TUDOR-IOAN CERLINCA, STEFAN-GHEORGHE PENTIUC

A NOVEL ARCHITECTURE FOR HAND GESTURE–BASED CONTROL OF MOBILE ROBOTS

Stefan cel Mare University of Suceava,
Universitatii str., 13, Suceava, Romania,
tel.: (+40)-230-524801, E-mail: tudor_c@eed.usv.ro

Abstract. The aim of this paper is to present a novel architecture for hand gesture-based control of mobile robots. The research activity was mainly focused on the design of a new method for hand gestures recognition under unconstrained scenes. Our method comes to solve some of the most important problems that current HRI (Human-Robot Interaction) systems fight with: changes of lighting, long distances, speed. Like any other HRI specific method, the one that we developed is working in real time/environment. It is a robust and adaptive method, being able to deal with changes of lighting. It is also capable of recognizing hand gestures from long distances. Another important issue we have focused upon was the integration of our method into a more complex HRI system, in which a human operator can drive a mobile robot only through hand gestures. In systems like these, the communication between human operators and robotic systems should be done in the most natural way. Typically, communication is done through voice and hands/head postures and gestures. Our method was designed in such a manner that will be able to recognize the hand gestures even if there are certain deviations from the ideal cases.
Key words: HRI (Human-Robot Interaction) systems, robust and accurate method.

INTRODUCTION

Human-Robot Interaction (HRI) can be considered as one of the most important Computer Vision domains. It has many applications in a variety of fields such as: search and rescue, military battle, mine and bomb detection, scientific exploration, law enforcement, entertainment and hospital care. HRI is the study of interactions between people and robots. In HRI based systems, the communication between human operators and robotic systems should be done in the most natural way. Typically, communication is done through voice and hands/head postures and gestures. The most important aspect related to HRI applications is the capability of running into unconstrained environments. There are also some other aspects that should be taken into consideration when developing a posture/gesture based HRI application:

- the way of increasing the maximum distance from which the system is still able to detect hands/head postures/gestures.
- the way of setting up the system. It is desirable that the set up procedure to be done only once. Then, the environment can be changed without affecting the applications results.

The method we developed proves to be robust to changes of lighting and can detect the hands gestures from long distances (up to 5 m). Also, it requires only a simple one-time set-up. Our method of hand gestures recognition was tested in the framework of an experimental robot control system.

HAND GESTURES RECOGNITION

Our method can detect up to 6 different hands gestures that corresponds to 4 different commands: LEFT, RIGHT, FORWARD, BACKWARD. Generally there are two different approaches to the problem of gestures recognition: fingers gestures recognition and hands gestures recognition. Because fingers cannot be accurately detected from long distances in unconstrained environments, we chose to detect/recognize the hands gestures.

The main idea is to estimate the hands location based on the head's location and size and then to verify if the location corresponds to a valid gesture. The hand gestures recognition algorithm is an 8 stages process: *Image acquisition, Head detection, Hands position estimation, Skin detection, Hands components detection, Hands components connectivity, Motion analysis for hands components and Gesture selection.*

HEAD DETECTION

Neither less to say, this is one of the most important stages. The main goal is to detect/estimate the location and size of the human head. If the head cannot be detected then no posture will be recognized. The detection algorithm works only for images that contain profile or frontal views of the face. A more detailed description of this algorithm could be found in [2] and [8]. The experimental results showed that sometimes the detection fails even if the images contain profile or frontal views of the face. In cases like these, our method will try to estimate the head's location on the base of the previous locations. The estimation will be done only for those images that fulfill the following condition: $t_{curr\_img} - t_{head} < Th_{MaxDiffTime}$, where: t_{curr_img} - date/time of the current image, t_{head} - date/time of the last successful detection and $Th_{MaxDiffTime} \in [0_{ms}, 200_{ms}]$ - the maximum elapsed time for the estimation to take place. These are the images with a high probability of head presence. Typically, it can be considered that the human operator will not move his head in $Th_{MaxDiffTime}$. A smaller $t_{curr\_img} - t_{head}$ value means a higher probability of head presence. Figures 1 and 2 show the results of this stage.

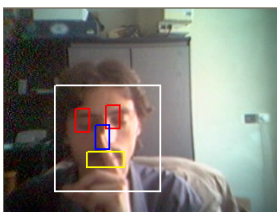

Figure 1. Head detection. One person

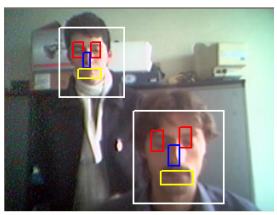

Figure 2. Head detection. Two persons

HANDS POSITION ESTIMATION

We consider a gesture to be valid only when the hands can be located in certain rectangular areas. For each hand, our method will determine 3 different rectangles. Figure 3 shows all these rectangles and the corresponding gestures. These rectangles are not fixed, but determined according to the actual head's location

and size. The estimation procedure is based on the Leonardo da Vinci's studies about human body proportions [3]. According to his studies, the size and position of neck, lower/upper arms and hands can be expressed as a function of head's size and location as follows:

- the neck space is 1/4 of a head length
- the hand is 3/4 of a head length
- the lower-arm is 5/4 heads length
- the upper-arm is 3/2 heads length

So, if the head can be detected, then, for each valid gesture, we can estimate the location and size of the hands. Because it is hard to determine the exact location and size of the hands, we chose to make these rectangles bigger than the actual estimated size of the hands. Note that this stage purpose is not to find the human hands but only to determine the rectangular areas in which to search for them.

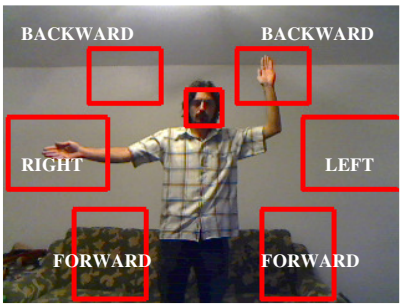

Figure 3. Hand position estimation

SKIN DETECTION

The main goal of this stage is to search for the human hands, in all 6 rectangular areas previously determined. This operation is done through the help of an adaptive skin detection algorithm that will determine the amount of skin-pixels for each rectangular area. It is considered that a hand is located in one of the 3 rectangular areas, only if there are sufficient skin-pixels. The threshold value is not fixed but determined according to the actual estimated size of the hands.

The skin detection algorithm operates in the HSV (Hue-Saturation-Value) [4] color space because of its similarities to the way that humans tend to perceive colors. Typically, a pixel can be considered a skin-pixel only if $H > h_{\min}$, $H < h_{\max}$ and $S > s_{\min}$, where *H*, *S*, *V* are the pixel's color components in the HSV color space and $h_{\min}$, $h_{\max}$, $s_{\min}$ the threshold values. The transformation between RGB and HSV color spaces is detailed described in [4]. In order to deal with changes of lighting or other environmental changes, the threshold values should be continuously updated. The only problem is to find something that will always reflect the environmental changes. We have done a lot of tests and came to the conclusion that the HSV histograms of face and hands looks pretty similar. So, if we can determine $h_{\min}$, $h_{\max}$ and $s_{\min}$ for skin-pixels in the face region, then we can use the same values to detect the hands skin-pixels. These values can be easily determined only by analyzing the HSV histogram of the head region.

Due to its high speed, our method of hands detection can be successfully used in those HRI applications that require real-time detection and recognition of hand gestures.

Mainly, this stage performs only a quantitative analysis of images. The overall results are good but the method fails when the rectangles contain skin-colored objects with areas comparable to the hands ones (bricks, cartoon boxes etc).

We have improved our method by performing a qualitative analysis also. It is a 3 steps analysis:

- hands components detection
- hands components connectivity
- hands components motion analysis

HANDS COMPONENTS DETECTION. MORPHOLOGICAL OPERATIONS

The main goal of this stage is to detect the skin-colored objects in all 6 rectangular areas. The detection was done through the help of a fast flood-fill algorithm that operates on the binarized image. The set of objects was filtered to remove the ones that cannot be considered a part of the human hands. The new set of objects $\boldsymbol{M}$ will be composed only from those that fulfill the following condition: $ThMinArea < Area(Obj) < EstimatedHandsArea$ where $ThMinArea$ is automatically computed on the base of the estimated hands size. Object detection is followed by an image closing operation, which fills the small gaps and discontinuities.

HANDS COMPONENTS CONNECTIVITY

This is the most important stage in the process of qualitative analysis of images. The main purpose is to connect together different objects that belongs to the human hands. The connection will take place only when the total number of skin-colored objects is below the limit *ThMaxNumObjects*. Otherwise, we can consider that a false detection occurred. It is considered that two objects O_1 and O_2 can be connected together if and only if the distance from theirs weight centers is below *ThMinWGCDist*. This threshold value is not fixed but determined according to the actual estimated size of the hands. The result of the connection stage is a new set $M_1 \subseteq M$ of objects that can be connected together. If the total area of objects in M_1 is greater than the actual estimated size of the hands (*EstimatedHandsArea*), then the objects with the smallest area will be consequently removed from M_1, till total area is approximately equal to *EstimatedHandsArea*. The new set M_2 fulfills the following conditions:

- $\forall O \in M_2, ThMinArea \le Area(O) \le EstimatedHandsArea$
- $\forall O_i, O_j \in M_2, \mathrm{Dist}(\mathrm{WGC}(\mathrm{O_i}), \mathrm{WGC}(\mathrm{O_j})) \le ThMinWGCDist$
- $Count(M_2) \le ThMaxNumObjects \quad \sum_{i=1}^{Count(M_2)} Area(O_i) \cong EstimatedHandsArea$

MOTION ANALYSIS FOR HANDS COMPONENTS

The performances have been increased but still, the system can give incorrect results. The problem with stationary skin-colored objects was solved by building and analyzing the motion history image (MHI) for all 6 rectangular areas. A gesture can be considered to be valid only when the skin-colored objects are moving and the motion gradient is uniform. Motion analysis is a 6 stages process:

1. motion detection for skin-colored objects
2. contour extraction for moving skin-colored objects
3. identification of movement direction
4. update of motion history image (MHI)
5. quantitative analysis of MHI
6. qualitative analysis of MHI

Motion detection was done through an adaptive background subtraction algorithm [5] [8]. The result of such an algorithm is represented in our case by a set of skin-colored moving objects. Then, for each moving object the contour is extracted and added to the motion vector. This operation takes approximately 1 second and a half. In the next step, our method takes the last two images and tries to detect the movement direction by comparing the contours pixels for all pairs of skin-colored moving objects. This way, at each step, we obtain a new moving direction.

Figures 4 and 5 present a moving object and the corresponding motion history image [5]. Figure 6 presents the motion gradient image (movement direction) for the MHI presented in figure 5 [7]. In the quantitative analysis stage, our method will only perform a check on the current amount of movements and determine if it can be associated with a moving hand or not. In the qualitative analysis stage, the method will determine the uniformity of the motion gradient. A non-uniform gradient can be associated with a chaotic movement of some skin-colored objects.

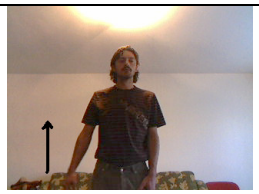

Figure 4. Moving object

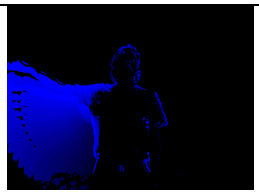

Figure 5. MHI of moving object

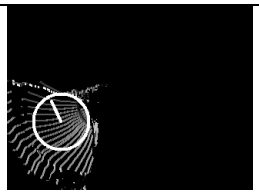

Figure 6. Motion gradient image

EXPERIMENTAL RESULTS

Figures 7-14 present the results of our hand gestures recognition method. The performances of our method were tested in 3 different conditions:

1. with sample images grabbed from static scenes in which the lighting and the distances from the video cameras were approximately constant. The test set was built using a total of 1800 different images (300 images for each gesture).
2. in unconstrained scenes with changes of lighting and different distances from the video cameras.
3. in the framework of an experimental robot control system. When a valid gesture is detected, a specific command will be sent to the mobile robot through the RS232 serial port. The command will be received by a microcontroller that will control the robot's engines.

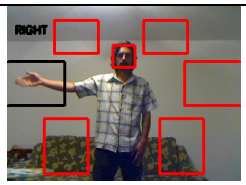

Figure 7. RIGHT command

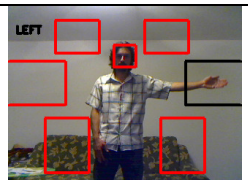

Figure 8. LEFT command

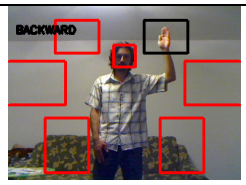

Figure 9. BACKWARD command

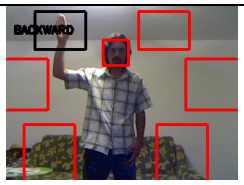

Figure 10. BACKWARD command

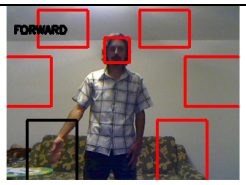

Figure 11. FORWARD command

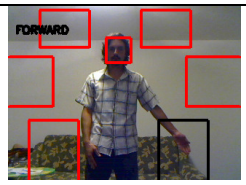

Figure 12. FORWARD command

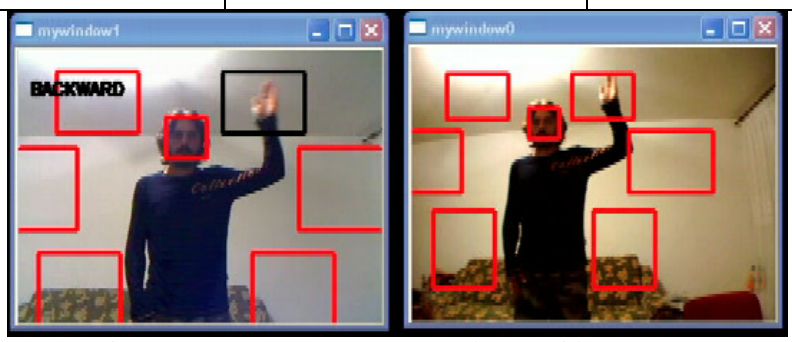

Figure 13. BACKWARD command. Two video cameras

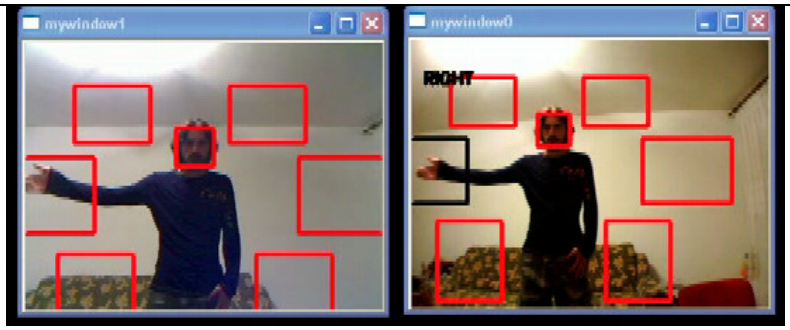

Figure 14. RIGHT command. Two video cameras

Tables 1, 2 and 3 show the detection rates for each case.

Table1.

Detection rates - static environment

	Positive	Negative
True	95%	98%
False	2%	5%

Table2.

Detection rates - unconstrained environment

	Positive	Negative
True	80-90%	95%
False	5%	10-20%

Table3.

Detection rates - mobile robot framework

	Positive	Negative
True	60-75%	90%
False	5%	25-40%

CONCLUSIONS AND FUTURE WORK

This paper presents a fast, robust and accurate method for hand gestures recognition under unconstrained scenes. Due to its advantages, our method can be extended (to recognize some other gestures) and used in various Computer Vision applications.

Our final goal was to integrate the method we have developed into a more complex HRI system, in which a human operator can drive a mobile robot only through hand gestures. The software application runs on Notebooks/Embedded Systems with Windows 2000/XP/Vista and communicates with the mobile robot through the RS232 serial port. The presented work is supported by CEEX-131/2006 funding grant.

Future work will be focused on recognition of more complicated gestures.

REFERENCES

1. http://en.wikipedia.org/wiki/Human_robot_interaction.
2. Tudor-Ioan Cerlinca, Stefan-Gheorghe Pentiuc, Marius Cerlinca, Radu Vatavu (2007), “Hand Posture Recognition for Human-Robot Interaction”, Workshop on Multimodal Interfaces for Semantic Interaction, Vol: 1, 1-2 Noiembrie, 2007, Japan.
3. http://en.wikipedia.org/wiki/Vitruvian_Man
4. http://en.wikipedia.org/wiki/HSV_color_space
5. Cerlinca Tudor-Ioan, "A Distributed System for Real Time Traffic Monitoring and Analysis", Advances

in Electrical and Computer Enginering, University Stefan cel Mare of Suceava, Romania ISSN:1582-7445, Vol 5(12), pp. 82-86, 2005.
6. http://www.intel.com/technology/computing/opencv/overview.htm
7. Pentiuc, Stefan-Gheorghe; Vatavu, Radu Daniel; Ungurean, Ciprian-Ovidiu; Cerlinca, Tudor Ioan (2008), "Techniques for Interacting by Gestures with Information Systems", Ecumict, Belgia 2008.
8. Mihai Horia ZAHARIA; An Alternative Approach to Grid Computing NorduGrid, Advances in Electrical and Computer Engineering, Suceava, Romania,ISSN 1582-7445, No 2/2004, volume 4 (11), pp. 60-63.

Надійшла до редакції 05.01.2009р.

TUDOR-IOAN CERLINCA - lecturer in Electrical Engineering and Computer Science at the Stefan cel Mare University of Suceava, Suceava, Romania, Phone: +40-727-713078,
E-mail: *tudor_c@eed.usv.ro*.