Omni-directional Vision based Human Motion Detection for Autonomous Mobile Robots

Hong LIU, Ning DONG, Hongbin ZHA
National Lab. on Machine Perception, Peking University,
Beijing, China, 100871
{ liuhong, dongning, zha}@cis.pku.edu.cn

Abstract - *This paper presents a novel human motion detection system using an omni-camera on a platform of autonomous mobile robot. Compared with other related works, the system using an omni-directional camera provides a larger FOV (Field of View) for motion detection, and makes a better effect using temporal differencing method based on compensation for ego-motion. Furthermore, the method, respectively combined with color information and feature points on human body, is employed to determine human motion contours, which consequently improves the performance of the temporal differencing method. Experimental results show that the system provides an efficient way to track human motion in an indoor environment.*

Keywords: Omni-directional camera; Autonomous mobile robot; Motion detection

1 Introduction

Human motion detection for autonomous mobile robots is widely used in many applications, such as human-robot interaction [1], smart surveillance [2] and motion analysis [3]. In related works, autonomous mobile robots can carry on various sensors such as ordinary cameras [4], active cameras [5, 6], and laser sensors [7, 8, 9]. Traditional visual analysis systems of human motion detection mainly employ active cameras and laser sensors, but they have actually limited FOV (Field of View), and all or a part of human body could move out of FOV. Therefore, mobile robot has to look for moving targets with searching algorithm to solve the trouble. In addition, how to deal with the registration of the images in multiple views is another difficult problem. The system in this paper employs an omni-directional camera [10], which is composed of a hyperboloidal mirror and a CCD camera, and can grab a 360º omni-directional image synchronously. It provides a 360º view angle of the environment in a single image. For its advantages of panoramic, direct and compact visual information, using omni-directional cameras for human motion detection is of great promising landscape.

Compared with the vision systems, which are mounted on robots fixedly, some new specific problems are introduced for detecting human motion when the vision system is migrated on a platform of mobile robot. The various methods for detecting human motion include background subtraction [11], template matching [12], optical flow [3] and temporal differencing [13]. Firstly, since the camera carried on the mobile robot moves everywhere, no certain background exists in the sequential images. Therefore, background subtraction is not available. Secondly, because shapes and positions of moving persons always change and is hard to be described by a template, template matching is not available, either. Thirdly, the optical flow method is not suitable for a real time system in respect with its computational complexity and low anti-noise capacity. In the same way, just using temporal differencing does not work in that the human motion could be influenced by robot motion. This paper puts forward a method to solve the problems. The system tracks the feature points between two consecutive images so that it builds up an affine transformation, which can compensate the robot's ego-motion from human motion. Based on the transformation results, the system would directly detect human motion using temporal differencing. Compared with the former methods, not only could our method solve the problem introduced by background "movement", but also increase the efficiency of motion detection with an omni-directional camera.

On the other hand, for solving the problems that omni-directional camera has introduced, for example, omni-directional image distortion, low resolution, etc, this paper presents new methods of employing other available information such as feature points on human body and color feature. Compared with regular camera, the 360º degrees information of the whole environment is included into an omni-directional image, so the local region of the mobile people tends to be a low resolution. At the same time, the motion characters are different with background and human. Therefore, we verify and compensate the human motion using the feature points on human body. In another way, color information, which is not sensitive for the image deformation such as change of rotation, translation and scale under a stable light condition, helps to verify the motion region that is acquired by the temporal differencing method.

2 System overview

The system of human motion detection in this paper is mounted on a Pioneer 2-DXe autonomous mobile robot, shown as Fig.1. By omni-vision system on the top of the mobile robot, robot grabs the sequential images in real time, and transfers the images to the processor embedded in the robot, which performs the human motion detection algorithm. The whole algorithm is implemented and tested independently by the autonomous mobile robot.

To investigate human motion in omni-images more conveniently, they need to be transformed into the cylindrical panoramic images [14], shown as Fig.2. The omni-image transformation is to provide the intuitionistic image for human-computer interaction and to process the omni-image conveniently, such as localizing mobile targets.

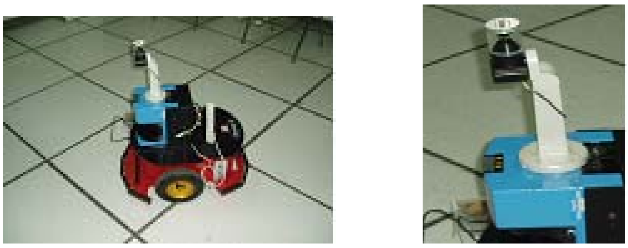


Fig.1 Human motion detection system implemented on Pioneer 2-Dxe robot

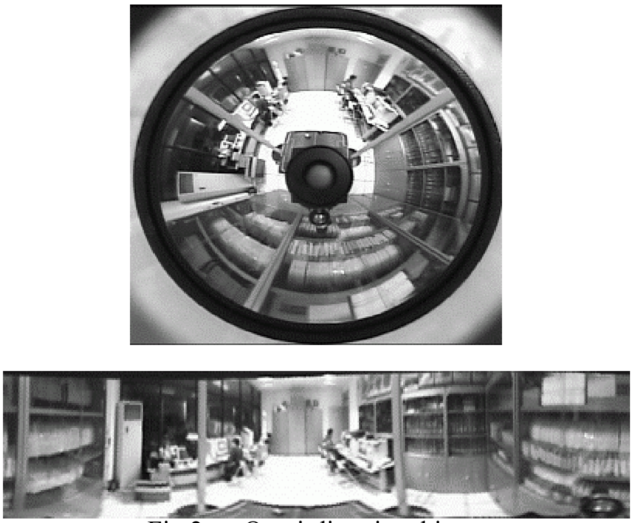


Fig.2 Omni-directional image and transformation results

In an image sequence, the system initially tracks the feature points between two consecutive images, and with feature points it builds up an affine transformation. Thus, the transformation compensates the background offset introduced by robot motion in the images. By temporal differencing, the system detects the human motion region. In this manner, according to different conditions, the system localizes the moving person respectively combined with color information and the feature points region on human body. The flow chart of the human motion detection process is shown as Fig.3.

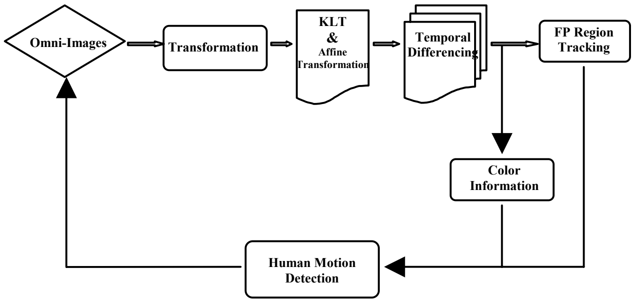


Fig.3 Flow chart of system process

3 Ego-motion compensation based on KLT method

Temporal differencing, which compares two consecutive frames and finds moving objects based on the difference, is perhaps the most intuitive and fast algorithm for moving object detection, especially when the viewing camera is static. However, when the camera moves (eg. when it is mounted on a mobile robot), straightforward differencing is not applicable because an obvious change is generated from moving the camera. There are two independent motions involved in the moving camera scenario: motions of moving objects and ego-motion of camera. Since the two motions are blended into a single image, the ego-motion of the camera should be eliminated so that the remaining motions, which are due to moving objects, can be detected. The method in this paper is that ego-motion of the camera can be estimated by tracking features between images. Ego-motion compensation is a transformation from the image coordinates of $\boldsymbol{I}^{t-1}$ to that of $\boldsymbol{I}^{t}$ so that the two images can be compared directly. The transformation can be estimated using two corresponding feature sets: a set of features in $\boldsymbol{I}^{t}$ and a set of corresponding features in $\boldsymbol{I}^{t-1}$.

We adopt Kanade-Lucas-Tomasi Method (KLT Method) for corresponding feature set selection, which is applied to select and track those feature points in the subsequent images $\boldsymbol{I}^{t}$ to find the corresponding set of feature points. There are three steps as follows:

1) Feature Selection

In fact, a single pixel cannot be tracked, so we track the windows of pixels that contain sufficient texture. The system extracts the high frequency part of pixels with Laplace transformation, and by threshold acquires the maximum ones as the feature points in a window. On the other hand, for reducing noise, the system sets the minimum distance to control the distribution of the features.

2) Feature Tracking

In sequential images, we define $\boldsymbol{I}(x, y, t)$ as the intensity of the pixel (x, y) when the time is t, and the motion of the pixel (x, y) is expressed as:

$$\boldsymbol{I}(x, y, t+\tau) = \boldsymbol{I}(x-\xi, y-\eta, t) \qquad (1)$$

The amount of motion $\boldsymbol{d} = (\xi,\ \eta)$ is called the **displacement** of the point at (x, y) between time instants t and $t+\tau$.

To solve for feature tracking, we should:

- detect the displacement so that estimate whether the windows in the previous image and the next one are the same. If not, we would discard the window.

- suppose the motion of the pixels in a window is the same, and express the motion as:

$$\boldsymbol{J}(\mathbf{X}) = \boldsymbol{I}(x, y, t+\tau) \tag{2}$$

$$\boldsymbol{I}(\mathbf{X} - \boldsymbol{d}) = \boldsymbol{I}(x-\xi, y-\eta, t) \tag{3}$$

$$\boldsymbol{J}(\mathbf{X}) = \boldsymbol{I}(\mathbf{X} - \boldsymbol{d}) + \boldsymbol{n}(\mathbf{X}) \tag{4}$$

Then the method minimizes the residue error $\boldsymbol{\varepsilon}$ defined by the following double integral over the given window W:

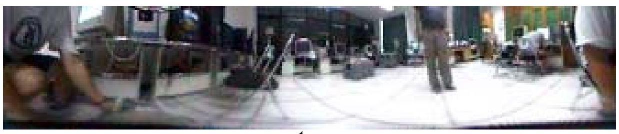
$$\boldsymbol{\varepsilon} = \int_W [I(X-d) - J(X)]^2 w\, dx \tag{5}$$

3) Solving for the Image Displacement

By minimizing the residue error $\boldsymbol{\varepsilon}$, the method solves for the image displacement. Therefore, the features in a window will be tracked well.

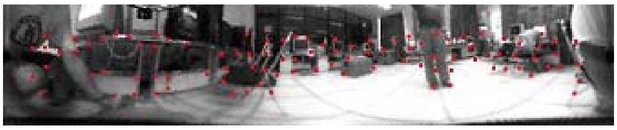
As a consequence, two corresponding feature sets are acquired: a set of features in $\boldsymbol{I}^{t-1}$ and a set of corresponding features in $\boldsymbol{I}^{t}$. By the corresponding features sets, a transformation model is estimated with Least Square between $\boldsymbol{I}^{t-1}$ and $\boldsymbol{I}^{t}$. In this paper, an affine transformation model is employed for this case. Thanks to the feature points on the background are relatively far away from mobile robot, they are approximately regarded as the feature points in the same plane regularly. Therefore, an affine transformation model is suitable to describe the ego-motion compensation. Thus, it compensates the background offset introduced by robot motion in the images. Fig.4 shows the influence introduced by ego-motion, and the results after the compensation made by our method.

(a) $\boldsymbol{I}^{t-1}$ image



(b) $\boldsymbol{I}^{t}$ image

(c) Temporal differencing between $\boldsymbol{I}^{t-1}$ and $\boldsymbol{I}^{t}$



(d) Feature points set extracted by KLT method

(e) Transformation by an affine model

(f) Temporal differencing after compensation

Fig.4 Eliminate the ego-motion on the robot by KLT method

4 Detection with temporal differencing

The approach of temporal differencing makes use of pixel-wise difference between two consecutive frames in an image sequence to extract moving regions. Temporal differencing is very suitable for dynamic environments, but generally does a poor job of extracting the entire relevant feature pixels, e.g., possibly generating holes inside moving entities.

Since the human motion detection research in this paper is based on mobile robot, the background acquired by vision system is always updated in real time. Therefore, temporal differencing is adaptive to detect the human motion after compensation based on KLT method. The temporal differencing employed in this paper as below steps:

1) Calculate the difference between the image $\boldsymbol{I}^{t-1}$ and $\boldsymbol{I}^{t}$.

2) Binarize the images differenced.

3) Smooth the image, and perform the morphologic processing including erosion and dilation. Hence, the system will eliminate the noise up to least.

4) Extract the human motion region.

The method in this paper is used in the regions with a rich enough texture, which is proposed to track corners, or windows with a high spatial frequency content, or regions where some mix of second-order derivatives was sufficiently high. Hence, the method is quite robust in the changing light condition. Experiments show that, when the mobile robot moves steadily, the accuracy of the system is more than 80%.

5 Improvement using feature points on human body

For compensating the robot motion from the image sequence, the motion detection based on KLT algorithm is employed into the system. Hence the feature points introduced by KLT are distributed into the mobile people region and non-mobile background region, which they have the different characters on motion. The experiments in some groups of image sequences indicate that, the displacement of mobile people tends to be larger than that of background.

In related work, B. Jung and G. S. Sukhatme proposed a method [13] to solve the problem, which makes use of temporal differencing combined with particle filter, so that the system detected the human motion effectively. On the same platform, their method was implemented and tested for several times, and its accuracy is up to 66.28%.

A novel method using feature points on human body is proposed in this paper, which will improve the effect based on the result using temporal differencing. The method is to track the feature point region on human body in the previous image, which is helpful to make certain the region brought by our method based on temporal differencing. The method is described as follows:

1) In the first frame of image sequence, the system selects N feature points with KLT method, and records the region of the feature points as A_1.

2) In the next image, we try to track the N feature points, and M feature points are successful to be tracked. Record the region of M feature points, as A_2. The information of human motion is obtained in the feature points.

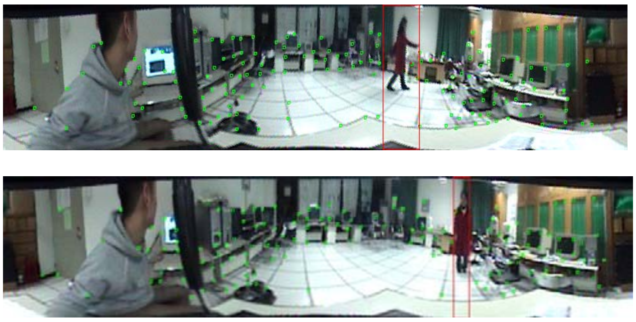
3) By temporal differencing, we obtain the results of human motion region in the image, and select the top 3 regions, K_1, K_2, and K_3, from the results, according to the area of the regions.

4) We compare the region A_2 with K_1, K_2 and K_3. The region matched best with A_2 is determined to the human motion region in the image. It is more helpful to eliminate the noise than other methods.

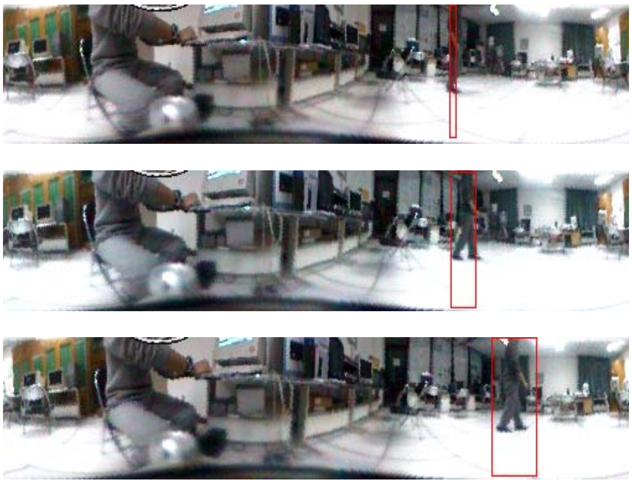
5) Repeat the step as above. We record the feature point region of human motion as A_n, which is regarded as the region compared with next image, and repeat the steps 1) - 5).

Furthermore, in the motion region obtained by the above method, the system computes the local transformation with the feature points in the above region. Then the corresponding compensation and temporal differencing are employed in the local region. In fact, the transformation tends to express the local motion more accurately.

Experiments show that, the improvement based on the feature points on human body does reduce the disturbance introduced by robot motion and noise, and increase the correct ratio effectively. Two groups of experimental results under different light conditions are shown as Fig.5:



(a) Experiment I (the feature points are shown)



(b) Experiment II (another person)

Fig.5 Results for human motion detection

The method that makes use of the feature points on the human body improves the performance, and compared with related works, the accuracy rate is increased to a large extent, which is up to 80% above.

6 Motion Detection using Color information

Though the temporal differencing based on KLT method brings the system a higher correct ratio, the result region detected by omni-camera tends to be roughly the position of human motion, since there exists some problems introduced by omni-camera itself, such as distortion, low-resolution, etc. Under the circumstance, the method is limited to localize the human motion region approximately, but the full region cannot be contained.

In fact, color information is not sensitive for the image deformation such as change of rotation, translation and scale. Under a stable light condition, motion detection benefits from the obvious color on human body. On the other hand, the color-based method provides a goal region at a high rate. Thus, when the system outputs a result after temporal differencing, we combine our method with color information of human body. In this way, not only would the method reduce the searching range and increase the efficiency, but also overcome the shortcoming that the omni-camera introduces. As a consequence, a human motion region that is more intact and more accurate will be acquired in our system.

A method of human motion detection, which combines temporal differencing method and color feature, is proposed in this paper.

6. 1 Specifying a target color

We define the initial color information by interaction in the first frame, within a region of interest (ROI) on human body. The first image is scaled by α and discretisized to the interval [0, 255] to form the normalized color components (NCC) as below:

$$r = \alpha \frac{R}{R+G+B} \tag{6}$$

$$g = \alpha \frac{G}{R+G+B} \tag{7}$$

The NCC values of each pixel within ROI are measured, which used to determine the initial color thresholds of the target object in the first frame. The distribution of ROI is regarded as a Gaussian function. Under the assumption of a normal distribution of the color values, we determine the thresholds (r_l; r_h; g_l; g_h) by computing the mean μ and the variance σ of color pixels using the equations:

$$r_{l/h} = \mu_r \mp \sigma_r \tag{8}$$

$$g_{l/h} = \mu_g \mp \sigma_g \tag{9}$$

These thresholds form a rectangular area in the NCC color space, which now specifies the color distribution to be tracked.

6. 2 Tracking object with color information

Traditionally, for each pixel at location (x, y) in the image frame the NCC values $r(x, y)$ and $g(x, y)$ are computed and compared to the target color values using the following criteria:

$$\boldsymbol{D_i} = \begin{cases} 1, \text{If } r_l < r(x, y) < r_h \text{ and } g_l < g(x, y) < g_h \\ 0, \text{else} \end{cases} \tag{10}$$

In this paper, we combine with the result of the temporal differencing method so that the performance of our system is improved. Based on the region A_K acquired by temporal differencing, we setup a region for searching with color information. The region tends to be a larger one that includes the region A_K. According to the initial color information, the method searches the points that meet the threshold requirement in the region, and the radius of the region is S. All the points are combined with different point sets by distance, which is marked as P_i, i=0, 1, 2, …, N. At the same time, a binary median filter is applied to eliminate noise and small false detections. The point set including most points is marked as the searching result P_{max}. P_{max} is recorded in an object *R_Object*, which includes the beginning and ending positions, width, length, color threshold and the information of all the points.

If the searching is failed in the region with radius S, we will go on next searching with radius S+n, and n is a step increment. The above steps are repeated on the frame, until a new *R_Object* is created. If the searching radius would reach the boundary of the image and we cannot find the point set that meets the requirement, the system returns "**Failure**".

6. 3 Adaptive color information update

When a new *R_Object* is created, the system updates the color information with the color threshold of the new *R_Object*. The system matches the result from temporal differencing and the one from color searching, and determines the final region of human motion.

Under the steady light condition and the obvious color feature, the method based on temporal differencing and color information makes the detecting results more accurate, especially with low-resolution images. Moreover, the detecting efficiency is increased. On the other hand, the method could be influenced by volatile color information under the environment. For example, the color of human body should meet that it is almost homochromous.

7 Experimental results

Computations in Pioneer robot are performed on embedded computers (Siemens C166), with 128 MB memory. The omni-camera Vstone-VCP2M/PCI is mounted on the robot can grab and process the images with 1.2 frames per second. The resolution of the original images is 640×480, and it becomes 806×166 after transformation.

Table 1 Comparison of experimental results about human motion detection method (Ratio = True / Motion)

	Light On/Off	*Motion*	*True*	*False*	*Lost*	*Ratio (%)*
B.J.	On	172	114	12	46	66.28
TD	Off	980	760	145	75	77.6
TD	On	980	776	137	67	79.2
CFP	Off	220	176	30	14	80
CFP	On	220	183	26	11	83.2
CC	Off	260	219	31	10	84.2
CC	On	260	242	15	3	93.1

Note: B.J. = Bang. Jung's method
TD = Temporal Differencing
CC = Combined with Color
CFP = Combined with Feature Points

The methods in this paper were implemented and tested on the platform of a Pioneer 2-DXe autonomous mobile robot, and their experimental results are shown as Table 1. ***Motion*** is the number of moving objects over the total number of frames. And ***True*** and ***False*** are the number of correct detections and the number of falsepositives. ***Ratio*** shows the percentage of moving objects correctly detected. Compared with related work, it is obvious that the system detects the human motion with higher correct ratio than other methods. Moreover, under the different conditions, the system improves the detection performance with the methods respectively.

According to the experimental results, comparison chart Fig.6 about the correct ratio of human motion detection as below:

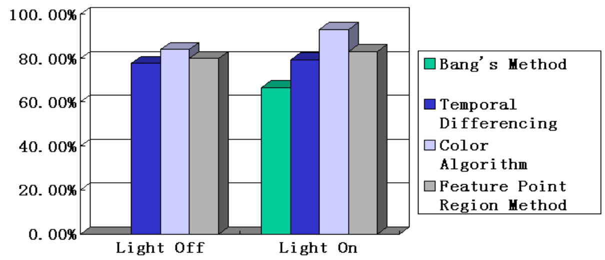


Fig.6 Comparison chart about the methods' performance

8 Conclusion

We propose a system to detect moving persons from a mobile robot using an omni-directional camera in an indoor environment. The omni-directional camera is used to obtain 360º view images of the scene so the system can enlarge the monitored area. Based on KLT method, the system compensates the background offset introduced by robot motion in the images, and by temporal differencing detects the human motion region. At the same time, for solving for the problems introduced by omni-directional vision, this paper employs color features and feature points region on human body so that the system improves the detection performance, and increases the accuracy of human motion detection. Experiments show that, in an indoor complicated environment, the system in this paper can detect human motion quickly and accurately.

References

[1] K. Tanaka, K. Yamano, K. Kondo, K. Kimuro, A vision system for detecting mobile robots in office environments, Proceedings of IEEE International Conference on Robotics and Automation, pp. 1583-1591, 2004.

[2] C.R. Wren, A. Azarbayejani, T. Darrell, and A.P. Pentland, Pfinder: Real-time tracking of the human body. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 780-785, July 1997.

[3] D. M. Gavrila, The visual analysis of human movement: A survey, Computer Vision and Image Understanding, 73(1): pp. 82-98, 1999.

[4] P. Nordlund, T. Uhlin, Closing the loop: Detection and pursuit of a moving object by a moving observer, Image and Vision Computing, pp. 328-331, 1996.

[5] D. Murray, A. Basu, Motion tracking with an active camera, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 541-542, 1994.

[6] G. L. Foresti, C. Micheloni, A robust feature tracker for active surveillance of outdoor scenes, ElectronicLetters on Computer Vision and Image Analysis, pp. 25-28, 2003.

[7] D. Schulz, W. Burgard, D. Fox, A.B. Cremers, Tracking multiple moving targets with a mobile robot using particle filters and statistical data association, Proceedings of IEEE International Conference on Robotics and Automation, pp. 1665 – 1670, 2001.

[8] D. Anguelov, D. Koller, E. Parker, S. Thrun, Detecting and modeling doors with mobile robots, Proceedings of IEEE International Conference on Robotics and Automation, pp. 3777 – 3784, 2004.

[9] B. Jensen, R. Siegwart, Using EM to detect motion with mobile robots ， Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1518-1523,2003.

[10] G. Cielniak, M. Miladinovic, D. Hammarin, L. Goranson, A. Lilienthal, T. Duckett, Appearance-based tracking of persons with an omnidirectional vision sensor, Omnivis Workshop with ECCV: Omnidirectional Vision and Camera Networks, pp. 84-91,2003.

[11] S. McKenna, S. Jabri, Z. Duric, Tracking groups of people, Computer Vision and Image Understanding, 80:pp. 42-56, 2000.

[12] A. Lipton, H. Fujiyoshi, R. Patil, Moving target classification and tracking from real-time video, In Proc. IEEE Workshop on Application of Computer Vision, Princeton, NJ, 1998.

[13] B. Jung, S. Sukhatme, Detecting Moving Objects using a Single Camera on a Mobile Robot in an Outdoor Environment, the 8th Conference on Intelligent Autonomous Systems, pp. 980-987,2004.

[14] Y. Onoe, K. Yamazawa, N. Yokoya, Visual surveillance and monitoring system using an omnidirectional video camera, Proc. IEEE Intl. Conf. on Pattern Recognition, Brisbane, Australia: IEEE Computer Society, pp. 588-592,1998.

[15] S. Arakil, T. Matsuoka, H. Takemura, N. Yokoya， Real-time tracking of multiple moving objects in moving camera image sequences using robust statistics， Proceedings of Fourteenth International Conference on Pattern Recognition, 1998.

[16] C. Tomasi, T. Kanade, Detection and tracking of point features, Technical Report CMU-CS-91-132, Carnegie Mellon University, Pittsburgh, PA, pp. 147-150, April 1991.

[17] C. Schlegel, C. Illmann, H. Jaberg, M. Schuster, R. Worz, Vision based person tracking with a mobile robot, Proceedings of the Ninth British Machine Vision Conference, pp. 418-427,1998.