7th WSEAS Int. Conf. on MATHEMATICAL METHODS and COMPUTATIONAL TECHNIQUES IN ELECTRICAL ENGINEERING, Sofia, 27-29/10/05 (pp255-258)

# Motion Analysis for Human-Robot Interaction

Kye Kyung Kim, Hae Jin Kim and Jae Yeon Lee
Electronics Telecommunications Research Institute, Korea

*Abstract* — **This paper is to present vision based motion analysis for human-robot interaction, which analyzes camera motion and human motion. Motion of camera is compensated by comparing edge features among consecutive image frames. Candidate regions of human motion are found by differencing between transformed $t^{th}$ image and $t\text{-}1^{th}$ image. Human motion is finally decided by image features and motion analysis. Gesture recognition module detects moving hand by motion analysis and skin color information obtained from face detection. The variation of hand location and the meaning gesture region are detected. We have experimented detection of moving object and gesture recognition with an active camera, which is pan/tilt/zoom and single camera that is mounted on mobile robot. Performance evaluation of gesture recognition has experimented using ETRI database and an encouraging recognition rate of 84% has been obtained.**

*Keywords* — **Detection of moving object, hand detection, gesture recognition.**

## 1. Introduction

Vision based motion analysis has many application fields such as surveillance system, intelligent traffic system, automatic control system, e*tc*. Especially, detecting moving object and recognizing gesture have been major research topic for interacting human-robot in recent years [1-11].

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 been challenged. To detect motion of moving object, estimation of camera motion or computation of optical flow method has been proposed [1-6].

Meanwhile, vision based gesture recognition [7-11] has been studied intensively. Motion analysis is necessary to recognize gesture that includes hand or arm motion detection. It has been developed to provide intelligent and natural communication between human and robot. Users generally use arm and hand gestures to give aid expression of their feelings and notification of their thought. Users usually use more simple gestures such as pointing gestures or command gestures rather than complex gestures. Especially, pointing gestures in mobile robot environment give directional information such as where to move a robot. Meanwhile command gestures give specific behavior such as stopping movement or identifying caller identification. Typical approaches using HMMs or neural networks have applied analytical methods and recognizing patterns. However, it is very difficult to recognize gestures because of diversity of gesture data from many operators and hard separation of meaning gesture from consecutive image frames. Gesture recognition is necessary for enhancement of communication and for good communication in noise environment.

This paper proposes vision based motion analysis that includes detection of moving object and gesture recognition 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 a compensated camera motion image. Moving object is determined using combined image features and aforementioned motion analysis. Corresponding features such as edge, color and shape are used to detect moving object. And also image features are used to recognize gestures. The image features such as skin color, shape and motion information are used to detect hand location and meaning gesture region. We have experimented with images captured by an active camera mounted on mobile robot and have tested performance of proposed motion analysis.

## 2. Configuration of motion Analysis

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

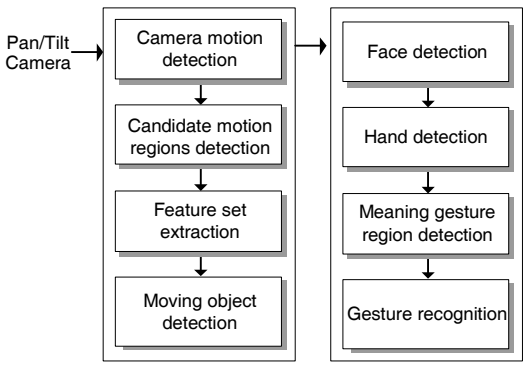


**Fig. 1. System configuration for motion analysis for HRI**

It is consisted of a pan/tilt camera, motion detection and gesture recognition modules. Camera motion and moving object motion is analyzed in motion detection module. Camera motion is estimated and decomposed from moving object motion because two kinds of motions are mixed in captured image when an active camera is used. 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.

Gesture recognition module includes hand detection, meaning gesture region detection and gesture recognition. To detect hand location, skin color information is used that is applied to detect face. Image features and motion information is used to detect hand location. Hand location is traced and transition value of hand coordinate from consecutive frames is calculated to recognize gesture. Pointing and command gestures are used to control mobile robot.

## 3. Motion detection module

Mobile robot includes pan/tilt/forward/backward/zoom camera movement. Camera motion and moving object motion are mixed in captured image because of movement of a mobile robot. 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. 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. The motion detection module processes 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.

### 3.1 Camera motion compensation

Camera motion is estimated and decomposed from moving object motion using image features. Camera motion parameters are computed to transform image coordinate, which is transformed by mobile robot movement. Fig. 2 shows blended motions of camera and moving object. To detect camera motion, features between adjacent image frames are extracted 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].

(a) (b)

**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).

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

$$T_2 = 0.8 \text{ x } W$$

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.

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.

## 3.2 Motion detection of moving object

Moving object region, human motion region, is decided from candidate motion regions, which is detected after camera motion is compensated. We have extracted image features to select human motion region because motion analysis is not sufficient to extract human motion. Motion detection is not trivial task in complex background and a mobile robot environment.

Temporal and spatial edge detection method has tried to detect human in both environment such as a static and an active camera environment. However, human detection has been challenged. Combined image features such as edge, color and shape have extracted. Motion region by image differencing and combined image features are used to detect human motion. The color features used to detect human motion region are appered in Fig. 3.

(a) (b)

**Fig. 3. Example of (a) color clustered image and (b) skin color regions.**

Fig. 4 shows human motion region, which is selected combined image features and motion analysis.

**Fig. 4. Result image detected by motion analysis and image features.**

# 4. Gesture recognition

The location of hand is detected by motion analysis by eq. (1) and skin color information. Hand color is estimated from face detection using open CV because of illumination effect. Extracted face region and hand location are shown in Fig. 5.

**Fig. 5. Example of extarction of face and hand.**

The transition values of consecutive hand coordidates are calculated to select meaning geature frame. If a transition value is above threshold, it considered as the start of gesture frame. The location of hand is traced and the transition values of consecutive hand coordidates are calculated for detecting the end of gesture frame.

The relative coordinates between face and hand are used as feature to recognize gesture. And also the coordinate of hand is used as feature vestor. We have implemented MLP to classify gesture. In this paper, a three-layer perceptron is implemented and trained by a well-known modified backpropagation algorithm which uses the instantaneous squared error [12].

# 5. Experiments

The proposed algorithm was implemented and tested on mobile robot under indoor environment. Mobile robot for experiment is shown in Fig. 6. An input image of 320x240 pixels was acquisitioned from an active camera on mobile robot. Tracking was able to process ten frames per second.

**Fig. 6. Mobile robot used for experiment.**

The performance of detection of moving object and gesture recognition are evaluated. The result images obtained in motion detection process are shown in Fig. 7. We have tracked a human motion by evaluating whether the estimated human shape region is an exact tracking target or not.

The proposed motion analysis is an efficient for interacting human-robot. However, we still have been challenged because of camera motion compensation and human form detection.

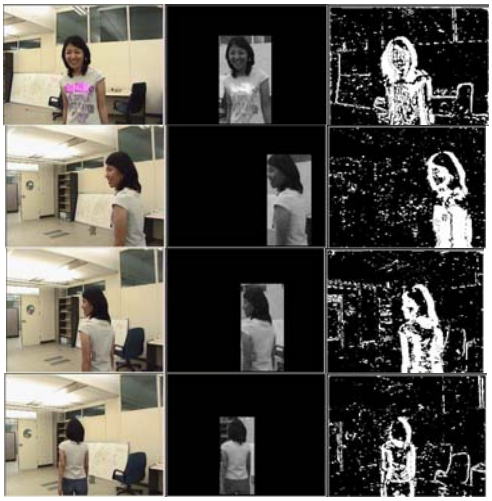


**Fig. 7. Example of images obtained in motion detection process (a) input images (b) detected object, and (c) motion image without camera motion compensation.**

Fig. 8 shows the type of gestures. Pointing gestures, right and left and command gestures, bow, handshake, circle and cross, are trained and tested. Each gesture is separated from continuous gestures in image frames. To evaluate gesture recognition method, we have carried out experiments with the gesture data of the ETRI database. A multi-layer perceptron has been trained with *x* and *y* hand coordinates from 9 persons and tested with 5 persons. The recognition rates according to six gesture types are shown in Table 1. Table 2 shows confusion matrix for gesture recognition results. The gestures of "bow" and "cross" have good recognition performances. But, "handshake", "right" and "left" have poorer results, because of diverse gesture data according to persons.

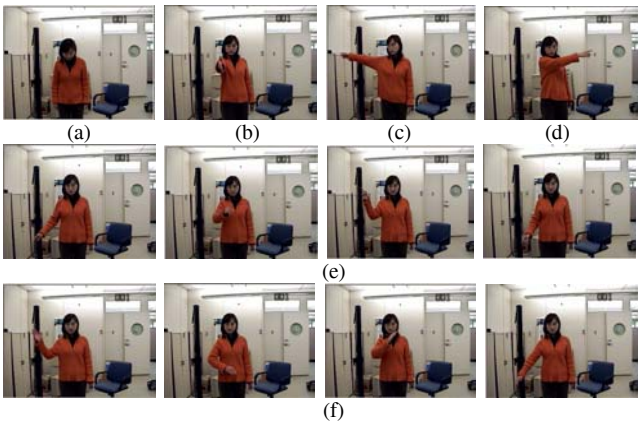


**Fig. 8. Six kinds of gesture types (a) bow (b) handshake (c) right (d) left (e) circle and (f) cross.**

**Table 1. The result of gesture recognition.**

| Target gesture | Recognition result of training DB | Recognition result of testing DB |
|---|---|---|
| bow | 100% | 100% |
| handshake | 100% | 86% |
| right | 100% | 71% |
| left | 100% | 67% |
| circle | 100% | 80% |
| cross | 100% | 100% |
| Total | 100% | 84% |

**Table 2. The confusion matrix for recognition result of Table 1.**

| Target gesture \ Recognized gesture | C[0] | C[1] | C[2] | C[3] | C[4] | C[5] | The number of target gestures | The number of recognized gestures |
|---|---|---|---|---|---|---|---|---|
| bow : C[0] | 8 | - | - | - | - | - | 8 | 8 |
| handshake : C[1] | - | 6 | 1 | - | - | - | 7 | 6 |
| right : C[2] | - | 1 | 5 | - | 1 | - | 7 | 5 |
| left : C[3] | - | 2 | - | 4 | - | - | 6 | 4 |
| circle : C[4] | - | - | - | - | 4 | 1 | 5 | 4 |
| cross : C[5] | - | - | - | - | - | 5 | 5 | 5 |
| Recognition rate (%) | 100 | 86 | 71 | 67 | 80 | 100 | 38 | 32 |

## 6. Conclusions

We have proposed vision based motion detection and gesture recognition method using a camera on mobile robot. To detect moving object, camera motion decomposition was used to separate camera motion from moving object motion. Candidate regions of moving object were detected using a rough motion compensation of camera. Moving object was detected by combined image feature set and motion analysis. Corresponding features such as edge, color and shape was used to detect moving object. Image features such as skin color, shape and motion information are used to detect hand location and meaning gesture region in gesture module.

We have experimented with images captured by an active camera mounted on mobile robot. In the future we are trying to detect motion with more combined image features and to recognize combined recognizer. 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.

[7] A. Ali and J. K. Aggarwal, "Segmentation and Recognition of Continuous Human Activity," IEEE Detection and Recognition of Events in Video, 2001.

[8] A. Bobick, " Real Time Online Adaptive Gesture Recognition, " IEEE ICPR, 2000.

[9] B. Li and H. Holstein, "Recognition of Human Periodic Motion- a Frequency Domain Approach, " IEEE ICPR, 2002.

[10] D. Ayers and M. Shah, "Recognizing Human Actions in a Static Room," IEEE WACV, 1998.

[11] A. Corradini, "Intergrated Dynamic Time Warping for Off-line Recognition of a Small Gesture Vocabulary," RATFG-RTS, 2001.

[12] B. Kosko, "Neural networks and fuzzy system," Prentice-Hall, Englewood Cliffs, NJ, 1992.