Detecting and Tracking Moving Object Using an Active Camera

Kye Kyung Kim, Soo Hyun Cho, Hae Jin Kim and Jae Yeon Lee
Electronics Telecommunications Research Institute, Korea
{kyekyung, marisan, jsh62860, leejy}@etri.re.kr

Abstract **— This paper is to introduce detecting and tracking moving object using an active camera, which is mounted on mobile robot. Motion of camera is analyzed and compensated by comparing edge features among consecutive image frames. Candidate regions of moving object are found by differencing between transformed t^{th} image and $t\text{-}1^{th}$ image. Moving object is finally decided by combined feature set and motion analysis. Object is tracked by matching object components in ROI. We have experimented detecting and tracking moving object with an active camera, which is pan/tilt/zoom and single camera.**

Keywords **— Detecting moving object, tracking object, an active camera.**

1. Introduction

Detecting and tracking techniques of moving object are used in many application fields such as surveillance system, intelligent traffic system, automatic control system, *etc.* Especially, detecting and tacking moving object has been major research topic for interacting robot-human in recent years. Detecting moving object with a camera on mobile robot is not trivial task because two kinds of motions are mixed; one is from camera mounted on robot and the other is from moving object.

Many methods have been proposed to detect motion of moving object using a static camera or an active camera. The former has fixed camera view, which is compared to the latter one. Motion detection using a static camera is not difficult work because the motion of camera is not included. The motion has been detected by differencing among consecutive images. Meanwhile, the motion detection of moving object using an active camera has challenged. To detect motion of moving object, estimation of camera motion or computation of optical flow method has been proposed [1-4]. To detect camera motion, features between adjacent image frames are tracked and coordinate between consecutive images is transformed. However, transformation between images or camera motion has estimated poorly because main motion of camera on mobile robot is different from pan/tilt camera motion [4]. Mobile robot includes pan/tilt/forward/backward camera movement. The motion of moving object without camera motion is not detected completely. Therefore, an algorithm for extracting only moving object motion is required from a complex image, which includes not only a moving object motion but also a background motion caused by a camera motion under an active camera environment.

Once moving object is detected, object has tracked using probabilistic filters such as a particle filter or a Kalman filter [5-6]. But probabilistic filter has complex computation.

In order for mobile robot to detect moving object, motion decomposition is needed that is for separation of camera motion from blended motions occurred from camera and moving object.

This paper proposed detection and tracking moving object using an active camera. Motion of moving object is detected after camera motion is decomposed. The motion of camera from the one of moving object is decomposed by extracting image features between consecutive frames. And camera motion is compensated roughly by differencing consecutive image frames. Candidate regions of moving object are detected from an image which is compensated camera motion.

Moving object is determined using combined feature set and aforementioned motion analysis. Corresponding features such as edge, color and shape are used to detect moving object. Once object is detected, interesting object is tracked in region of interest (ROI). In order to process on real time, interesting region of motion is estimated and tracked. Blob information for the detected moving object, the location, size, shape, and color distribution is obtained.

We have estimated transformation between successive image coordinate systems. However, motion of moving object using an active camera is detected by combined feature set and motion analysis. We have experimented with images captured by an active camera mounted on mobile robot.

2. Configuration of motion detection

Configuration of the motion detection is shown in Fig. 1.

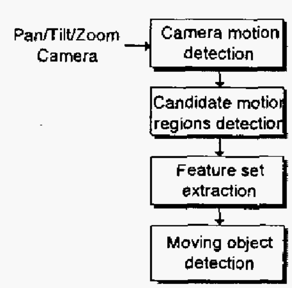


Fig. 1. System configuration for motion detection.

It is consisted of a pan/tilt/zoom camera and four processes. Two kinds of motion are mixed in image when an active camera is used. First of all, camera motion is estimated and decomposed roughly using image features. Camera motion parameters to transform image coordinate are computed in camera motion detection process. Candidate moving object regions are found after camera motion compensation. To detect moving object among candidate moving object regions, skin color and shape information is used.

The motion detection, human motion detection, and tracking algorithm can be described by six steps as follows:

(1) An image acquisition from an active camera on mobile robot.
(2) An extraction of motion information from the image.
(3) A separation between a human motion and a background movement caused by moving camera from some motion areas of image.
(4) A deletion of background movement caused by camera motion from the human motion area
(5) A detection of shape feature information of human which is independent of camera motion by using edge, color and shape information.
(6) A detection of the shape of moving human by using the motion information and the human shape information.
(7) A continuous tracking of the extracted object.

3. Camera motion compensation

When a static camera is used, motion in image is occurred by moving object. Moving object is detected by differencing consecutive image frames *t-1* and *t* that is a fast and simple method to detect motion. However, when an active camera is used, camera motion and moving object motion are mixed in image. Fig. 2 shows blended motions of camera and moving object.

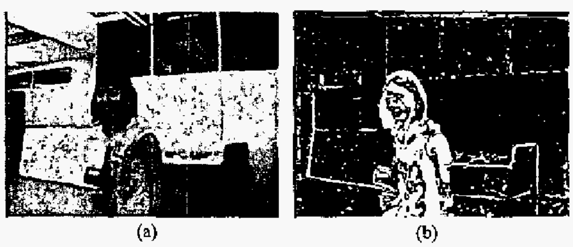


Fig. 2. Example of blended motion of an active camera (a) input image and (b) blended motion image.

To detect motion of moving object, estimation of camera motion is needed. We determine whether camera moves or not. If difference of pixel intensity, $f(x,y)$, among consecutive images, $t-1$, t, $t+1$, is above threshold, the pixel is considered as a motion pixel and is obtained using following *eq*. (1).

$$|f_t(x,y) - f_{t-1}(x,y)| > T_1 \;\&\; |f_{t+1}(x,y) - f_t(x,y)| > T_1 \qquad (1)$$

where f_{t-1}, f_t, f_{t+1} are $t-1$, t and $t+1$ image frames, respectively. T_1 denotes threshold and is calculated empirically.

Camera motion pixels are determined by analyzing blob analysis and following *eq*. (2).

$$\max_x [M_t^c(x,y)] - \min_x [M_t^c(x,y)] > T_2 \qquad c = 1, 2, 3, \cdots n \qquad (2)$$

$$T_2 = 0.8 \times W$$

where M_t^c is connected component for motion pixel in t^{th} image. W and T_2 denote width of image and threshold, respectively.

Meanwhile, if intensity value of motion blob is below threshold, it is considered as motion blob of moving object without camera motion. Small blobs are removed by comparing the number of pixels of connected components.

To decompose camera motion, motion parameters of camera are detected. The initial motion of camera is occurred by pan or tilt movement of camera. Therefore, motion parameter of camera is obtained by pan and tilt movement of camera. Edge features between consecutive two images are detected. The motion parameter for pan movement of camera is calculated using *eq*. (3).

$$h_t(x) = \sum_{y=0}^{h} \sum_{k=-p}^{p} \sum_{l=0}^{dh} \sum_{j=p}^{w-p} (E_t(y+l, j) - E_{t-1}(y+l, j+k))$$

$$pan = \arg\min h_t(x) \qquad (3)$$

where $h(x)$ is defined to find camera movement of x coordinate and and $E(y, j)$ is edge function. In eg. (3), difference between edge components of t^{th} image and $t\text{-}1^{th}$ image is obtained. Fig. 3 shows $h(x)$ of t^{th} image.

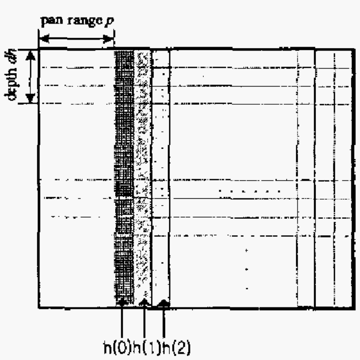


Fig. 3. Example of h(x) of t^{th} image.

Meanwhile, the motion parameter for tilt movement of camera is calculated using *eq*. (4).

$$v_t(y) = \sum_{x=0}^{w} \sum_{k=-t}^{t} \sum_{l=0}^{dv} \sum_{i=t}^{h-t} (E_t(i, x+l) - E_{t-1}(i+k, x+l))$$

$$tilt = \arg\min v_t(y) \tag{4}$$

where $v(x)$ is defined to find camera movement of y coordinate and is obtained by similar method like *eq*. (3).

Two motion parameters detected using the above formulas are used to transform t^{th} image coordinate. The pixel (x', y') of t^{th} image is transformed as $x' = x + pan$ and $y' = y + tilt$.

A motion image is extracted by using an estimated difference value among three cascaded frames which are eliminated pan/tilt camera motion.

4. Motion detection of moving object

Candidate motion regions can be extracted by using the motion image, edge, color and shape. The sub regions of moving object, especially face and hands of human, are extracted by skin color information. And also human motion is obtained by color clustering algorithm. Face and hand regions are used to recognize gesture for controlling mobile robot. Image feature set is consisted of edge, color and shape and is used to extract human motion using following *eq*. (5) to (7).

$$E_t = \sqrt{(E_t^v)^2 + (E_t^h)^2} \tag{5}$$

$$\begin{aligned} E_t^v = & -f_t(x-1, y-1) + f_t(x+1, y-1) - 2 \times f_t(x-1, y) \\ & + 2 \times f_t(x+1, y) - f_t(x-1, y+1) + f_t(x+1, y+1) \end{aligned}$$

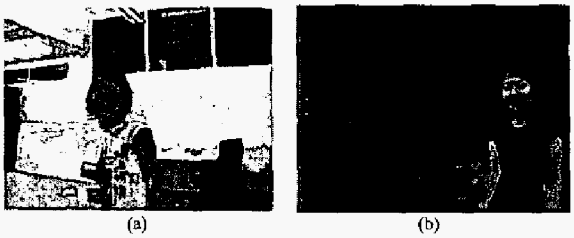
$$\begin{aligned} E_t^h = & -f_t(x-1, y-1) - 2 \times f_t(x, y-1) - f_t(x+1, y-1) \\ & + f_t(x-1, y+1) + 2 \times f_t(x, y+1) + f_t(x+1, y+1) \end{aligned}$$

where E_t is edge component of t^{th} image is obtained by horizontal and vertical edge components.

. The conditions for satisfying skin color region are showed in *eq*. (6).

$$\begin{aligned} & [C_t^b(x,y) > \alpha \times C_t^r(x,y) \ \& \ C_t^b(x,y) > \beta \times C_t^r(x,y) \ \& \\ & C_t^g(x,y) > \gamma \times C_t^r(x,y) \ \& \ C_t^g(x,y) > \delta \times C_t^r(x,y)] \ \& \\ & [\varepsilon < H[x + w \times y] < \eta] \ \& \ [\omega < S[x + w \times y] < \xi] \end{aligned} \tag{6}$$

.where C_t^r, C_t^r and C_t^r are r, g, b values of t^{th} image, respectively. $H(\cdot)$ and $S(\cdot)$ denote hue and saturation values obtained from C_t^r, C_t^r and C_t^r. If $f(x,y)$ satisfies above condition, it is considered as skin color pixel.



(a) (b)

Fig. 4. Example of (a) color clustered image and (b) skin color regions.

Fig. 4. shows image applied color clustering algorithm and regions applied skin color detection algorithm.

Moving object region, human motion region, is detected by motion region detected using above camera motion compensation. And also human motion region is detected by combined image features. The detected human motion region is appeared in Fig. 5.

Fig. 5. Result image detected by motion and image features.

5. Tracking moving object

Moving human motion is detected in t^{th} image and blobs of the human motion are analyzed. The shape of human is tracked using block matching algorithm only for detected human shape region in region of interest. Human motion is detected by above motion detection process. Blob information for the detected human motion, the location, size, shape, and color distribution for human motion blobs is obtained. In order to process on real time, interesting region of motion is estimated. Blobs to be tracked are shown in Fig. 6.

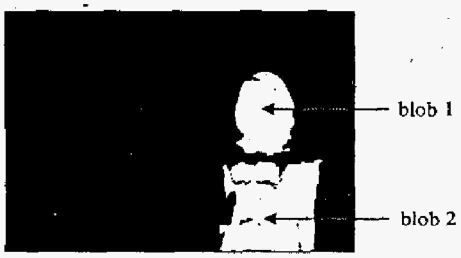


Fig. 6. Example of blobs to be tracked.

It is possible to tack in real time because human motion to be tracked is detected by combined factor of motion analysis and image features not probabilistic model. And also proposed tracking method provides good performance when object was occluded or disappeared in real world environment.

6. Experiments

The proposed algorithm was implemented and tested on mobile robot under indoor environment. Mobile robot for experiment is shown in Fig. 7. An input image of 320x240 pixels was acquisitioned from an active camera on mobile robot. Tracking was able to process ten frames per second.

The performance for detecting and tracking moving object was evaluated by tracking target object marked manually. The result images obtained in motion detection process are shown in Fig. 8.

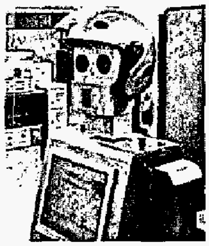


Fig. 7. Mobile robot used for experiment.

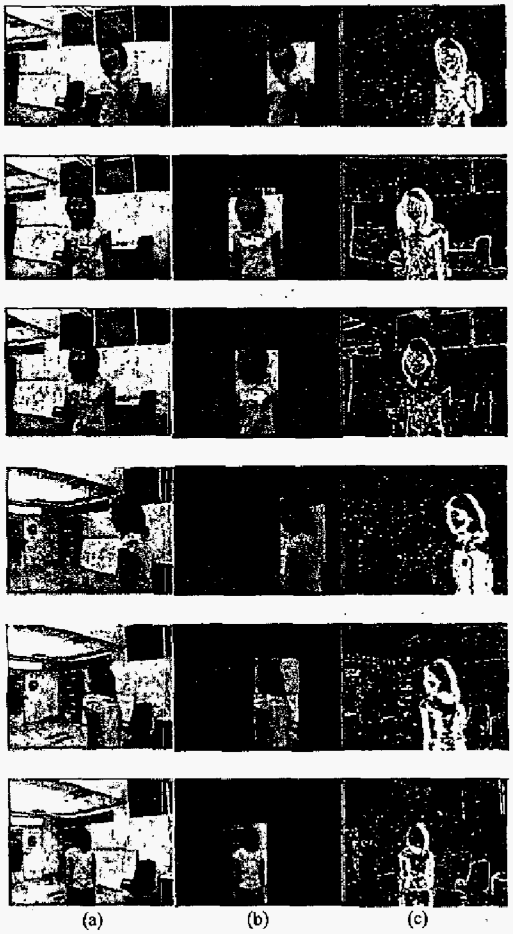


Fig. 8. Example of images obtained in motion detection process (a) input images (b) detected object, and (c) motion image without camera motion compensation.

We have tracked a human motion by evaluating whether the estimated human shape region is an exact tracking target or not.

7. Conclusions

We have proposed motion detection and tracking method using an active camera on mobile robot. In order for mobile robot to detect moving object, camera motion decomposition was used that was process for separation of camera motion from moving object motion. Camera motion from the motion of moving object is separated by extracting image features and by differencing consecutive image frames.

Candidate regions of moving object were detected using a rough motion compensation of camera. Moving object was detected by combined feature set and motion analysis. Corresponding features such as edge, color and shape was used to detect moving object.

Once moving object was detected, object was tracked in ROI. In order to process on real time, interesting region of motion was estimated and tracked. The proposed method had estimated transformation between consecutive image coordinate systems. However, motion of moving object using an active camera was detected by combined feature set and motion analysis. It provided good performance for motion detection and object tracking even though object was occluded or disappeared.

We have experimented with images captured by an active camera mounted on mobile robot. Main goal of proposed algorithm is robustness on mobile robot in real environment.

References

[1] D. Murray and A. Basu, "Motion tracking with an active camera", IEEE Trans. On Pattern Analysis and Machine Intelligence, 16(5), pp. 449-459, May, 1994.

[2] M. Irani, R. Rousso, and S. Peleg, "Recovery of ego-motion using image stabilization", In Proc. Of the IEEE Computer Vision and Pattern Recog., pp. 454-460, March, 1994.

[3] A. Censi, A. Fusiello, and V. Roberto, "Image stabilization by features tracking", In Proc. of the 10th Int. Conf. on Image Analysis and Processing, pp. 665-667, Venice, Italy, Sep., 1999.

[4] B. Jung and G. S. Sukhatme, "Detecting moving objects using a single camera on a mobile robot in an outdoor environment", In the 8th Conf. on Intelligent Autonomous Systems, pp. 980-987, Amsterdam, The Netherlands, March 10-13, 2004.

[5] C. Hue, J. P. L. Cadre, and P. Perez, "A partical filter to track multiple objects", In IEEE Workshop on Multi-object Tracking, pp. 61-68, Vancouver, Canada, July, 2001.

[6] J. Kang, I. Cohen, and G Medioni, "Continuous multi-views tracking using tensor voting", In Proc. Of the IEEE Workshop on Motion and Video Computing, pp. 181-186, Orlando, Florida, Dec., 2002.