# DEVELOPMENT OF AUDIO-VISUAL MODEL AND ALGORITHM FOR HUMAN CONTROL OF MOBILE ROBOT MOTION WITH COMBINED RECOGNITION OF VOICE AND GESTURE COMMANDS

Snezhana Pleshkova, Zahari Zahariev, Aleksander Bekiarski, Shima Sehati Dehkharghani

Technical University of Sofia, Bulgaria, St. Kliment Ohridski 8

Tel.: +359 965 3300; E-mail: snegpl@tu-sofia.bg

*Abstract*

*The existence of audio and visual sensors in the intelligent robots, is the real challenge to develop audiovisual models and algorithms for human control of mobile robot motion with recognition of voice and gesture commands. There are a lot of developments in this area of researches. Usually they consider separately only audio or only visual information from human voice or gesture commands to control mobile robots motion, but the precision is not sufficient. Therefore, the goal of the article is to extend this precision applying together the recognized voice and gesture commands with the proposed model and algorithm for human mobile robots motion control. The results for achieved precision in the mobile robots motion control, using together the recognized human voice and gesture commands, are presented and demonstrated from carried out a lot of experiments.*

## 1. INTRODUCTION

There exists two types of interactions between mobile robot and human. They can be considering separately or in combinations:

- mobile robot follow human movement, voice or gestures [1];
- human commands send to control mobile robot motion [2, 3].

This paper considering the second type of interactions between mobile robot and human. This task can be solved using different methods, but in all of them is necessary to take in account the existing and mounted sensors [4] on the mobile robot platform like tactile, ultrasound, GPS, sound, image, wheel sensors, etc. Only the sound and image sensors are useful in the special cases of human mobile robot interactions and there are a lot of developments in this area of researches [5]. Most of them consider separately only audio or only visual information from human voice [6] or gesture commands [7] to control mobile robots motion, but the precision is not sufficient. In this article the goal is to increase this precision applying and analysing together the recognized audio and visual information in human voice and gesture commands. To do this it is proposed a suitable model and corresponding algorithm for mobile robots motion control. The experiments, to test the proposed model and algorithm, are carried out using the mounted on for mobile robot, the embedded IoT module myRIO [8] and with connected to it sound and image sensors, i.e. microphone and camera. The programmer code for realizing the proposed model and algorithm is developed using visual programming language LabView [9], which is compatible and can be embedded in myRIO module.

The achieved precision in the experiments of human mobile robots motion control is presented using together both the recognized human voice and gesture commands and is compared with the precision achieved in the cases of using separately only voice of only gesture commands for mobile robot motion control.

## 2. THE PROPOSED AUDIO-VISUAL MODEL FOR HUMAN CONTROL OF MOBILE ROBOT MOTION

The general view of the proposed audio-visual model for human control of mobile robot motion with combined recognition of voice and gesture commands is presented on Figure 1. Audio-visual information for voice and gestures commands is received from Microphone and Camera mounted on the Mobile Robot Platform.

The main parts in the proposed audio-visual model are the Voice and Gesture Command Recognition blocks.

The recognized voice and gesture commands are transformed in the next Voice and Gesture Commands Combined Interpretation block as corresponding motor control signals. Then, from these signals, the next Mobile Robot Motor Control block prepare the necessary execution signals from each of the motors on the wheels of the Mobile Robot Moving Platform, shown on Figure 1.

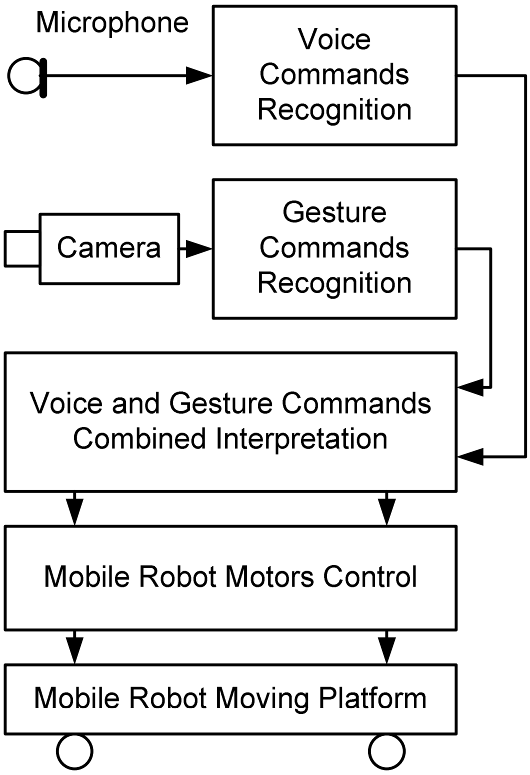


**Figure 1.** General audio-visual model for human control of mobile robot motion

## 3. DEVELOPMENT OF ALGORITHM FOR AUDIO-VISUAL HUMAN CONTROL OF MOBILE ROBOT MOTION

The developed algorithm for audio-visual human control of mobile robot motion is presented as audio (Figure 2), video (Figure 3) parts and its combination (Figure 4). Usually the voice only recognition part of algorithm, shown on Figure 2, is started first, because the voice human interaction with mobile robot is more natural and more common used.

After definition of initial or current mobile robot position is necessary to check the voice existence for activate the voice recognition. When the voice command recognition is correct, mobile robot do movement, according to recognized voice command.

If the recognition of voice command is not correct, mobile robot do not movement, increasing the collected number of voice error by one. In this case the algorithm can go to wait for new voice command or, depending to human decision, go to try the algorithm for recognition of the same not recognized voice command, but as a gesture command.

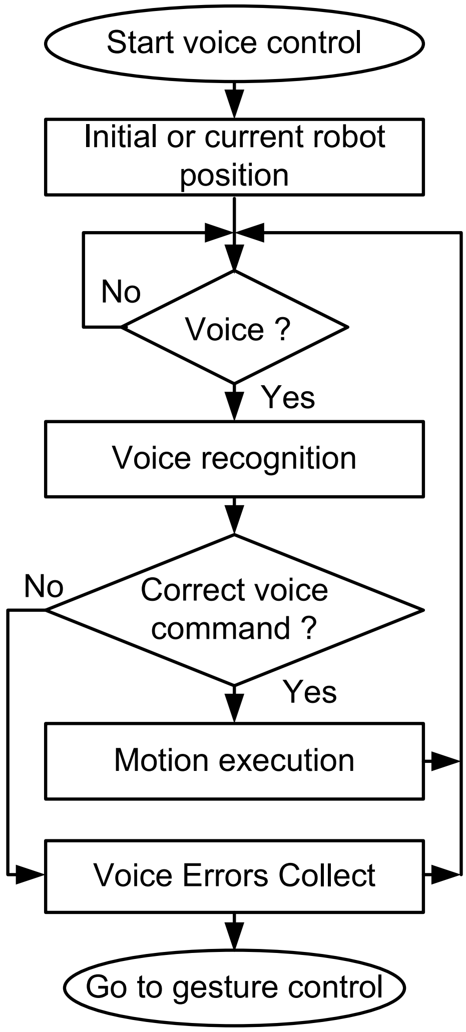


**Figure 2.** Algorithm for voice only human control of mobile robot motion

The algorithm for gesture only commands recognition (Figure 3) is similar, except that in the case, when the recognition of gesture command is not correct, the algorithm can go to wait for new gesture command or, depending to human decision, go to try the algorithm for combined voice and gesture human control of mobile robot motion.

The algorithm for combined voice and gesture human control of mobile robot motion is presented on Figure 4. In this algorithm are carried out at the same time the voice and gesture commands recognition and if some of them is not correct recognized, then it can go to wait for new not recognized, but as the opposite, gesture or voice command.

Also, if the voice and gesture commands are correctly recognized, they are checked for matching. If the matching fail, then the collected number of combined voice and gesture error is increased by one, mobile robot do not movement and algorithm go to wait for other voice or/and gesture commands. In the case, when the matching of the voice and gesture commands is successful, then mobile robot do movement, according to the recognized voice and gesture commands or the algorithm is ending, if there is a recognized voice and gesture stop commands, or if it is the time to stop motion of mobile robot, depending to human decision.

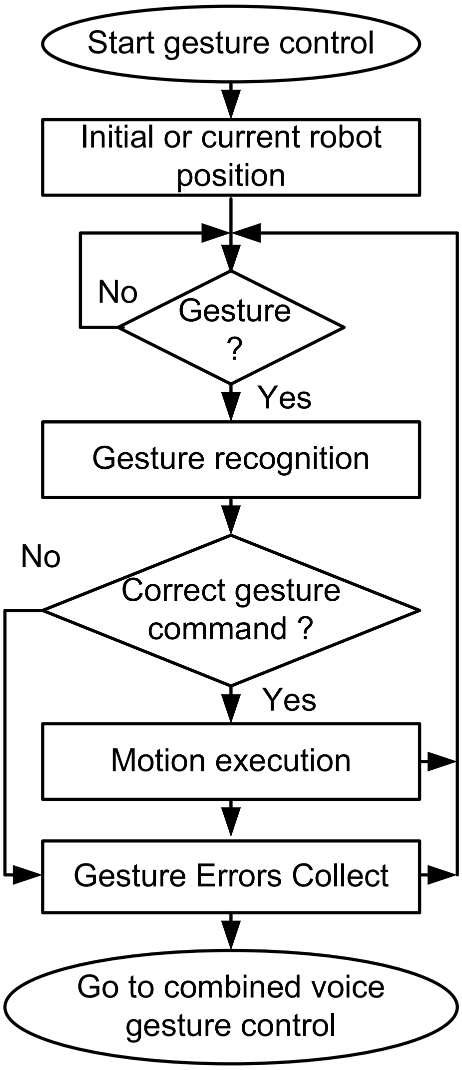


**Figure 3.** Algorithm for gesture only human control of mobile robot motion

## 4. EXPERIMENTAL RESULTS

In the experiments are defined different, but with the constant number $N_S$ of moving steps (for example $N_S=60$), tasks for mobile robot to follow the predefined trajectories of motion. The number of collected and averaged errors for voice only $EN_V$, gesture only $EN_G$ and combined voice and gesture $EN_{VG}$ commands are shown on Table 1. In accordance with the collected errors, in Table 1 are presented also the determined in percentage average precision of human mobile robots motion control using only voice recognized, or only gesture recognized commands, or together recognized voice and gesture commands.

**Table 1.** Determined average numbers of error and precision

| Tests | Errors | Precision,% |
|---|---|---|
| Voice only | $EN_V=12$ | 82 |
| Gesture only | $EN_G=16$ | 78 |
| Combined voice and gesture | $EN_{VG}=9$ | 93 |

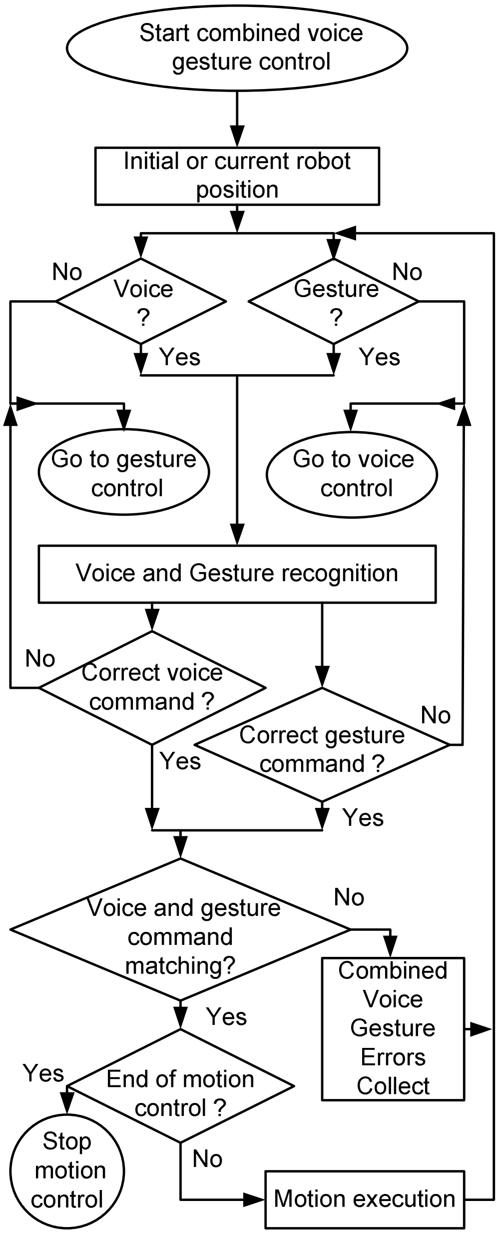


**Figure 4.** Algorithm for combined voice and gesture human control of mobile robot motion

## 5. CONCLUSION

As the conclusion the results, summarized in Table 1, can be are analysed in the following comparative way:

- there exists relationship between the collected and averaged recognition errors and the precision of human mobile robots motion control;
- the precision of human mobile robot motion control with voice only commands recognition is greater than the precision of the control with gesture only commands recognition, may be because the gesture recognition is more difficult to realize and is the source of more erroneous recognitions of gestures;
- the precision of human mobile robot motion control with combined voice and gesture commands recognition is superior than the precision of both voice only and gesture only commands recognition.

## References

[1] *N. Bellotto, and H. Hu, "Multisensor-Based Human Detection and Tracking for Mobile Service Robots", IEEE Transactions on Systems, Man, and Cybernetics - Part b: Cybernetics, Vol. 39, no. 1, February 2009, pp. 167-181.*

[2] *M. Kim, J. Lee, and S. Kang, "A Self-Organizing Interaction and Synchronization Method between a Wearable Device and Mobile Robot",* Sensors *2016,* 16*(6), 842, pp. 1-15.*

[3] *Sn. Pleshkova, Al. Bekiarski, Sh. Sehati, and K. Peeva, "Perception of Audio Visual Information for Mobile Robot Motion Control Systems", Chapter 6 from the Book Perception of Audio Visual Information for Mobile Robot Motion Control Systems, Springer, June 2014.*

[4] *L. Pérez, Ín. Rodríguez, N. Rodríguez, R. Usamentiaga, and D. García, "Robot Guidance Using Machine Vision Techniques in Industrial Environments: A Comparative Review", Sensors, 2016, Volume 16, 335, pp.1-16.*

[5] *K. Kosuge, "Human-Robot Interaction", IEEE International Conference on Robotics and Biomimetics, Shenyang, China, 22-26 August 2004.*

[6] *R. Jain, and S. Saxena, "Voice Automated Mobile Robot", International Journal of Computer Applications, Volume 16– No.2, February 2011, pp.32-35.*

[7] *K. Cekova, N. Koceska, and S. Koceski, "Gesture Control of a Mobile Robot using Kinect Sensor", International Conference on Applied Internet and Information Technologies, 2016, pp. 251-258.*

[8] *Embedded Device myRIO. National Instruments. http://www.ni.com/en-rs/support/model.myrio-1900.html*

[9] *LabView. National Instruments. http://www.ni.com/en-rs/shop/labview/labview-details.html*