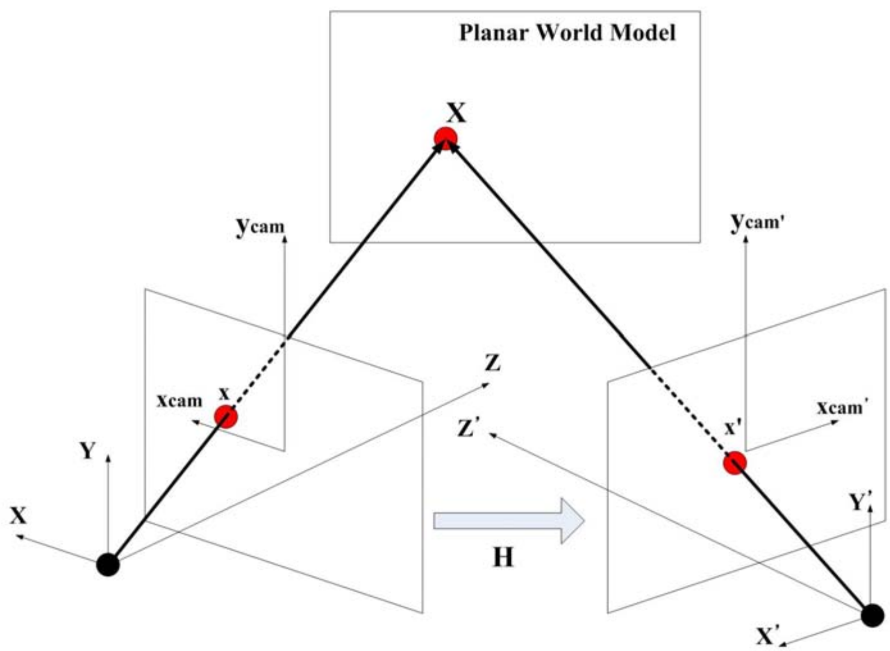


Figure 5: The concept of homography

4.3 Correspondence among Multiple Views

In the section, we present the correspondence objects in structured space by the multiple cameras. We use the principal axis of a detected object and the homography constraint to match objects under multiple views. Fig. 6 describes the geometrical relationship across multiple cameras. This relationship is used to overcome the problem correspondence under multiple cameras system [Hu et al. 2006].

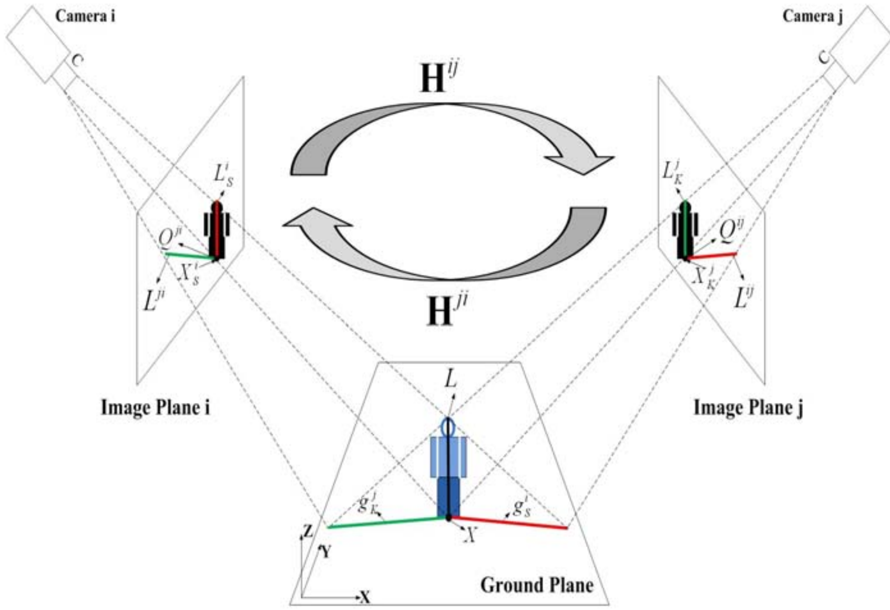


Figure 6: The geometrical relationship among multiple views

A line L_S^i is the principal axis of a person $\mathbf{S}$ in the camera i. $\mathbf{H}^{ij}$ is the homography from image plane i to image plane j. Also, $\mathbf{H}^{ji}$ is the homography from image plane j to image plane i. L^{ji} is the line in image plane i to by transforming L_K^j from image plane j to image plane i using homography $\mathbf{H}^{ji}$. L_K^j is the principal axis of person $\mathbf{K}$ in camera j. Q^{ji} is the intersection of L_S^i and L^{ji} in image plane i. This point is the intersection of principal axis L and ground plane to the same point X in 3D space. L is the principal axis of a person in 3D space. g_S^i is the line acquired by projecting L onto the ground plane in 3D space from the direction of camera i. Obviously, L^{ij} is the projecting of g_S^i on image plane j. It is obvious that if person $\mathbf{S}$ in camera i and person $\mathbf{K}$ in camera j correspond to the same person in 3D space, X corresponds to the two points Q^{ji} in image plane i and Q^{ij} in image plane j.

A point Q^{ji} in image plane i and Q^{ij} in image plane j are corresponding point. X_S^i is a point in image plane i by transforming Q^{ij} from image plane j to image plane i using homography $\mathbf{H}^{ji}$. X_K^j is similarly defined.

$$\begin{aligned} X_S^i &= \mathbf{H}^{ji} Q^{ij} \\ X_K^j &= \mathbf{H}^{ij} Q^{ji} \end{aligned} \tag{17}$$

We calculate the geometrical error between the transformed X_S^i and the intersection Q^{ji}. The point pairs of X_K^j and Q^{ij} are similarly defined. A point $Q^{ji}(x_1, y_1)$, $X_S^i(x_1', y_1')$, $Q^{ij}(x_2, y_2)$, and $X_K^j(x_2', y_2')$ are pairs of corresponding points. The geometrical error is given as

$$\begin{aligned} e_1(Q^{ji}, X_S^i) &= \sqrt{(x_1 - x_1')^2 + (y_1 - y_1')^2} < \alpha_1 \\ e_2(Q^{ij}, X_K^j) &= \sqrt{(x_2 - x_2')^2 + (y_2 - y_2')^2} < \alpha_2. \end{aligned} \quad (18)$$

If $e_1(Q^{ji}, X_S^i)$ and $e_2(Q^{ij}, X_K^j)$ are smaller than threshold valueα_1 and α_2, the person **S** in image plane i corresponds as the person **K** in image plane j. Thus, we track a person by these corresponding points on the ground plane. We use the correspondence information to track a person in each single camera view when tracked human is within the ground plane of any two views. As the principal axis of a person in each view can be detected robustly and accurately, the intersection of the principal axis of the person in one view and the line obtained by transforming the principal axis of the person from another view to the first view is robust and accurate. We use this intersection to track person in single view. As shown in Fig. 6, L_S^i and L_K^j correspond to the same person, the intersection Q^{ij} is used to track a person in camera i.

5 Experiment

We demonstrate that human tracking under multiple cameras system. Fig. 7 describes the configuration of structured space by multiple cameras. We use the five cameras that labeled C1–C5. We use 5000 images (5×1000).

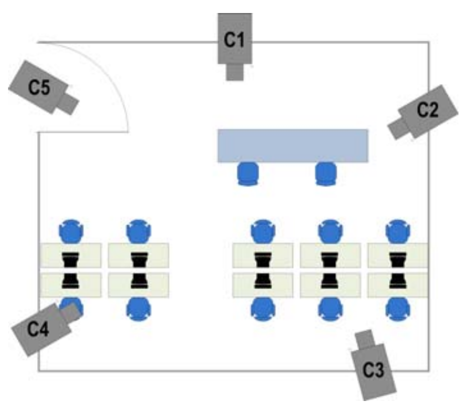


Figure 7: The configuration of multiple cameras

We can see that the result of multiple background model in the Fig. 8. From the top to the bottom, the camera numbers are, respectively, C1–C5. Each in row is generated by clusters for each pixel. Therefore the first column means the highest density clusters and other columns are orderly described for second, third, and fourth clusters. We apply 100 frames to generate the initial multiple background model.

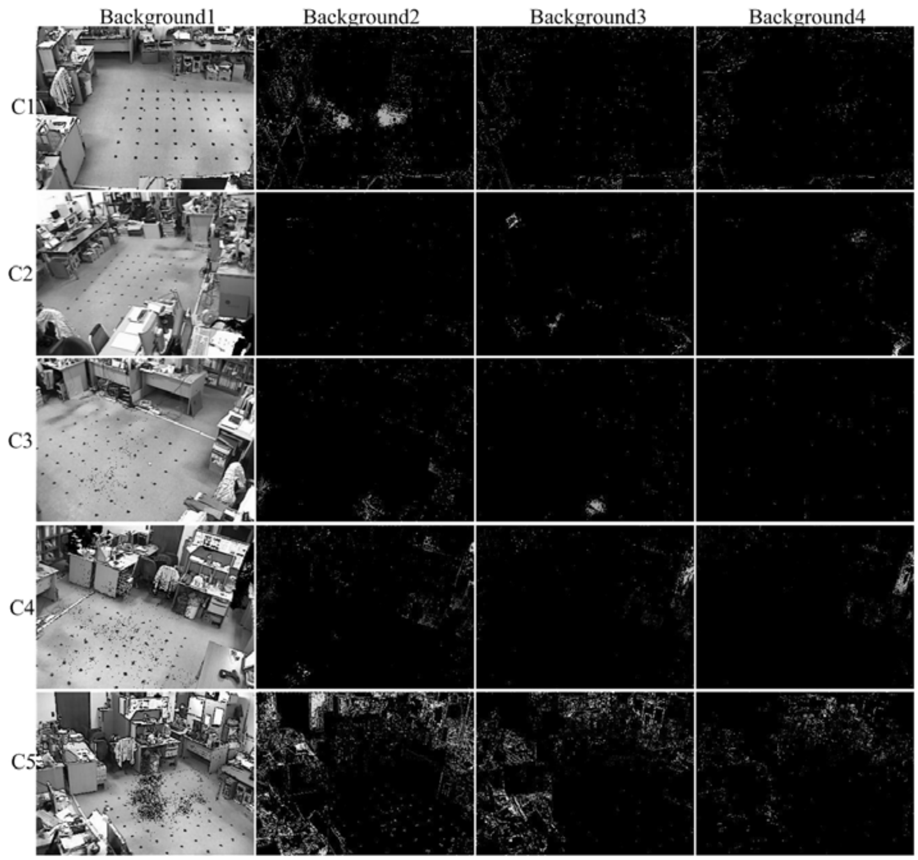


Figure 8: The multiple background model

To generate the median background model, we use the 31 frames. We apply the background models to detect moving objects. Fig. 9 shows the result of the median background. Using this background model, we reduce some noises under a process of motion segmentation.

Figure 9: The median background model

We apply the background models to the background subtraction. Using the multiple background model, we detect a moving object. And then we use the median background model to reduce noises. After detecting object, we compute the principal axis of a detected object by the principal axis moments. Fig. 10 shows the result of principal axis of a detected object. The green bounding box in each image is the detected object. Also, the yellow line in each image is the principal axis of a detected object by the principal axis moments. The red points are the centroid of detected objects.

After detection of the principal axis of a detected object, we can obtain to match object between two views using the homography constrain. Fig. 11 shows the result of the homography between two views. The bounding boxes of green and yellow are the detected moving object. The lines of green and yellow are the principal axis and the transformed principal axis by the homography constraint. The red points are the centroid and intersection with the principal axis and the transformed principal axis.

After getting the principal axis of a detected object, we find correspondence object among multiple views using the homography constraint. Fig. 12 shows the result of correspondence between two views. The first column is the correspondence between C1 and C2. The second column is the correspondence between C3 and C4. And the last column is the correspondence between C5 and C1. The bounding boxes of green and yellow color are detected a person for each camera. The lines of green and yellow are the detected principal axis and transformed principal axis using homography. The red points are the centroid point and the intersection point which principal axis and transformed principal axis.

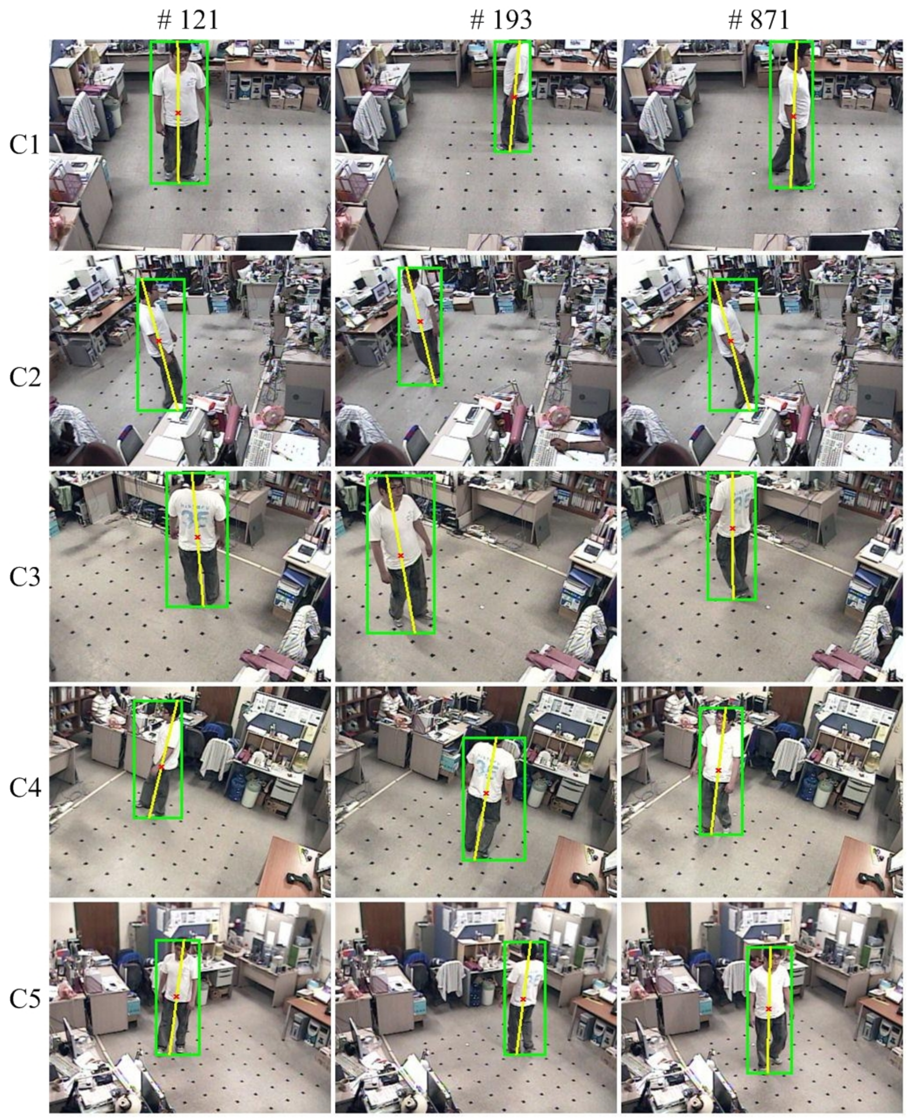


Figure 10: The result of principal axis of a detected object

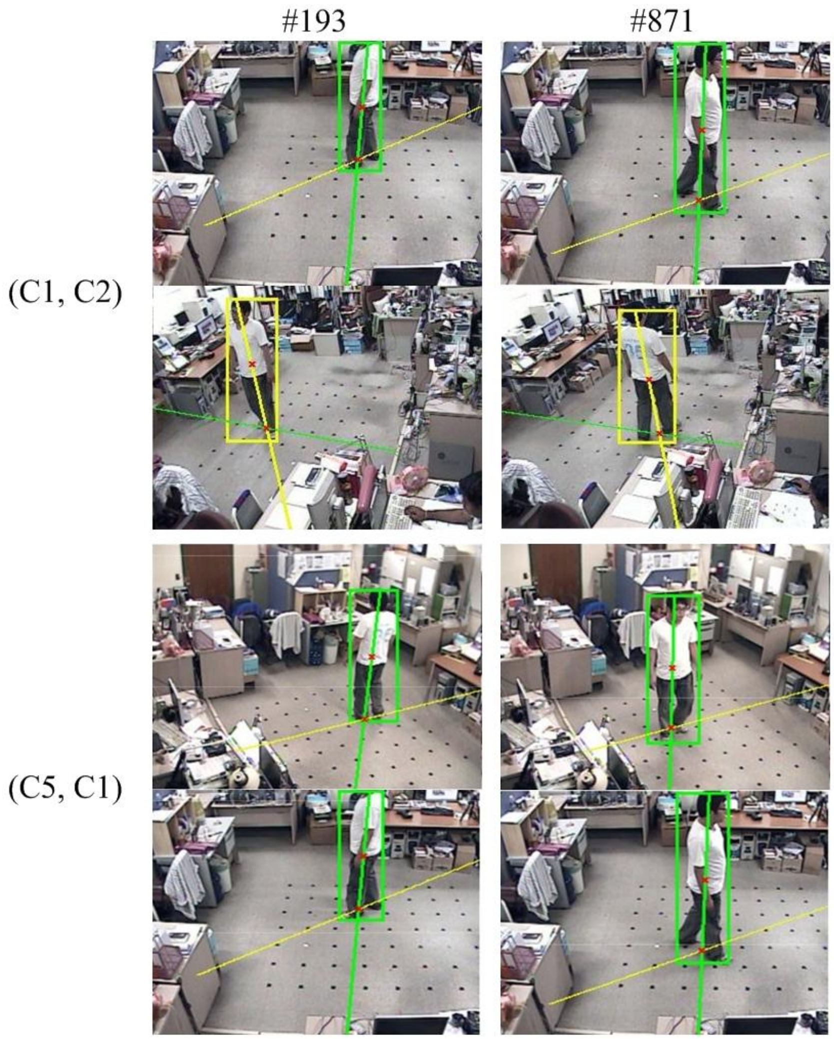


Figure 11: The result of homography between two views

Table 2 shows the rate of human region detection. We use 901 images in each camera. We can see the detected result over 99 %. If non-human region is detected larger than human region, human region process is failed. Also, the result of incorrectly detected region is 1.05 %. We can get the accurate result. But we consider only one person to stay in structured space by multiple cameras. Table 3 shows the rate of correspondence between two views. In the structured space, total object means stayed moving objects at the same time. We can see the corresponding rated over 97 %. From this table, we know that the result obtained accuracy.

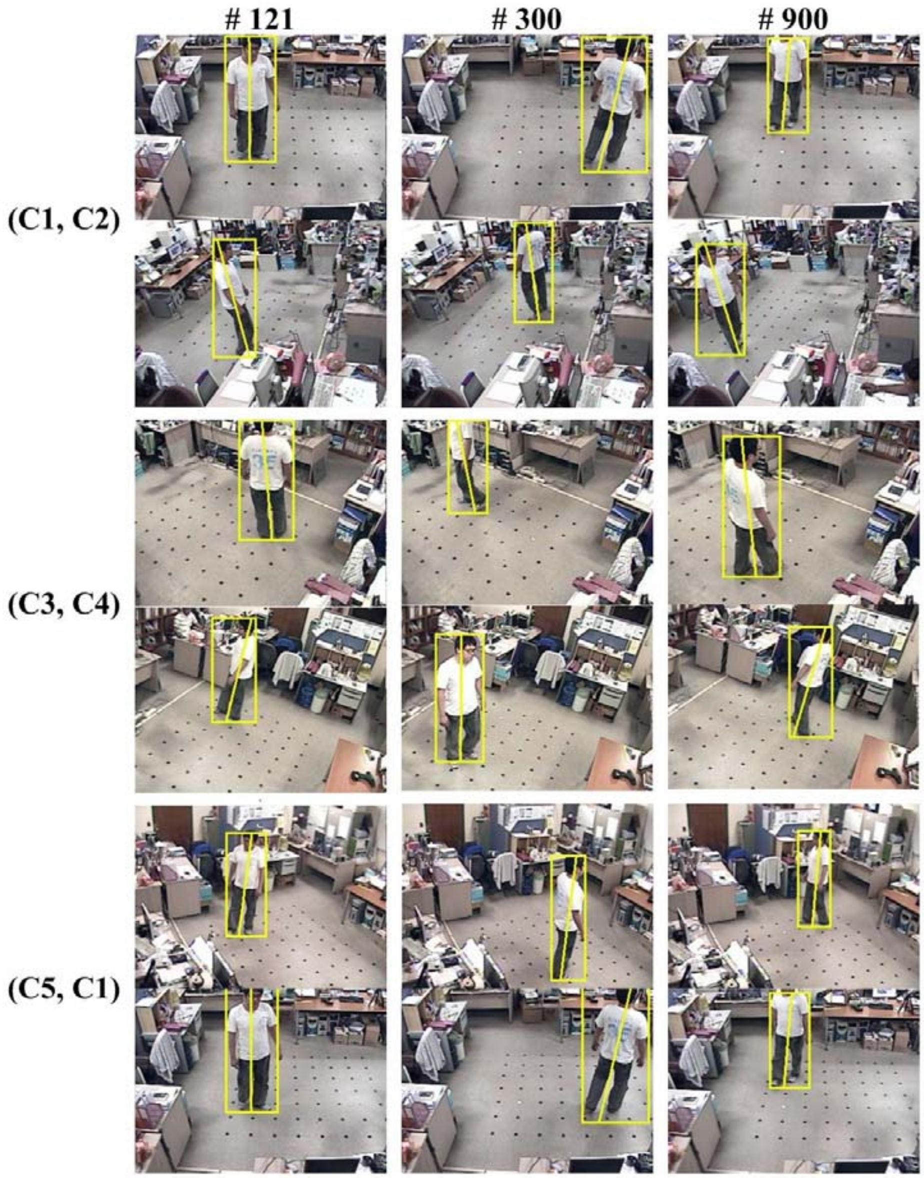


Figure 12: Result of correspondence between two views

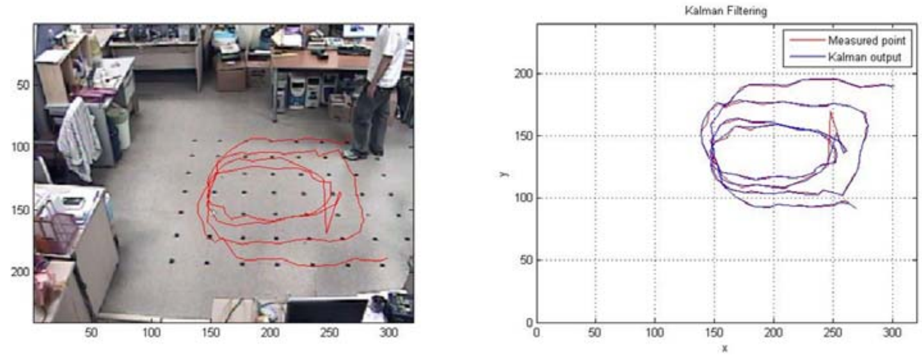
Table 2: The rate of human region detection

Camera Number	Total Image	Detected Region	Incorrectly Detected Region
C1	901	893/901 99.11 %	13/893 1.46 %
C2	901	885/901 98.22 %	13/885 1.47 %
C3	901	901/901 100.00 %	12/901 1.33 %
C4	901	901/901 100.00 %	3/901 0.33 %
C5	901	901/901 100.00 %	6/901 0.67 %
Total	4505	4481/4505 99.47 %	47/4481 1.05 %

Table 3: The rate of correspondence between two views

Stereo Image	Total Object	Corresponding Rate
(C1, C2)	887	867/887 97.75 %
(C3, C4)	901	875/901 97.11 %
(C5, C1)	893	875/893 97.98 %
Total	2681	2617/2681 97.61 %

Fig. 13 shows the result of Kalman filter. [Welch and Bishop 2001] We track a person by corresponding points on the plane. (a) is the trajectory in camera 1. (b) is the result of Kalman filtering. The red line is the measuring points. The blue line is the kalman output.



(a) Trajectory in C1

(b) Kalman filtering

Figure 13: The result of human tracking in 2D space

6 Conclusion

Tracking in multiple cameras system is necessary to match object among multiple views. We described how to correspond objects among multiple images which are taken by multiple cameras. The correspondence among multiple cameras system is finding a similarity among simultaneously existed objects in different views. We apply principal axis and homography to find correspondence objects among multi-views. To get a principal axis of object, we need to detect objects in the single view. For getting the silhouette of objects, we apply a background subtraction using multiple background model. The several variations for each pixel in time sequence are successfully contained by multiple background model. After motion segmentation, we use the principal axis moments to calculate a direction of principal axis for detected silhouette. For getting the homography, we used some landmarks on the ground plane. Using principal axis and homography, we found correspondence objects among multiple views. In the 3D space, we have the only one intersection between the principal axis of object and the ground plane. The intersection is an invariant point in different views. Therefore we estimate the invariant point of each view by principal axis of objects in one view and transformed principal axis in another view to the same view using the homography constraint. We show the detected intersections in the experimentation result. The correspondence is constructed according to relationship between intersection points. The experimental result shows the accurate correspondence among multiple views. In our future work, we will use these result to apply the recognition of human gait. So, we will consider accurate detection of moving person for gait recognition.

Acknowledgments

The authors would like to thank to Ulsan Metropolitan City, MOCIE, KOTEF, and MOE of Korean Government which partly supported this research through the NARC, the Human Resource Training project for regional innovation, and post BK21 project at University of Ulsan.

References

[Chang and Gong 2001] Chang, T. H., Gong, S.: "Tracking Multiple People with a Multi-Camera System"; Proc. of 2001 IEEE Workshop on Multi-Object Tracking, 2001, 19-26.

[Elgammal et al. 2002] Elgammal, A., Duraiswami, R., Harwwod, D., and Davis, L.: "Background and Foreground Modeling using Nonparametric Kernel Density Estimation"; Proc. of the IEEE. 90(7), 2002, 1151-1163.

[Elias et al. 2003] Elías, H. J., Carlos, O. U., and Jesús, O.: "Detected Motion Classification with a Double-background and a Neighborhood-based Difference"; Pattern Recognition Letters, 24(12), 2003, pp. 2079–2092

[Hu 1962] Hu, M. K.: "Visual Pattern Recognition by Moments"; IRE Trans. Info. Theory, vol. IT-8, 1962, 179-187.

[Hu et al. 2006] Hu, W., Hu, M., Zhou, X., Tan, T., Lou, J., and Maybank, S.: "Principal Axis-Based Correspondence between Multiple Cameras for People Tracking"; IEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 4, 2006, 663-671.

[Khan et al. 2001] Khan, S., Javed, O., Rasheed, Z., and Shah, M.: "Human Tracking in Multiple Cameras"; Proc. of IEEE Int'l Conf. on Computer Vision, 2001, 331-336.

[Kim and Jo 2007] Kim, Taeho and Jo, Kang-Hyun: "Detection of Moving Object: Background and Foreground under Moving Camera"; Int'l J. of Information Acquisition, vol. 4, no. 3, 2007, 227-236.

[Krumm et al. 2000] Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., and Shafer, S.: "Multi-Camera Multi-Person Tracking for EasyLiving"; Proc. IEEE Int'l Workshop Visual Surveillance, 2000, 3-10.

[Mittal and Davis 2002] Mittal, A., Davis, L. S.: "M2Tracker: A Multi-View Approach to Segmenting and Tracking People in a Clutted Scene using Region-Based Stereo"; Proc. Euro. Conf. Computer Visoin, 2002, 18-36.

[Morioka et al. 2006] Morioka, K., Mao, X., and Hashimoto, H.: "Global Color Model Based Object Matching in the Multi-Camera Environment"; Proc. of the 2006 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems, 2006, 2644-2649.

[Orwell et al. 1999] J., Orwell, P., Remagnino, and Jones, G. A.: "Multiple Camera Color Tracking"; Proc. IEEE Int'l Workshop Visual Surveillance, 1999, 14-24.

[Utsumi et al. 1998] Utsumi, A., Mori, H., Ohya, J., and Yachida, M.: "Multiple Human Tracking using Multiple Cameras"; Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition, 1998, 498-503.

[Wang et al. 2004] Wang, C. K., Cheng, M. Y., Liao, C. H., Li, C. C., Sun, C. Y., and Tsai, M. C.: "Design and Implementation of a Multi-purpose Real-time Pan-tilt Visual Tracking System"; Proc. of the 2004 IEEE Int'l Conf. on Control Applications, 2004, 1079-1084.

[Welch and Bishop 2001] Welch, G. and Bishop, G.: "An Introduction to the Kalman Filter"; SIGGRAPH 2001.