Motion Detection for Multiple Moving Targets by Using an Omnidirectional Camera

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

Abstract

This paper proposes a new real-time omnidirectional vision system for tracking multiple targets in indoor environments. Since traditional cameras suffer from the problem of having limited visual fields, an omnidirectional camera is used in our system to obtain the 360º view images of the global scene. First, omnidirectional images are changed into the cylindrical panoramic images. Then an adaptive background subtraction method is utilized to segment the moving regions. Human bodies qua the targets are tracked and identified by using their color information. Experimental results show that the proposed system performs well in indoor, complex environments.

1. Introduction

Computer vision systems for moving targets tracking are widely used in many applications such as smart surveillance, virtual reality, advanced user interfaces, motion analysis and model-based image coding [1]. One of the most important issues in visual surveillance is to expand the view fields of visual sensors.

Most conventional visual systems used single fixed camera to take image sequences of the scene [2, 3]. The disadvantage of using a fixed single camera is that the area captured by the camera is relatively narrow due to the limitation of its view fields. While a person across the outer side of the camera's view field, the tracking may fail.

Using multiple regular cameras located in several areas can partly solve the above problem [4, 5]. When a moving target disappears from the capture of one camera, another one can continuously track it. However, since the images captured by different cameras are in different spatial coordinates, establishing feature correspondence between them is rather difficult.

Another solution for enlarging view fields is to use mechanically controlled single camera such as a rotational camera [6]. The sequence of the images acquired by rotation is stitched to a panoramic view of the scene. Also, there is a fatal disadvantage in this solution due to the mechanical control. It cost too much time to obtain a panoramic image that it is hard to track targets in real-time.

Recent years, several researchers developed a new type of visual sensor, omnidirectional camera. It provides a 360^{o} view angle of the environment in a single image. It has been applied to some applications such as autonomous navigation [7] and surveillance [8, 9]. For its advantages of panoramic, compact visual information and with directive features, using the omnidirectional camera for targets tracking is of great promising.

Our system employs a hyperboloidal omnidirectional camera [10], which is composed of a hyperboloidal mirror and a CCD camera. The camera is set in the center of our lab, to "see" in all directions at any instant in time. In our system, the targets are assumed to be human bodies. Before tracking, the scene model is first built by camera's observing the whole scene without moving targets in it. If a human body was detected to enter the scene, the system builds up a color model for him/her.

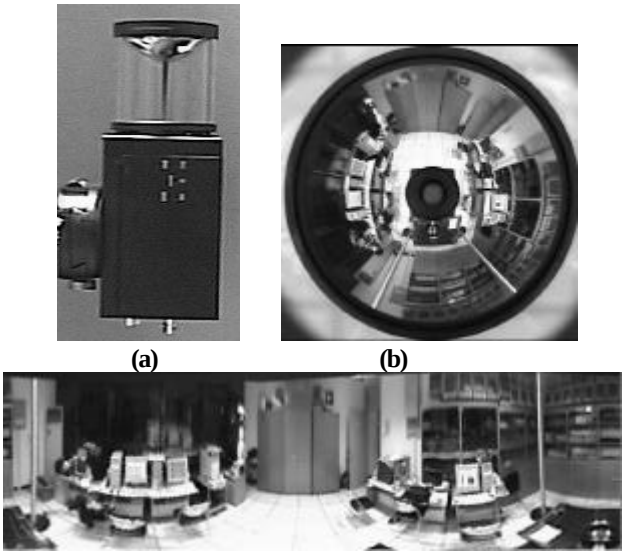
While there are multiple human bodies moving in the scene, they can also be tracked whether they are in groups or separated from one another, according to their color models.

The remainder of this paper is organized as follows: Section 2 describes the algorithm to transform the omnidirectional image into its cylindrical panoramic view. An adaptive background subtraction method for obtaining the moving regions is described in section 3. Section 4 presents a color-based model to segment and track human bodies while they are in groups or separated from one another. Experiments and conclusions are given in section 5 and section 6, respectively.

2. Omnidirectional image transformation

The omnidirectional camera employed in our system is composed of a hyperboloidal mirror and a CCD camera, as illustrated in Fig.1 (a). Since the circular image directly acquired by the omnidirectional camera is naturally distorted, it needs to be transformed into the cylindrical panoramic image for target analysis.

A circular omnidirectional image and its cylindrical panoramic image are shown in Fig.1 (b) and (c).



(a) **(b)**

(c)

Figure 1. (a) Omnidirectional camera (b) Omnidirectional image (c) Cylindrical panoramic image

Each pixel in the cylindrical panoramic image can be exactly calculated from the circular omnidirectional image, through complex equations related to the parameters of the camera [9]. Since the computation costs much time, we present a fast method for the transformation. For each pixel in the perspective image, which is represented by coordinate (x, y), its corresponding coordinate (x_1, y_1) is located in the omnidirectional image (see Fig.2).

This transformation yields the following equations.

$$q = x / r_1 \quad (1)$$

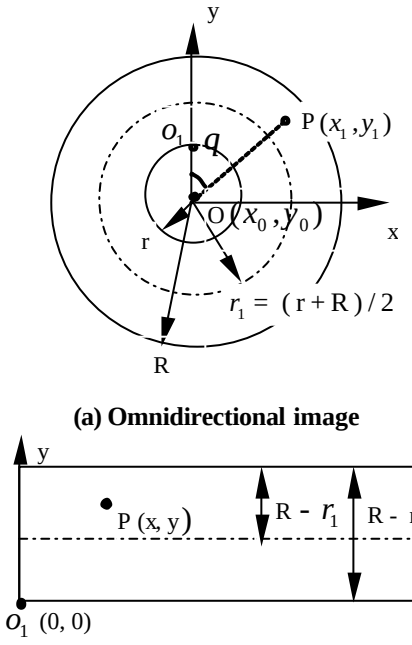
$$r_1 = (r+R)/2 \quad (2)$$

$$x_1 = x_0 + (r + y)\sin q \quad (3)$$

$$y_1 = y_0 + (r + y)\cos q \quad (4)$$

Here, q is the intersection angle between OP and y-axis, $O(x_0, y_0)$ is the center of the circles in the omnidirectional image. R and r is the radius of the large circle and small circle. Only the fields between the large circle and the small circle are valid for the transformation. Here, r_1 is the radius of the medium circle marked with broken line, which is in the middle of the large and small circle.

During the transformation, the torus region of the original image is "cut" through y-axis, and stretched to a rectangle, i.e. the cylindrical image. The width of the cylindrical image equals to the perimeter of the circle marked with broken line in the original image. The purpose of the transformation is to make the cylindrical image looking more concordant, since the torus fields between the large and medium circle will be compressed, while the fields between medium and small circle will be extended. Experiments show that this method of transformation is fast and effective.



(a) Omnidirectional image

(b) Cylindrical panoramic image

Figure 2. Coordinate transformation for omnidirectional images to cylindrical panoramic images

3. Background modeling and foreground regions detection

In our system, the omnidirectional camera is located fixedly in a laboratory room, which may contain some small dynamic factors such as floating of curtains, blinking of computer screens, and variations of illumination. Also, when people moving in the room, their mirror images maybe appear on the glass windows. These problems make it hard to obtain the accurate foreground region by simply subtracting a new frame from a settled background. Here, we describe an adaptive background method to produce a foreground segmentation mask, which is similar to the work in [11, 12].

3.1. Background modeling

Every point in the background is assumed to have a mean color value and a distribution about that value. Before any person entering the environment, the omnidirectional camera observes the scene for several seconds, and then the initial background model can be built up as follows. We define m_i as the mean color value of a point i, and s_i^2 as the covariance of that

point's distribution. Thus, (m_i, s_i^2) can be stored as the color background model for the point *i*. Since a color pixel has three components of R, G and B, m_i and s_i^2 is defined as vectors:

$$m_i = (m_i(r), m_i(g), m_i(b)) \quad (5)$$

$$s_i^2 = (s_i^2(r), s_i^2(g), s_i^2(b)) \quad (6)$$

The initial background model cannot be expected suitable for long periods of time due to the variations of the scene. For each new frame *t*, $y_i(t)$ is the current color of pixel *i*. The background model is update on-line using the following formulas:

$$m_i(t+1) = \begin{cases} (1-\alpha) m_i(t) + \alpha y_i(t+1), & \text{If } i \text{ in background} \\ m_i(t), & \text{If } i \text{ in foreground} \end{cases} \quad (7)$$

$$s_i^2(t+1) = \begin{cases} (1-\alpha) s_i^2(t) + \alpha (y_i(t+1) - m_i(t+1))^2, & \text{If } i \text{ in background} \\ s_i^2(t), & \text{If } i \text{ in foreground} \end{cases} \quad (8)$$

Here, the constant α ($0 < \alpha < 1$) controls the adaptation rate.

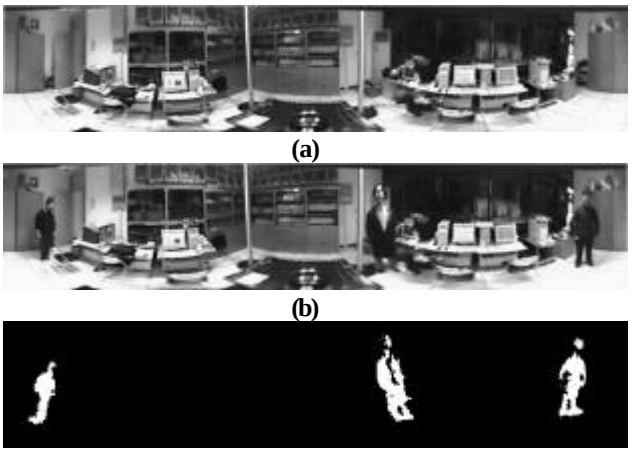
3.2. Foreground region detection

Once we obtain the adaptive background model, each new frame can be subtracted from it to determine foreground regions.

D_i is defined as the binary value of the subtract result of pixel *i*. If D_i=1, the pixel is classified as the foreground. Otherwise, the pixel is classified as the background. We obtain the D_i as following:

$$D_i = \begin{cases} 1, \text{If } (y_i(r) - m_i(r) > 3s_i(r)) \\ \quad \text{Or } (y_i(g) - m_i(g) > 3s_i(g)) \\ \quad \text{Or } (y_i(b) - m_i(b) > 3s_i(b)) \\ 0, \text{else} \end{cases} \quad (9)$$

The binary images obtained directly from subtractions usually contain isolated points or lines caused by those dynamic factors. Mathematical morphological filters are used to erase them. First a 3 × 3 erode filter is used to wipe off the isolated points and lines, then a 3×3 expand filter is used to recover the exact foreground region. Fig. 3 shows the result of background subtraction.



(a)

(b)

(c)

Figure 3. Adaptive background subtraction (a) Background (b) The scene with persons (c) Foreground

4. Tracking multiple human bodies

Tracking persons in our system is to find the correspondence between the human bodies and the foreground regions, and mark them with occlusions.

Persons are assumed entering the scene one by one, thus his/her model can be initialized through his/her position and color information. Vector HB (ID, x, y, r, g, b) is defined to represent a human body. Here, ID is used to identify each person. Each person gets a new ID once he/she entering the scene, and when he/she leaves, the ID will be deleted in the system. While the person is moving in the scene, (x, y) record the coordinate of the person's cg (center of the gravity), and (r, g, b) record the mean color value of all the pixels belong to him /her.

While a person moving along in the room, its corresponding foreground region can be obtained simply by matching the human body's cg to the nearest foreground region. Then update the human body's cg by the cg of the region. If more than two human bodies share one region, they are considered to be in one group. When anybody leaves the group, his/her ID can be recognized through its color model.

5. Experiments

The experimental system is set in our lab environment, which contains many small dynamic factors. The moving targets are assumed to be human bodies. An omnidirectional camera is settled in the center of the room. Before tracking, the environment scene model is built by camera's observing the whole scene without people for several seconds. If human bodies are detected entering the scene, their corresponding detected regions are marked by bounding boxes, until they leave the room.

Furthermore, the trail of each human body's cg can be obtained.

Omnidirectional images acquired in 24-bit RGB model are in size of 480×480 resolution, and cylindrical panoramic images after transformations are in 824×166. The system runs at average 8Hz on a Pentium IV 1.8GHz PC.

Detection results for multiple moving human bodies are shown in Fig. 4. Three persons are walking, meeting and separating in the room. Each human body assigned by a unique ID, is marked through an occlusion. Trail of each human body's cg is given in Fig. 5. The frames marked with (a), (b), (c), and (d) in Fig. 5 correspond to the four images in Fig. 4.

6. Conclusions

We have proposed a new real-time visual system based on omnidirectional vision, for detecting and tracking multiple targets in indoor environment. An omnidirectional camera is used to obtain 360º view images of the scene to enlarge the monitored area. An adaptive subtraction method is utilized to improve the robustness of the system in dynamic environments. Moreover, a model of human body is presented for tracking multiple people, whether they are in groups or separated from each other. Experiments show that the system can track multiple moving targets with large view fields in a dynamic environment in real-time.

Detections may fail on the occasions that if two people dressed in similar color move together to form a group. When they separated from each other, the system cannot recognize the exact ID of each one. This is the problem we plan to solve in our future work.

Acknowledgements:

This work is supported by Chinese "863" Project (Project No.: 2001AA422200) and NSFC (Project No.: 60175025), P.R. China.

References

[1] D.M. Gavrila, "The visual analysis of human movement: A survey", *Computer Vision and Image Understanding*, 1999, vol. 73, no. 1, pp. 82-98.

[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*, July 1997, vol. 19, no. 7, pp. 780-785.

[3] I. Haritaoglu, D. Harwood, and L.S. Davis, "W4: Real-time surveillance of people and their activities", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000, vol.22, no. 7, pp. 809-830.

[4] A. Utsumi, H. Mori, J. Ohya, and M. Yachida, "Multiple-view-based tracking of multiple humans", *Proc. of IEEE Intl. Conf. on Pattern Recognition*, Brisbane, Australia, 1998, pp. 597-601.

[5] Q. Cai and J.K. Aggarwal, "Tracking human motion using multiple cameras", *Proc. IEEE Intl. Conf. on Pattern Recognition*, Vienna, 1996. pp. 68-72.

[6] A. Krishnan, and N. Ahuja, "Panoramic image acquisition", *Proc. IEEE Computer Vision and Pattern Recognition*, 1996, pp. 379-384.

[7] L.Delahoche, C. Pegard, B. Marhic, and P. Vasseur, "A navigation system based on an omnidirectional vision sensor", *Proc. Intl. Conf. on Intelligent Robotics and Systems*, 1997, pp. 718-724.

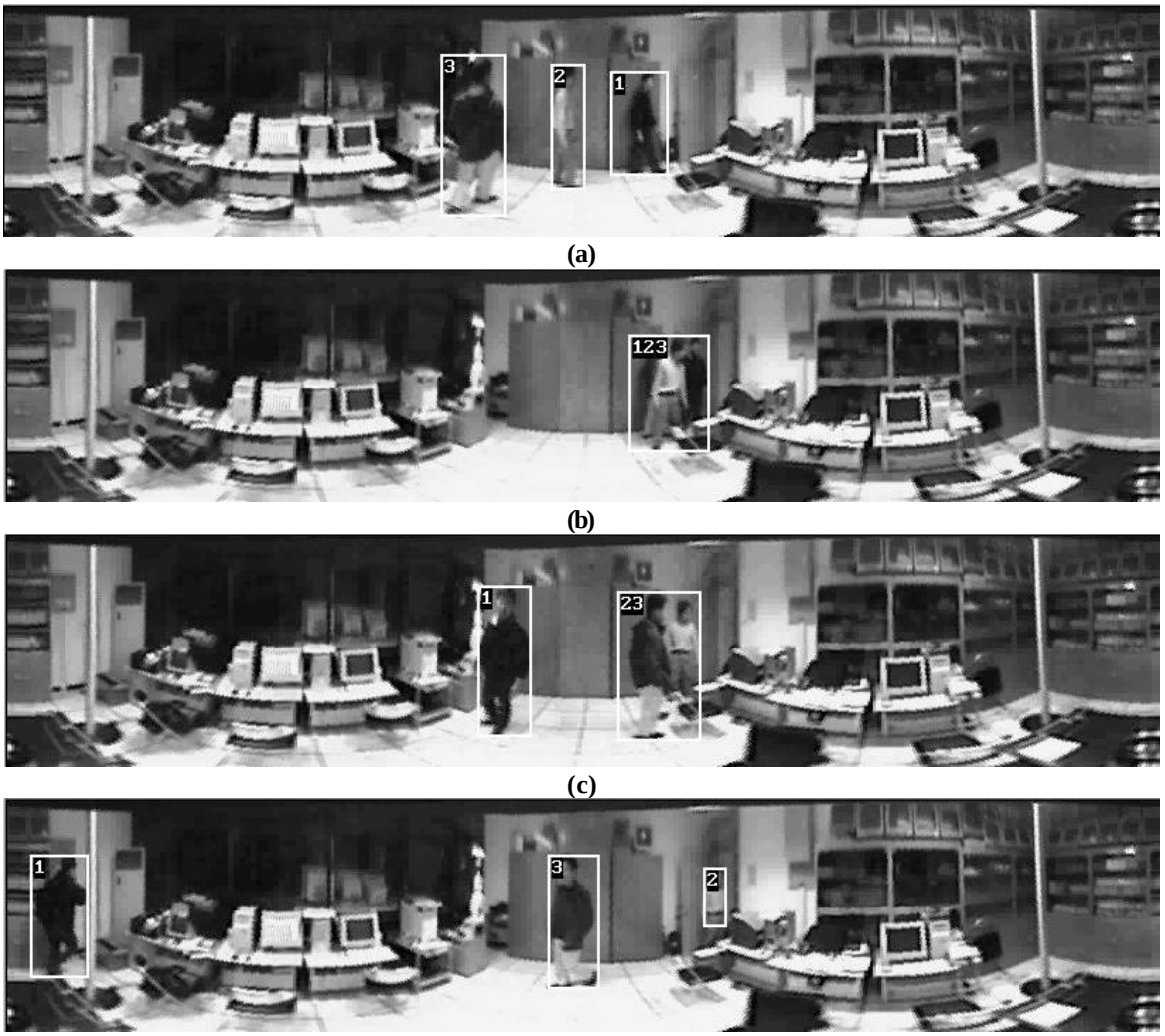
[8] Y. Onoe, K. Yamazawa, N. Yokoya, and H. Takemura, "Visual surveillance and monitoring system using an omnidirectional video camera", *Proc. IEEE Intl. Conf. on Pattern Recognition*, 1998, pp. 588-592.

[9] T. Boult, R. Micheals, A. Erkan, P. Lewis, C. Powers, C. Qian, and W.Yin, "Frame-rate multi-body tracking for surveillance", *Proc. DARPA Image Understanding Workshop*, Monterey, California, 1998.

[10] H. Ishiguro, "Development of low-cost compact omnidirectional vision sensors and their applications", *Proc. of Intl. Conf. on Information Systems, Analysis, and Synthesis*, 1998, pp. 433-439.

[11] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people", *Computer Vision and Image Understanding*, 2000, vol. 80, pp. 42-56.

[12] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1999, pp. 246-252.



(a)

(b)

(c)

(d)

Fig. 4. Results of detection of moving human bodies

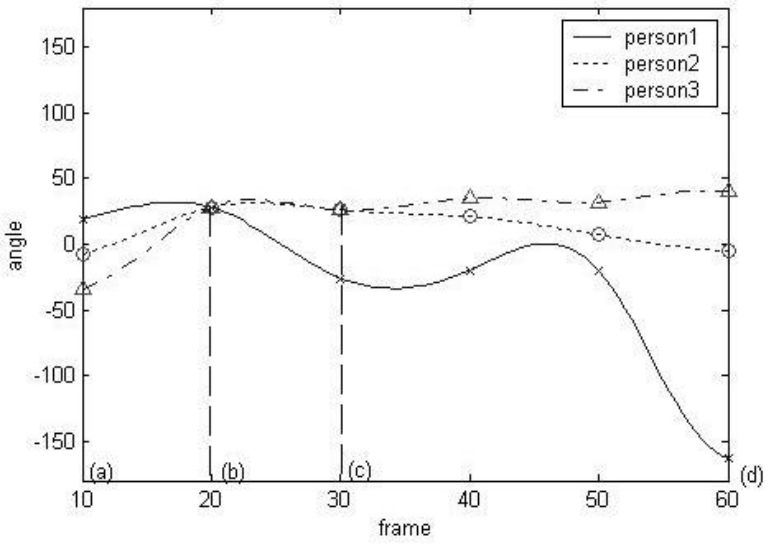


Fig. 5. Trails of moving human bodies