International Journal of Innovative Computing, Information and Control
Volume **x**, Number **0x**, x **2008**
 ISSN 1349-4198
pp. **0–0**

REAL TIME TRACKING AND IDENTIFICATION OF MOVING PERSONS BY USING A CAMERA IN OUTDOOR ENVIRONMENT

Budi Sugandi, Hyoungseop Kim, Joo Kooi Tan and Seiji Ishikawa
Graduate School of Engineering
Kyushu Institute of Technology
1-1 Sensui, Tobata, Kitakyushu 804-8550, Japan
kim@cntl.kyutech.ac.jp

Abstract. *A new method for detecting and tracking of moving persons based on low resolution image employing peripheral increment sign correlation image and identifying the moving persons by their color and spatial information is proposed in this paper. Many tracking algorithms have better performance under a static background in indoor environment. It is, however, most of the tracking algorithms are applied in outdoor environment with noisy background instead of indoor environment. Since a low resolution image has a property that it can remove the small size pixels, it is adopted to solve the problem of the noisy background. In the tracking of a target object, many applications have problem when object occlude each other. A block matching technique based on peripheral increment sign correlation image is utilized to solve this problem. The identification of a target object is performed using color and spatial information of the target object. The experimental results prove the feasibility and usefulness of the proposed method*
Keywords: Person Tracking, Person Identification, Block Matching Technique, Peripheral Increment Sign Correlation Image

1. **Introduction.** In recent years, with the latest technological advancements, visual surveillance and security system receive a great interest among many researchers. Until recently, video surveillance and security system was mainly a concern for military or large-scale companies. However, due to increasing crime rate, necessitates are taking better precautions in security-sensitive areas, like country borders, airports or government offices. Even individuals are seeking for personalized security systems to monitor their houses or other valuable assets. The vast amount of data acquired from video imagery should be analyzed by an intelligent and useful autonomous structure. This intelligent system should have the capacity to observe the surrounding environment and extract useful information for subsequent reasoning, like detecting and analyzing the activity, or identifying the objects entering the scene. Besides, monitoring should be done 24 hours a day, without any interruption. This sort of a system will achieve the security system task more accurately and effectively and saving a great amount of human effort.

Recently, various image processing methods for detecting and tracking of moving persons have been proposed in the past, such as, background subtraction, frame difference, optical flow and probability based approach. Liu *et al.* [1] proposed a background subtraction technique to detect the moving objects in an image by taking the difference between current and reference background image. It is extremely sensitive to the change in dynamic scenes derived from lighting and extraneous events etc. Lipton *et al.* [2] proposed a frame difference method that use of the differences between two or three successive frame images to extract the moving regions. This method is very adaptive to dynamic environment, but generally does a poor job of extracting all the relevant pixels, e.g., there may be holes left inside moving entities. Meyer *et al.* [3] proposed an optical flow method by computing the displacement vector field to initialize a contour based tracking algorithm, called active rays, for the extraction of articulated objects. The optical flow method can be used to detect independently moving objects even in the presence of camera motion. However, most flow computation methods are computationally complex and very sensitive to noise. The method proposed by Paragios *et al.* [4] considered a probability and statistics problem and used the observed information to obtain a classification equation of probability to segment image. These detection models can detect and track moving objects perfectly, but they suffer from the high computing cost. So it is hard to be used in real-time surveillance application. Davis *et al.* [5] proposed the W4 method that designed a simplified statistics model to make real-time processing possible. First, the proposed method learned the background by the luminance of image and used a statistical cycle to observe the variance range of every pixel in the frames that have no moving objects. After the process, the method can build a background model by the observed information and compute the inter-frame difference of the current frame and background model to detect the moving pixels. The result of W4 is good performance and does not need much computing cost, but the background model has to be often re-established.

To overcome those problems, we propose a new method for detecting and tracking of moving persons based on low resolution image [6] employing peripheral increment sign correlation (PISC) [7] image. A low resolution image has a nice property that it can remove the small size region which contained image noise having pepper and salt noise. The PISC image is an image matching method with robust performance, high accuracy and high computational efficiency. The PISC image is constructed from the trend of the brightness changes in the neighborhood of the pixel under consideration. The PISC image is used in block matching process to determine the target object among the moving objects emerging in the background. We also proposed an algorithm of feature extraction and person identification using spatial and color information of the tracked persons [8].

This paper is organized as follows. The detail of our proposed methods and some image processing techniques for detecting and tracking of the moving persons are presented in section 2. Experimental results employing our proposed method are illustrated in section 3. And finally, some discussions and conclusions are given in section 4.

2. **Methods.** In this paper, we categorize our study in three stages; object detection, object tracking and object identification. On each stage, we apply our new method and evaluate the effectiveness of our method.

The first stage is begun by capturing the image with a camera. In this stage, we propose a low resolution image to remove the image noise having pepper and salt noise and any small motion in the background. Then, some image pre-processing techniques are performed to detect the moving objects. The second stage is performed to track a target object among the moving objects emerging in the background. We perform a block matching algorithm to track the target object. The block is made based on PISC image that considers the brightness change in the pixels of the block relative to the considered pixel. As a last stage, we identify the target object by extracting the color and spatial information of the tracked object. In this paper, we use means and standard deviation value as image feature of the tracked object. To evaluate the identification process, we use similarity equation that is performed by comparing the image feature of the target and the reference object. The most similar feature is then compared with a certain threshold value. The details of each technique are described as follows.

2.1. **Object detection.** This stage is aim at detection the moving objects appear in the scene. In order to detect the moving objects, we perform some image processing

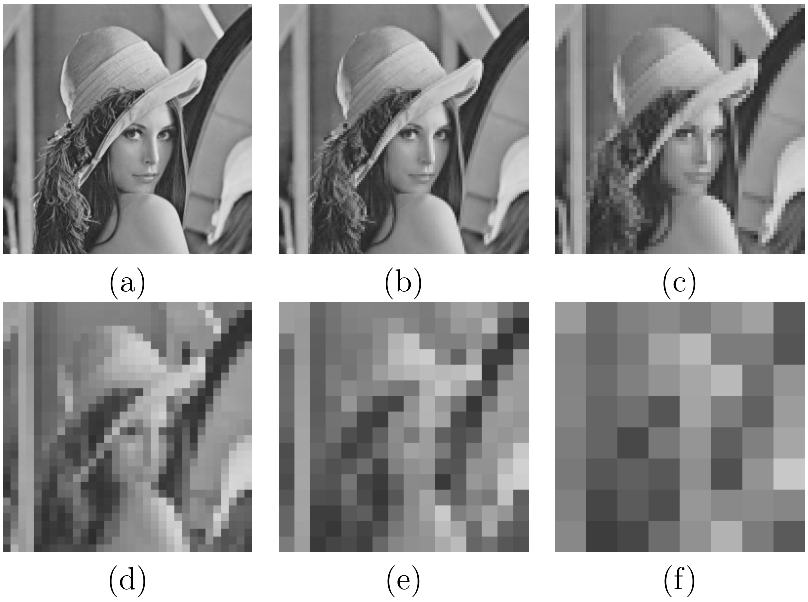


FIGURE 1. Low resolution image in various resolutions : (a) original, (b) 1/2 from original, (c) 1/4, (d) 1/8, (e) 1/16, (f) 1/32

techniques such as, low resolution image, frame difference, morphological operation and labeling technique.

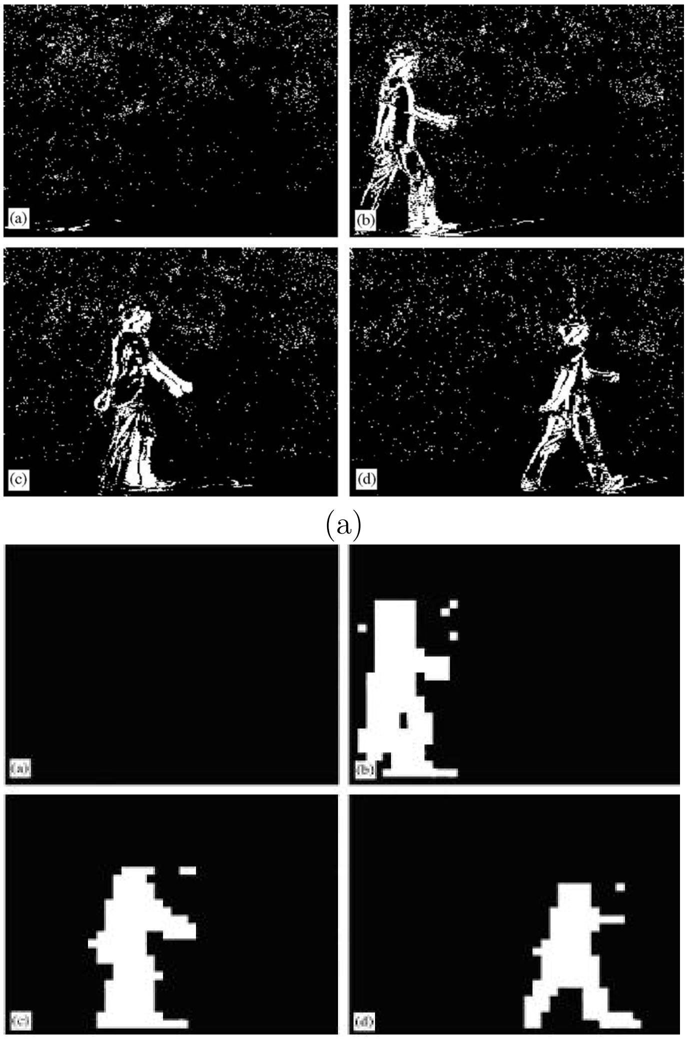
A low resolution image is made by reducing the spatial resolution of an image with keeping the image size [6, 9]. We make the low resolution image by averaging the pixels value of its neighbors, including itself. The low resolution image is used to remove the scattering noise having salt and pepper noise and any small motion appear in the background. FIGURE 1 shows the low resolution image in different resolutions.

Frame difference image [2, 11] is performed by taking the difference image between successive frame images. In this paper, frame difference method is performed on the three successive frames of the low resolution images. The comparison of moving object detection results using a conventional method and a low resolution image is shown in FIGURE 2. As shown in that figure, by using the conventional method (frame difference technique) the detected moving object is still greatly affected by the noise such as moving leaves. However, by using a low resolution image, the noise is removed completely.

Morphological operation is performed to fill small gaps inside the moving objects and to remove the noise remained in the moving objects [12]. We implement the morphological operations such as dilation and erosion. Dilation adds pixels to the boundary of the object and closes isolated background pixel while erosion removes isolated foreground pixels.

The next step is performing connected component labeling method. This method is performed to label each moving object emerging in the scene. The connected component labeling groups the pixels into components based on pixel connectivity (same intensity or gray level) [6]. We perform the connected component labeling method by comparing a pixel with the pixels in the four neighbors (the neighbors (i) to the left of the pixel, (ii) above it, and (iii and iv) the two upper diagonal terms) from top-left to bottom-right and from bottom-right to top-left. A pixel labels as neighbor's label when it has at least one neighbor with the same label.

2.2. **Object tracking.** In this stage, the target object will be tracked from the detected objects by performing a block matching method. The blocks are made based on the PISC image that considers the brightness changes on the image. The blocks are defined



(a) (b)

FIGURE 2. Comparison of moving object detection technique : (a) by using a conventional technique, (b) by using a low resolution image

by dividing the image frame into non-overlapping square parts. FIGURE 3 shows an example of a block in PISC image with block size of 9×9 pixels.

The blocks of the PISC image in the previous frame are defined in equation (1). Similarly, the blocks of the PISC image in the current frame are defined in equation (2). The matching blocks between two successive frames (the previous and current frame) is determined by using correlation value that express in equation (3). The high correlation value shows that the blocks are matched each other.

$$b_{np} = \begin{cases} 1 & \text{if } f_{np} \geq f(i,j) \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

$$b'_{np} = \begin{cases} 1 & \text{if } f_{np} \geq f(i,j) \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

$$corr_n = \Sigma_{p=0}^{k-1} b_{np} * b'_{np} + \Sigma_{p=0}^{k-1} (1 - b_{np}) * (1 - b'_{np}) \quad (3)$$

$f_{i-4,j-4}$	$f_{i-3,j-4}$	...	...	...	...	...	$f_{i+3,j-4}$	$f_{i+4,j-4}$
$f_{i-4,j-3}$	$f_{i-3,j-3}$	...	...	...	...	...	$f_{i+3,j-3}$	$f_{i+4,j-3}$
...	...	...	...	...	...	...	...	...
...	...	...	$f_{i-1,j-1}$	$f_{i,j-1}$	$f_{i+1,j-1}$	...	...	...
...	...	...	$f_{i-1,j}$	$f_{i,j}$	$f_{i+1,j}$	...	...	...
...	...	...	$f_{i-1,j+1}$	$f_{i,j+1}$	$f_{i+1,j+1}$	...	...	...
...	...	...	...	...	...	...	...	...
$f_{i-4,j+3}$	$f_{i-3,j+3}$	...	...	...	...	...	$f_{i+3,j+3}$	$f_{i+4,j+3}$
$f_{i-4,j+4}$	$f_{i-3,j+4}$	...	...	...	...	...	$f_{i+3,j+4}$	$f_{i+4,j+4}$

FIGURE 3. An example of PISC image with size of 9 × 9 pixels.

Here *b* and *b'* are the block in the previous and current frame, *n* is number of block, *p* is pixel and *k* is number of pixels of block, respectively.

The tracking process is illustrated in Figure 4. First off all, to reduce the processing time, blocks are made only in the moving object area as shown in Figure 4(a) and 4(b). An object is determined as a target object when it has matching blocks more than certain threshold. The tracking area can be enlarged, as shown in Figure 4 by dash line, when the target object cannot be determined in the previous criteria. As shown in Figure 4, block A of one object with size 9 × 9 pixels in previous frame (Figure 4(a)) will search the matching block in the blocks of others objects in current frame (Figure 4(b)) by using correlation value as express in equation (3). When the number of matching blocks of an object in the current frame is more than certain threshold, we determine this object as target object. Then, it will be tracked by our system.

2.3. **Object identification.** As a last stage, we identify the target object by extracting the spatial and color information of the tracked object. In this paper, we use means and standard deviation value as image feature of the target object. To evaluate the identification process, we use similarity equation that is performed by comparing the image feature of the target and the reference object.

After obtaining the target object in previous stage, we extract the spatial information of object such as position. For this purpose, we use bounding box of the object determined by computing the maximum and minimum value of x and y coordinates of the target object according to the following equations:

$$B^i_{min} = \left\{ \left(x^i_{min}, y^i_{min} \right) | x, y \in O^i \right\} \quad (4)$$

$$B^i_{max} = \left\{ \left(x^i_{max}, y^i_{max} \right) | x, y \in O^i \right\} \quad (5)$$

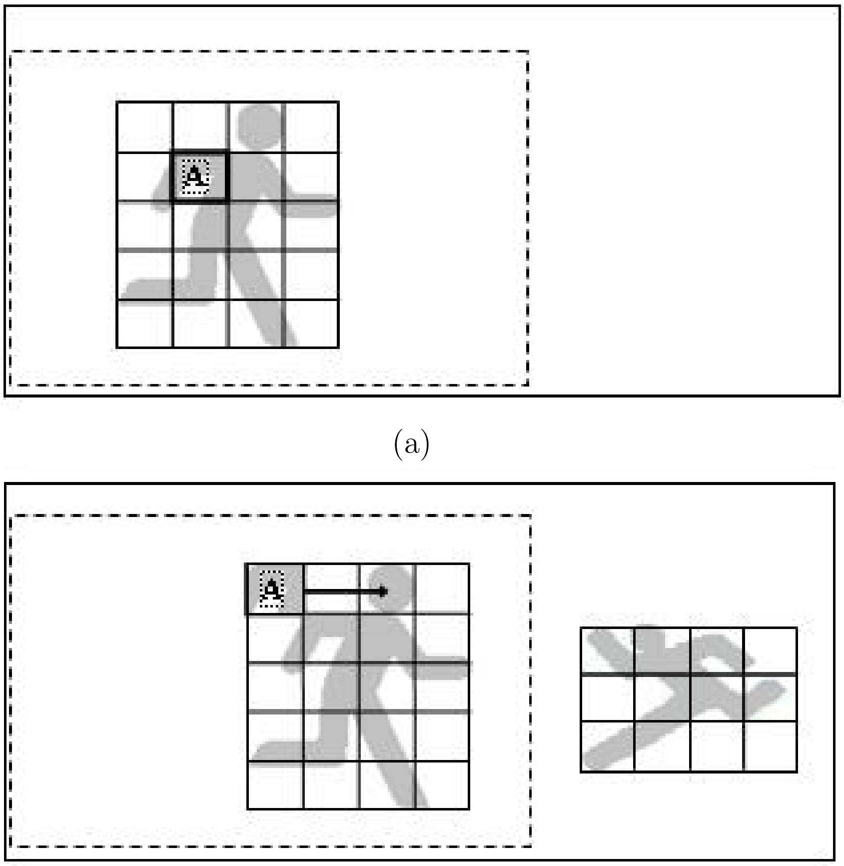


FIGURE 4. Tracking process : (a) previous frame, (b) current frame

Here, Oi denotes the set of the coordinate of points in the target object i, B^i_{min} is the left-top corner coordinates of the target object i and B^i_{max} is the right-bottom corner coordinates of the target object i, respectively.

We use RGB color as color information of moving objects. We can extract the color information from upper and lower part of the human body to obtain more color information. For that purpose, we can divide the human object into three parts; the head, the upper and the lower part of the body. The ratio of these three parts can be defined as shown in FIGURE 5. Because it is difficult to get valuable color information from human head, we only calculate the color information of upper and lower part of the human body.

The first color information calculated is mean value of each human body part as calculated by equation (6) and (7). The mean value is calculated for each color component of RGB space.

$$\mu_{f_k}^{O_U^i} = \frac{\Sigma_{x=x^i_{min}}^{x^i_{max}} \Sigma_{y=y^i_{min}}^{y^i_{max}} f_k(x,y)}{\#O_U^i} \tag{6}$$

$$\mu_{f_k}^{O_L^i} = \frac{\Sigma_{x=x^i_{min}}^{x^i_{max}} \Sigma_{y=y^i_{min}}^{y^i_{max}} f_k(x,y)}{\#O_L^i} \tag{7}$$

where i is number of the moving objects and (x, y) is the coordinate of pixels in moving object. (x^i_{min}, y^i_{min}) and (x^i_{max}, y^i_{max}) are the coordinates of the bounding box of moving object i, $f_k(x, y)$ denotes pixel value for each color component in RGB space of the current frame, O^i_U and O^i_L denote the set of coordinates of upper and lower part of human body of moving object i and $\#Oi$ is the number of pixels of moving object i, respectively.

Standard deviation is a statistical term that provides a good indication of volatility. It measures how widely the values are dispersed from the average. Dispersion is the difference between the actual value and the average value. The larger the difference between the actual color and the average color is, the higher the standard deviation will be and the higher the volatility. We can extract more useful color features by computing the dispersed color information from upper and lower part of body as shown in equation (8) and (9).

$$SD^{O^i_U}_{f_k} = \sqrt{\frac{\sum_{x=x^i_{min}}^{x^i_{max}} \sum_{y=y^i_{min}}^{y^i_{max}} \left(f_k(x,y) - \mu^{O^i_U}_{f_k}\right)^2}{\#O^i_U}} \tag{8}$$

$$SD^{O^i_L}_{f_k} = \sqrt{\frac{\sum_{x=x^i_{min}}^{x^i_{max}} \sum_{y=y^i_{min}}^{y^i_{max}} \left(f_k(x,y) - \mu^{O^i_L}_{f_k}\right)^2}{\#O^i_L}} \tag{9}$$

where $SD^{O^i_U}_{f_k}$ and $SD^{O^i_L}_{f_k}$ denote the standard deviation of each color component of RGB space for upper and lower part of the human body of moving object i, respectively.

After the feature of the object is extracted, we can represent the moving object i by the following feature vectors;

$$F^i = \left(\mu^{O^i_U}_{f_k}, \mu^{O^i_L}_{f_k}, SD^{O^i_U}_{f_k}, SD^{O^i_L}_{f_k}, B^i_{min}, B^i_{max}\right) \tag{10}$$

To identify a moving object, a feature queue is created to save the features of the moving objects. When a new object enters the system, it will be tracked and labeled and the features of the object are extracted and recorded into the queue. Once a moving object i is detected, the system will extract the features F^i of the object i and identify it from the identified objects in the queue by computing the similarity $S(F^i, F^j), j = 1...n$, where j is one of the n identified objects. The similarity, $S(F^i, F^j)$ is computed as following,

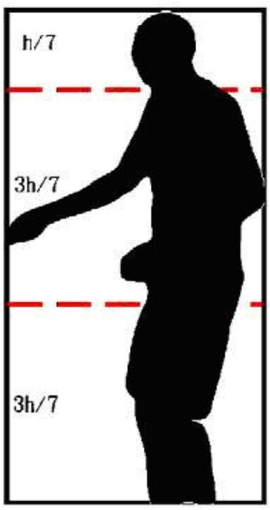


Figure 5. The definition of a human body ratio

$$S\left(F^i, F^j\right) = \begin{array}{l} Mc\left(\left|\mu_{f_k}^{O_U^i} - \mu_{f_k}^{O_U^j}\right|\right) + Mc\left(\left|\mu_{f_k}^{O_L^i} - \mu_{f_k}^{O_L^j}\right|\right) + \\ Msd\left(\left|SD_{f_k}^{O_U^i} - SD_{f_k}^{O_U^j}\right|\right) + Msd\left(\left|SD_{f_k}^{O_L^i} - SD_{f_k}^{O_L^j}\right|\right) + \\ 0.5Mp\left(\left|B_{min}^i - B_{min}^j\right|\right) + 0.5Mp\left(\left|B_{max}^i - B_{max}^j\right|\right) \end{array} \quad (11)$$

where Mc and Msd are the membership function of color information and standard deviation and Mp is the membership function of spatial information, respectively, as defined in Equation (12), (13) and (14).

$$Mc(x) = \begin{cases} 1 - x/Thr & \text{if } x < Thr \\ 0 & \text{if } x \geq Thr \end{cases} \quad (12)$$

$$Msd(x) = \begin{cases} 1 - x/Thr & \text{if } x < Thr \\ 0 & \text{if } x \geq Thr \end{cases} \quad (13)$$

$$Mp(x) = \begin{cases} 1 - 3x/W & \text{if } x < W/3 \\ 0 & \text{if } x \geq W/3 \end{cases} \quad (14)$$

where Thr is threshold which is obtained empirically and W is the width or height of the image frames. We compare the features of an object with those of the objects in the feature queue. The one with the maximum similarity is identified as same as target object.

3. **Experimental Results.** We have done the experiments by using a camera in outdoor environment and real time condition. The experiments are implemented on Pentium 4 with 2.53 GHz CPU and 256 MB RAM. Our method can detect and track the target object at 2 frames/second over 320 × 240 pixels image. In this paper, we use the block size of 9 × 9 pixels. We assume that the object coming firstly will be tracked as the target object. We tracked the target object from two and three moving objects that occluded between each other. The experimental results are shown in the selected frames in FIGURE 6 - FIGURE 8. The rectangle area around the moving person shows the extracted target person.

On the first experiment as shown in FIGURE 6, two objects are moving in the same direction. At first, a man wearing the blue shirt enters the scene from the left side. This person is assumed as a target object and will be tracked by our system. Then on the next frame, the second man wearing the white shirt enters the scene from left side also. They move in the same direction and overlap each other. As shown in the figure, we successfully track the target object in all sequence of the frames.

On the second experiment as shown in FIGURE 7, two moving persons are moving in the different direction. At first, a man wearing the white shirt enters the scene from the left side. This person is assumed as a target object. Then on the next frame, the second man wearing the blue shirt enters the scene from right side. They move in the different direction and overlap each other in the middle of the scene. We successfully track the first moving person as target person as shown in that figure.

On the third experiment as shown in FIGURE 8, we track the target object among three moving objects appear in the background. Here, we assume a man wearing a white shirt as a target object. We also successfully track the target in all sequence of frames.

The experimental results of the object identification by using color and spatial information of moving object are shown in TABLE 1. On this experiment, we evaluate the similarity of the object features between the detected object and the target object on each frame. The correct identification is determined when the detected object on each frame is similar to the target object. Then, the identification rate is determined from any correct identification and the detected object on each frame. From the table, we notice that some objects are not correctly identified in some frames of each case. The wrong identification occurs in two types. First, it occurs when the moving object just enters or leaves the scene. When moving object is just entering or leaving the scene, the system can detect the moving object. However, the moving object is detected and tracked at the border of the scene; the extracted features of the moving object in this case cannot represent the moving object very well. The main reason is that only partial of the features of the moving objects are extracted. The second kind of erroneous identification occurs when the moving object moves slowing down. In this situation, the frame difference image of the moving object will become smaller and we will get smaller bounding box and less moving pixels. Therefore, the extracted features will lose its representative.

4. **Discussion and Conclusion.** This paper proposed a new method for detecting and tracking of moving object based on low resolution image employing PISC image for real time application and an identification method by using color and spatial information of the moving object. The proposed method detects the objects appear in the background, tracks the target object and identifies the tracked object. By using our new method, the satisfactory results are achieved. The results improved the previous methods. The salt and pepper noise and the small motion in the background are completely removed, while the object regions are extracted. The proposed method can track the target object in occlusion condition by another object. Furthermore, the identification result of the tracked object shows the effectiveness of our method with the identification rates more than 90%.

However, the proposed methods still have limitations. The low resolution image cannot detect the small object appears in the background due to low resolution image removed the small size area of the object. These limitations can be solved by adding other information to the target object such as flow of moving object based on optical flow, dimension or another feature. The tracking method has difficulty when the objects appear in the same time. We cannot track the target object as our assumption. We can solve this problem by extracting the skin information or the optical flow of the target object. Moreover, the object identification method cannot identify the object when it just enters or leaves the scene and when it moves slowing down. This problem arises due to the system cannot extract the feature correctly. This problem can be improved by a better feature selection method and moving object detection method. Taking them into consideration could lead to some improvement to our methods. These are remaining for our future works.

REFERENCES

[1] Y. LIU , A. Haizho and Xu Guangyou, Moving object detection and tracking based on background subtraction, *Proc. of the Society of Photo-Optical Instrumentation Engineers*, pp.62-66, 2001.

[2] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, Moving target classification and tracking from real-time video, *Proc. of the IEEE Workshop Applications of Computer Vision*, pp.8-14, 1998.

[3] D. Meyer, J. Denzler and H. Niemann, Model based extraction of articulated objects in image sequences for gait analysis, *Proc. of the IEEE Int. Conf. on Image Processing*, pp.78-81, 1998.

[4] N. Paragios and R. Deriche, Geodesic active contours and level sets for the detection and tracking of moving object, *IEEE Trans. on Pattern Analysis and Machine Inteligent*, vol.22, no.3, pp. 266-280, 2000.

[5] I. Haritaoglu, D. Harwood and L. S. Davis, W4: real-time surveillance of people and their activities, *IEEE Trans. on Pattern Analysis and Machine Inteligent*, vol.22, no.8, pp.809-830, 2000.

[6] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA., 2001

[7] Y. Satoh, S. Kaneko and S. Igarashi, Robust object detection and segmentation by peripheral increment sign correlation (PISC) image, *System and Computer in Japan*, vol.J84-D-II, no.12, pp.2585-2594, 2001.
[8] F. Cheng and Y. Chen, Real time multiple objects tracking and identification based on discrete wavelet transform,*Journal of the pattern Recognition Society*, vol.39, pp.1126-1139, 2006.
[9] B. Sugandi, H.S. Kim, J.K. Tan and S. Ishikawa, Tracking of moving object by using a low resolution image,*Proc. of the International Conference on Innovative Computing, Information and Control*, CDROM 4 pages, 2007.
[10] C. Stauffer and W. Grimson, Adaptive background mixture models for real-time tracking, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, vol.2, pp.246-252, 1999.
[11] S. M. Desa and Q. A. Salih, Image subtraction for real time moving object extraction,*Proc. of the International Conference on Computer Graphics, Imaging and Visualization*, pp.41-45, 2004.
[12] E. Stringa, Morphological change detection algorithms for surveillance applications, *Proc. of the British Machine Vision Conf.*, pp.402-412, 2000.

FIGURE 6. Two persons move in the same direction

FIGURE 7. Two persons move in the different direction

FIGURE 8. Three persons move on the scene

Table 1. Object identification result

Case	Object detected	Correct identification	Identification rates [%]
1	126	117	92.8
2	148	136	91.9
3	154	141	91.6