Development of System Model for Audio Visual Control of Mobile Robots with Voice and Gesture Commands

Snejana Pleshkova[1)], Zahari Zahariev[2)],

[1,2)] Faculty of Telecommunications,
Technical University - Sofia, Sofia, Bulgaria,
snegpl@tu-sofia.bg

Abstract*: There are many methods for controlling the movement of mobile robots. Some of them are based on received information from the audio and video sensors of the robot. Some methods are using separately the audio or the video information, but the purpose of the current article is to combine both, using audio and video input in parallel, and by this to increase the effectiveness of controlling the movement of the mobile robot. This is very important in the cases when the human operator using both gestures and voice commands to navigate and control the robot.*

1. INTRODUCTION

There are many methods for mobile robots movement control [1]. For the purposes of this article it is looking at the audio – visual methods for controlling a mobile robot, for which there are numerous publication on the topic [2], but in the current one is taking in account the specific way of parallel control using both visual gesture commands and audio voice commands. For developing the algorithms in this article are used algorithms for gesture recognition and voice recognition.

2. BLOCK SCHEMA OF SYSTEM MODEL FOR AUDIO VISUAL CONTROL OF MOBILE ROBOTS MOTION

On Fig.1 is shown the functional block schema of the proposed system model for movement control of mobile robot with gesture and voice recognition.

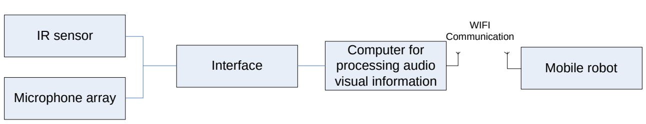

Fig. 1. Block schema of the system model.

The functional blocks for gesture and voice commands of a mobile robot are: IR sensor, Microphone array, Interface, Computer, Mobile robot and Wi-Fi Communication. There are many possibilities to realize practically the blocks on Fig.1.

For capturing gesture send to the robot from the human is proposed to use not an ordinary and expensive thermal IR camera, but a cheaper IR sensor (Fig.1).

Microphone array can be realized as linear or circular type with choosing the suitable number of microphones as elements of microphone array.

Interface provided the correct transfer to the computer the captured thermal images from the IR sensor and audio information from microphone array. The choice of the interface type can be from the existing standard interfaces, but strongly depend from the specific characteristics of the data (in this case thermal images and audio information), which are transferred and from hardware of IR sensor and microphone array.

Computer receive, process the audio and thermal information and send (via Wi-Fi Connection) the results from processing as voice and/or gesture commands, which are executed from the robot in its current steps of motion.

2. ALGORITHM OF SYSTEM MODEL FOR AUDIO VISUAL CONTROL OF MOBILE ROBOTS MOTION

For the proposed on Fig.1 system model is developed the corresponding algorithm (Fig.2) for interpretation and processing of audio and visual information in tasks of mobile robot motion control with recognized gesture and voice commands send from human to the robot.

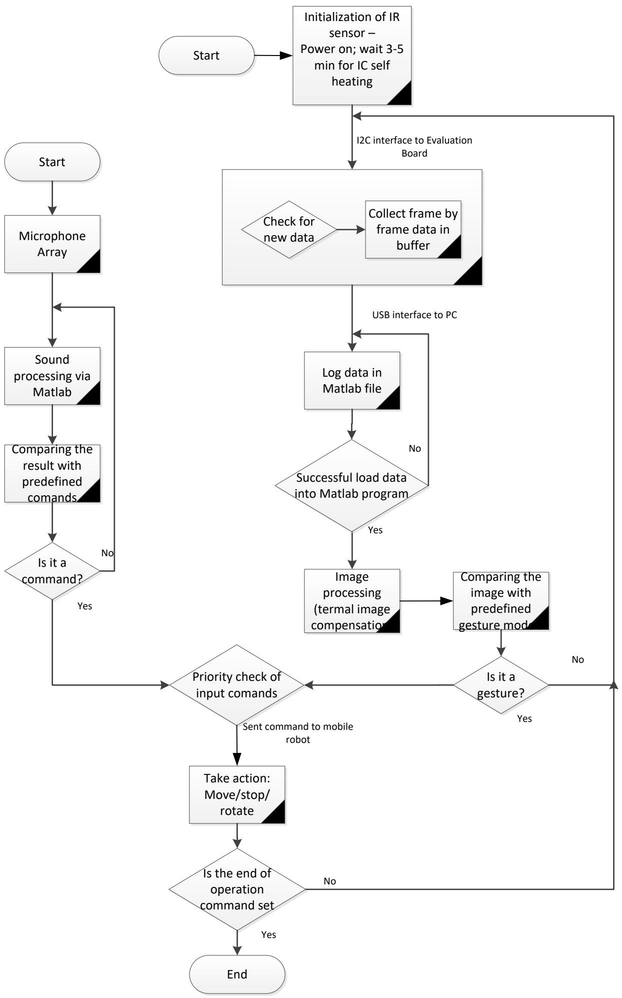

Fig. 2. Algorithm for processing and interpretation of audio and visual information

In this paper is take a closer look on the gesture recognition and thermal gesture image processing.

For the voice commands processing the algorithm (on the left side in Fig. 1) is as follow: collect some Etalon voice commands from the operator (teaching phase); store the collected commands in the memory (SD card, HDD, SSD); when there is incoming command it is compared to the etalon ones; if the comparison shows equality under certain limit (bot are similar more than 95 %) the specific action is taken.

Gesture recognition algorithm is separated in two different stages – learn stage of algorithm (presented on Fig.3) and test stage of recognition of received gesture commands algorithm (presented on Fig.4).

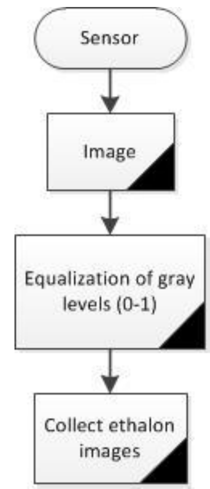

Fig. 3. The learn stage of the gesture recognition algorithm.

From IR sensor are taken the IR images of the gestures. After that the captured IR images are sent to computer (shown on Fig.1) for brightness levels adjustment using the appropriate grading of pixel intensity from “0” (cold temperature corresponding to background) to “1”(warm temperature corresponding to the example of hand gesture). After that step a new image is stored in a memory of Etalon images.

For the next step is needed to evaluate the similarity of the gestures. This is why it is repeat the etalon gesture several times. Using Matlab implementation of the algorithm shown on Fig. 2 it is proposed in the experiments presented in this article to prepare direct comparison of the observed images. For other new input images of the same gesture the difference between Etalon ones can be calculated and used as parameter for decision whether or not the Image difference – “mean(mean(A)) new input images belong to desired gesture. If this parameter is less than some threshold with the values determined experimentally (0.03 in the carried out here experiments and based on real measurements – 10 times repeated the same gesture), therefore the images belong to the same gesture.

3. EXPERIMENTAL RESULTS

Based on the proposed block schema shown on Fig. 1 and on the developed algorithm in Fig. 2 are prepared the experiments as software simulations using audio and thermos-visual information from real audio thermo-visual sensors mounted on moving platform of mobile robot.

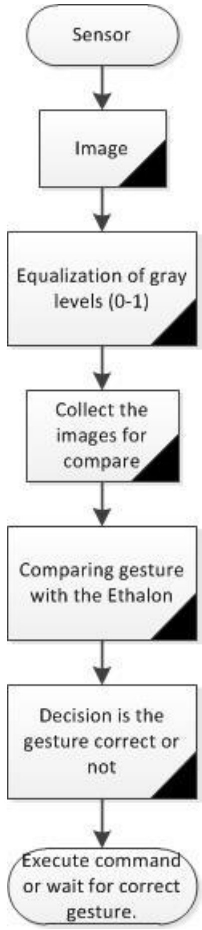

Fig. 4. Test stage of the gesture recognition algorithm.

The experiments are carried out with the following types of devices shown generally as blocks in Fig.1: Mobile robot Surveyor SRV-1 (Fig.5 left) with IR sensor and Wi-Fi module [7]; IR – Sensor MLX90640 (Fig.5 right) with resolution 32x24 pixels [8]; Microphone array with MEMS microphones type STEVAL-MKI155v2 (Fig.6 left) on evaluation board STEVAL-MKI 138v1 (Fig.6 right) [9].

The arranged with the above mentioned devices schema for receiving and processing audio and thermo-visual information is tested and used in the experiments for interpretation and processing of audio and visual information in tasks of mobile robot motion control with recognized gesture and voice commands send from human to the robot. The results from these experiments are summarized and described here briefly (only algorithm for the gesture recognition shown in right side in Fig. 2) to demonstrating the proper preliminary processing of audio and thermos-visual information and translating this results as commands to the mobile robot.

Fig. 5. Mobile robot Surveyor SRV-1

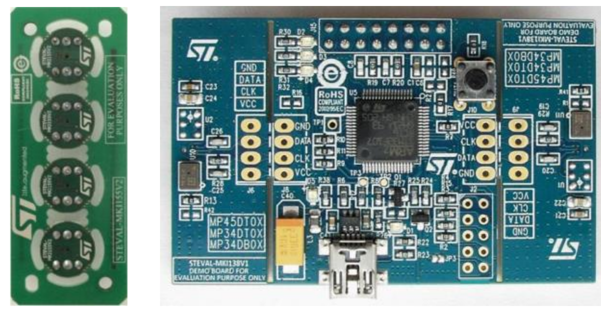

Fig. 6. MEMS microphone array STEVAL-MKI155v2 on evaluation board STEVAL-MKI 138v1

On Fig.7 is presented as example the received from IR – Sensor MLX90640 thermo-visual image (as numerical data shown in pseudo colors) of a chosen gesture.

Fig. 7. Thermo-visual image (as numerical data shown in pseudo colors) chosen as gesture received from IR – MLX90640 thermo-visual sensor.

On Fig.8 and Fig 9 are shown the results from processing of audio and thermo-visual information to achieve the correct gesture recognition using the proposed algorithm on Fig.2 implemented in Matlab [10]. All thermo-visual gesture images shown on Fig.8 and Fig 9 are presented after pseudo colors to black and white transformation. The correct gesture recognition is demonstrated on Fig. 8. On the left is thermo-visual gesture image used as reference image in the recognition. The middle on Fig. 8 is the presented the same (but with little variations) thermo-visual gesture image capture from IR sensor of the mobile robot used as test to correct recognition. On the right is thermo-visual gesture image as difference between two thermo-visual gesture images for which the calculated value of difference is very small - 0.023 and is used as criterion of decide for a correct chosen gesture recognition.

Fig. 8. Correct gesture recognition of a chosen thermal image: on the left - reference gesture; on the middle - variation of the same gesture; on the right - image difference between two gestures for which the calculated value of difference is very small - 0.023, i.e. the recognition is correct.

The example of wrong gesture recognition is demonstrated on Fig. 9. On the left is shown also the reference thermo-visual gesture image used in the recognition. The middle image on Fig. 9 is also the same (but with larges variations) image of gesture capture from IR sensor of the mobile robot used as test to wrong recognition. On the right is image difference between two gestures for which the calculated value of difference is bigger - 0.13 and is used as criterion for decision of wrong gesture recognition.

Fig. 9. Wrong gesture recognition of a chosen thermal image recognition: on the left - reference gesture; on the middle - variation of the gesture; on the right - image difference between two gestures for which the calculated value of difference is bigger - 0.13, i.e. the recognition is wrong.

3. Conclusion

The briefly presented results demonstrate the ability of the proposed on Fig.2 algorithm to process of thermo-visual information for correct gesture recognition. The information for recognized gestures can be combined with the information from recognition of voice commands (not presented here – the algorithm for the voice recognition is shown on left side in Fig. 2) and send as is shown on Fig. 1 as joined gesture-voice commands from human to the mobile robot, which is the goal of this article - to correctly control the mobile robot motion with gesture-voice commands from human. Of course, the proposed algorithm is the subject of future improvements and enhancements. The improvements in the proposed algorithm can be in terms of training with various and numerous variations of gestures and voice commands, and also to investigate the influence of the joint usage of gestures and voice commands in enhancement the accuracy of mobile robot motion control from human commands.

Acknowledgment

This work was supported by Technical University – Sofia inner program to research projects under 172PD0035-07: “Development of methods and algorithms for recognition of gestures and speech, and processing to interpret the received commands”.

References

[1] Everett. H., "Sensors for Mobile Robots: Theory and Application”, *Wellesley*, MA, 1995.

[2] Gabriel G., C. Nascimento, and Ed. Yagyu, “Mobile Robot Path Control Using Landmark Information”, *ABCM Symposium Series in Mechatronics*, Vol. 2, 2006, pp. 2-37.

[3] Malima, A., E. Ozgur, et al., “A Fast Algorithm for Vision-Based Hand Gesture Recognition for Robot Control”, *Signal Processing and Communications Applications, 2006 IEEE 14th.*, 2006.

[4] Nickel K., R. Stiefelhagen, “Visual recognition of pointing gestures for human–robot interaction Interactive Systems Labs”, *Universitaet Karlsruhe, 76131 Karlsruhe, Germany. Sciencedirect journal*, Vol. 25, Issue 12, 3 December 2007, pp. 1875-1884.

[5] Davis J., M. Shah, “Recognizing hand gestures”, *in: Proceedings of the European Conference on Computer Vision, ECCV*, 1994, pp. 331-340.

[6] Waldherr, S., Romero, R., and S. Thrun, “A gesture based interface for human-robot interaction”, *Autonomous Robots*, Vol. 9(7), 2000, pp. 151–173.

[7] Surveyor SRV-1 Blackfin Robot Manual. http://surveyor.com

[8] https://www.melexis.com/en/product/MLX90640/Far-Infrared-Thermal-Sensor-Array

[9] http://www.st.com/mems-and-sensors/

[10] https://www.mathworks.com/