MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
13-1

Cooperative Usage of Monocular Camera and Omnidirectional Camera for Segmenting Moving Humans

Thatsaphan Suwannathat and Masahide Kaneko
The University of Electro-Communications
1-5-1 Chofugaoka, Chofu-shi
Tokyo 182-8585, JAPAN
touch@radish.ee.uec.ac.jp, kaneko@ee.uec.ac.jp

Abstract

This paper presents a system to track moving humans in an indoor environment. We propose a cooperative approach of monocular camera and omnidirectional camera to the problem of finding humans by a mobile robot. A cooperation of fixed and moving platforms is proposed for this approach, which features background subtraction, pixel classification, color blob merging, color blob relation and segmentation on a moving platform. Experimental results are given for detecting moving humans using two cooperative platforms.

1 Introduction

Flexible vision systems can provide an extremely rich sensing modality for sophisticated robot platform. We propose a cooperative approach to the problem of finding humans in an indoor environment. A popular approach to tracking humans using a stationary camera is to look for regions of change in the scene. This can be done using consecutive frame differencing [1, 2], or more popularly, by comparing the current frame against a back ground model [3, 4]. The background difference methods differ from each other in the way the background model is built. In [4] the model is built assuming a normal distribution for the gray values at each pixel. Once change detection is done, either by comparing with a background model or by subtracting consecutive frames, most approaches, e.g:[3, 4] perform a correspondence step to label each of regions as an object.

Since a conventional camera is limited in its field of view, a real-time omnidirectional camera which can acquire an onmidirectional (360 degree) field of view could be applied in a variety of fields such as autonomous navigation for mobile robot [5,6] and human interaction problems [7].

In this paper, we describe a framework to detect segment and track a sequence of human moving in an indoor environment. The common set of assumptions in tracking human motion includes small image motion and non-rigid of the moving object. To detect a moving object, we focus on cooperative behavior involving cameras residing on different platforms. Figure 1 illustrates the proposed system.

Our system consists of fixed and moving platforms as shown in Figure 1. Fixed platform uses a stationary monocular camera to look for regions of change.

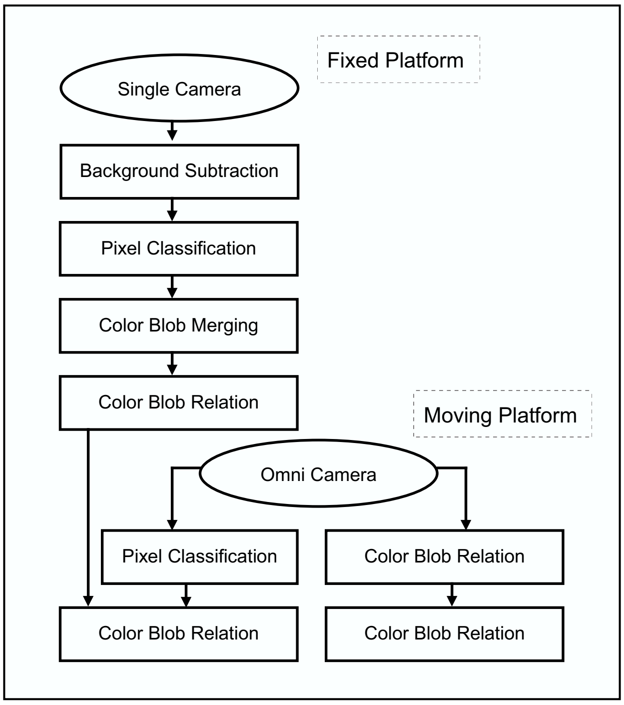


Figure 1. Proposed system

In this process, 8 colors classification is used to classify individual pixels into coherent blobs. The blobs are then grouped to form the moving human. The blob level data which is detected in this process will send to a moving platform (mobile robot) which navigates by wall following detection using an image acquired from omnidirectional camera. Using the blob level data which is acquired from fixed platform, the moving platform can detect the moving humans that appear in an omnidirectional camera.

By taking advantage of the unique feature and properties of the conventional and omnidirectional camera, we hope to overcome the inherent disadvantages of each, resulting in a combined sensor system which is more effective than either sensor is individually.

2 Background Subtraction

Background subtraction is performed in each frame to segment the foreground image region, as follows: at each image pixel (x,y) of a given input frame, the change in pixel intensity is evaluated by computing the Mahalanobis distance from the Gaussian background model.

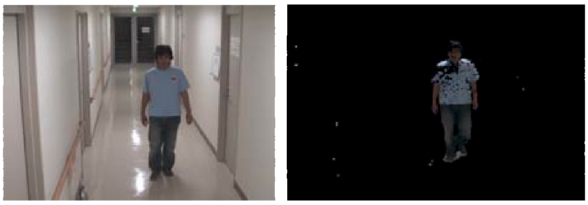
$$\delta(x,y) = \frac{v(x,y) - \mu(x,y)}{\sigma(x,y)} \qquad (1)$$

where $v(x,y)$ is color distribution of each pixel at image coordinate (x,y), $\mu(x,y)$ is mean and $\sigma(x,y)$is standard deviation of each color channel

Foreground image $F(x,y)$ is defined by the maximum of the three distance measures, δ_H, δ_S and δ_V for H, S, V channels

$$F(x,y) = \max[\delta_H(x,y), \delta_S(x,y), \delta_V(x,y)] \qquad (2)$$

F is then thresholded to make a binary image. At this stage, morphological operations are performed as a processing step to remove small regions of noise pixels. Figure 2 shows an input image and the result of background subtraction.



(a) (b)

Figure 2. (a) Input image, (b) Background subtraction

3 Classification of Individual Pixel

The objective for classification of individual pixel is to classify each RGB pixel in a foreground image as having a color in the specified range. In this paper, we classified RGB pixel value into 8 colors as shown in Figure 3.

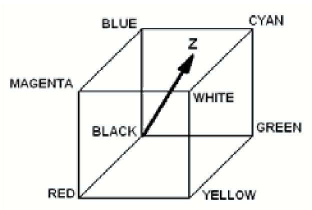


Figure 3. RGB vector classification

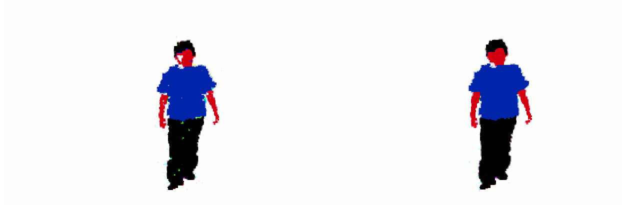
In order to perform this classification, it is necessary to have a measure of similarity. We used Euclidean distance. Let $\boldsymbol{z}$ denote an arbitrary point in RGB space, $\boldsymbol{a}$ denote average color we wish to classify. We say that $\boldsymbol{z}$ is similar to $\boldsymbol{a}$ if the distance between them is less than a specified threshold D_0. The Euclidean distance between $\boldsymbol{z}$ and $\boldsymbol{a}$ is given by

$$D(\boldsymbol{z},\boldsymbol{a}) = \|\boldsymbol{z} - \boldsymbol{a}\|$$

$$= [(\boldsymbol{z}-\boldsymbol{a})^T(\boldsymbol{z}-\boldsymbol{a})]^{\frac{1}{2}}$$

$$= [(z_R - a_R)^2 + (z_G - a_G)^2 + (z_B - a_B)^2]^{\frac{1}{2}} \qquad (3)$$

where subscripts R,G and B, denote the RGB components of vectors $\boldsymbol{a}$ and $\boldsymbol{z}$, respectively.



(a) (b)

Figure 4. (a) pixel classification, (b) Color blob merging.

4 Color Blob Merging Procedure

Merging blobs is a region growing procedure. Two blobs are merged by the following criteria; blobs A_i and A_j are merged only if the following criteria are satisfied.

1. Two blobs should be similar in color, where the similarity is defined by Mahalanobis distance δ_Φ of color feature Φ between the blobs A_i and A_j as follows.

$$\delta_\Phi = (\Phi_i - \Phi_j)^T \Sigma_\Phi^{-1} (\Phi_i - \Phi_j) \qquad (4)$$

$$\Phi = [\mu_R, \mu_G, \mu_B]^T$$

   where Σ_Φ is the covariance matrix of color channels for all the blobs in the image. If δ_Φ is less than a threshold T_Φ, blobs A_i and A_j are similar in color.
2. Two blobs should be adjacent.
3. A small blob surrounded by a single large blob is merged to it.

Figure 4 shows an example of pixel-level classification and color blob merging.

5 Color Blob Relation

A set of adjacent blobs constitutes a human body part. In order to assign the blobs to a certain human body part, we use a hierarchical human body model [8]. A human body is represented by two body parts: upper body and lower body.

The two body parts are initialized in the image as follows; the upper body is defined by the vertical range from the top of 0.45 times the height of human. The lower body is defined by the rest of vertical range. To assign each blob to a body part, comparing the centroid of each blob to the human body parameters.

6 Moving Human Segmentation on a Moving Platform

Using blob level data which is acquired from the fixed platform, a moving human that appears in an omnidirectional image which is obtained from an omnidirectional camera on a moving platform can be detected.

An omnidirectional image is classified into the color class according to the blob level data which is obtained from a fixed platform. Each color blob data is projected both vertical and horizontal by a 1D Gaussian. A mixture of the 1D Gaussian is used to model a moving color blob. Frame-to-frame update of these Gaussian parameters is then segmented into a set of the upper and lower body translation of a human.

7 Robot Navigation by Wall Following using Omnidirectional Camera

7.1 Concept

To navigate the mobile robot in a structured environment (like corridor, hallway, etc.) with a single omnidirectional camera, we use the wall following [6] method. The omnidirectional image is obtained from an omnidirectional sensor, which has the prominent feature in sensing the surrounding image at the same time. When the wall image from the camera is transformed by the polar coordinate transformation, the straight lines between the wall and the floor appear in the curve lines after transformation. The peak point represents the distance and the direction between the robot and the wall.

7.2 Method

The mobile robot navigation method by using the polar coordinate transformation from omnidirectional image can be described as follows.

STEP1 Polar Mapping
STEP2 Peak Detection
STEP3 Robot Controlling

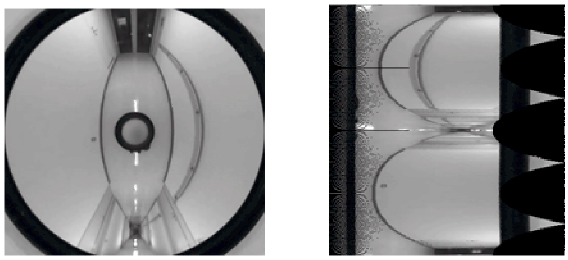
7.2.1 Polar Mapping

The omnidirectional image, obtained from an omnidirectional sensor, is transformed from Cartesian coordinate to Polar coordinate using

$$r = \sqrt{(x - x_h)^2 + (y - y_h)^2} \quad (5)$$

$$\theta = \arctan\frac{y - y_h}{x - x_h} \quad (6)$$

where r is the distance from the center of image to a coordinate, θ is azimuth angle, (x,y) is a coordinate on image plane, and (x_h,y_h) is the center of image plane. By the above equations, the pixel at (x,y) is converted into (r, θ) coordinate as shown in Figure 5



(a) (b)

Figure 5. (a) Scene from camera, (b) Polar coordinate.

7.2.2 Peak Detection

This process starts from making a polar coordinate image into binary image. At the moment some noise has to be eliminated by morphological erosion and dilation. The peak point can be detected using *x-y* projection. The peak point means the shortest distance from the centre of camera as shown in Figure 6.

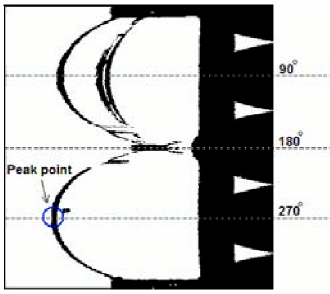


Figure 6. Peak point and reference angle.

7.2.3 Robot Controlling

In controlling, the robot is always kept for running parallel to the wall that is on left or right side. While robot is running, the peak point on polar plane has to maintain to be located at 90 or 270 degree angle.

8 Experiment and Results

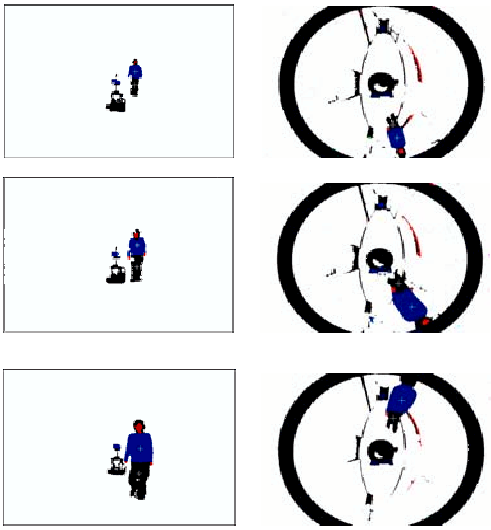
In our experimental system, we mounted an omnidirectional camera on a mobile robot used as moving platform equipment. And the other conventional camera mounted on a camera stand was used as fixed platform. The communication between two platforms is through a client server model over a wireless Ethernet link. System programs run separately on the two platforms. Only color blob parameter data was transmitted from fixed platform to moving platform.

Figure 7 shows an overview of our experimental equipments and environment. We have tested our system for human walking situation. The images used in this work were 720x480 pixels in size, obtained at a rate of 15 frames/sec.

Figure 8 shows result image sequence obtained from fixed and moving platform. The first column of Fig.8 shows the subsampled frames of the output from fixed platform. And the second column represents the corresponding frames of the output from moving platform. The upper and lower body of moving human are correctly detected both in the fixed and moving platform.

Figure 9 shows example sequences of moving humans walking cross each other. The first column shows the original input sequences acquired from omnidirectional camera on a moving platform. The second column is the corresponding sequences of panoramic image of first column which shows individual humans are correctly segmented and tracked.

Figure 7. Experimental equipments and environment.

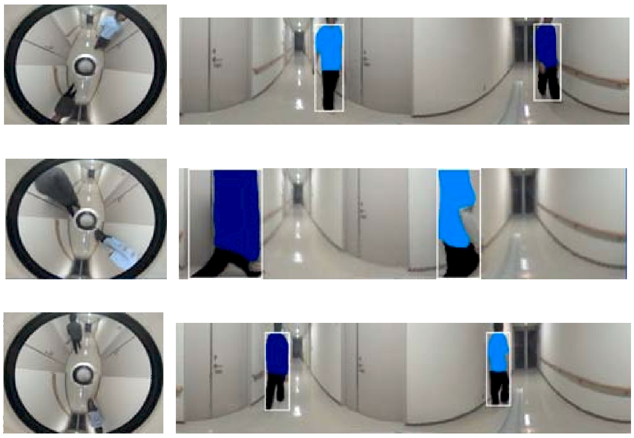


(a) (b)

Figure 8. (a) Sequences of tracked human from fixed platform. (b) Sequences of tracked human from moving platform.

9 Conclusion

This paper has presented a cooperative approach to the problem of finding human in an indoor environment. The contributions of this paper are as follows; 1) Moving human detection on a fixed platform by comparing the current frame against a back ground model, 2) Moving human detection on a moving platform using color blob data obtained from a fixed platform, 3) Mobile robot navigation using wall following detection. Experiments have shown that the moving human is tracked across the sequence. On-going and future work includes the following topics: 1) Improvement of the color blob relation, 2) Tracking of multiple humans under occlusion environment.



(a) (b)

Figure 9. Example input of moving humans walking cross each other. (a) Original image sequence acquired form omnidirectional camera. (b) Tracked humans image sequence (panoramic image).

References

[1] Q. Zang and R. Klette, "Robust Background Subtraction and Maintenance," *Pattern Recognition, 17th International Conference on Pattern Recognition(ICPR'04)* , Vol. 2, pp. 90-93, 2004.

[2] Lipton, A.J., H. Fujiyoshi, R.S. Patil, "Moving Target Classification and Tracking from Real-time Video," *Proc. of ARPA Image Understanding Workshop*, pp. 129-136, 1998.

[3] S. Park and J.K. Aggarwal, "Segmentation and Tracking of Interacting Human Body Parts under Occlusion and Shadowing," *Workshop on Motion and Video Computing (MOTION'02)*, pp. 105-111, 2002.

[4] Kanade, T. *et. al.,* "Advance in Cooperative Multi-Sensor Video Surveillance," *Proc. of ARPA Image Understanding Workshop*, pp. 3-24, 1998.

[5] H. Koyasu, J. Miura and Y. Shirai, "Mobile Robot Navigation in Dynamic Environments using Omnidirectional Stereo," *Proc. of IEEE Int. Conf. on Robotics&Automation,* pp. 893-898, 2003.

[6] T. Juchim and K. Chamnongthai, "Mobile Robot Navigation by Wall Following Using Polar Coordinate Image from Omnidirectional Image Sensor," *IEICE Trans. Inf.&Syst.*, Vol. E85-D, No. 1, pp. 264-274, 2002.

[7] Y. Yagi and M. Yachida, "Omnidirectional Sensing for Human Interaction", *Proc. of Third Workshop on Omnidirectional Vision (OMNIVIS'02),* pp. 121-127, 2002.

[8] S. Park and J.K. Aggarwal, "Recognition of human interaction using multiple features in grayscale images," *Proc. of IEEE Int. Conf. on Pattern Recognition (ICPR'00),* Vol. 1, pp. 1051-1054, 2000