VISUAL DESIGN OF MOBILE ROBOT AUDIO AND VIDEO SYSTEM IN 2D SPACE OF OBSERVATION

Alexander Bekiarski, Snejana Pleshkova-Bekiarska

Technical University – Sofia, Bulgaria
St. Kliment Ohridski, 8
Tel.: +359 965 3300; E-mail: snegpl@tu-sofia.bg

Abstract

The mobile robot consists from a lot of subsystems. The main system is moving system, which is a combination of mechanical, electric and electronics parts. But the information, which control the movements the robot is collecting, processing from the sensors system. The most natural sensors in sense, that the mobile robot is similar to people, are audio and visual sensors. They are most informative between all mobile robot sensors, but it is very difficult to process and to separate with these sensors only the information useful for the movement of the robot.

This difficulty increase, if it is necessary to add the conditions and characteristics of the area or space of observation of the mobile robot. Therefore the goal of this article is to facilitate the design of audio and visual robot system with the developing a user friendly tool as a visual interface in the steps of simulation and modeling the audio and video robot system.

1. INTRODUCTION

The mobile robots perform the audio and visual information from the corresponding sensors [1], [6]. Most popular audio and video sensors are microphone arrays and video cameras. They are placed on the mobile robot moving device and collect the sounds and images from the robot space of observation respectively. Sounds are transformed from each of the microphones in the array as sound signals [2] and images from the video camera (or cameras in case of stereo robot vision) as video signals [3]. The main goals in the processing of sound and video signals are:

– to separate the signal of interest (SOI) from the combination of sound signals, received from microphone array (usually this signal is speech signal from a speaker talking to the robot in area or space of observation;
– to calculate or estimate the sound or speech signal direction of arrival (DOA) (this is a kind of sound localization);
– to separate the visual object from the received pictures (usually the speaker in space of observation);
– to calculate the position of separated visual object from the received pictures;
– finally, to joint the information as co-ordinates of speaker, calculated from audio and video system, and use this joined information for choosing the right direction for the control of the robot moving system.

Some or all of these mentioned tasks can be simulated separately or in combination, but in all cases exist the problem of a fast and simple modification of the characteristics of space or area of observation, room dimensions, mobile robot position, speaker places, etc. This mean, that it is necessary to have a tool in computer simulation programme, which gives a visual representation of robot space of observation with the possibility to interactive changes of room dimensions, robot and speakers places, etc. The simple case for design a visual interactive tool in simulation of a robot audio visual system is to represent the robot space of observation as a empty rectangle, i.e. as a 2D space of observation and then place in this rectangle the initial robot and speaker positions. This is the goal in this article.

2. VISUAL 2D REPRESENTATION OF ROBOT SPACE OF OBSERVATION

In general case the robot space or area of observation can be arbitrary, but here it is chosen to use an empty room as a concrete space of robot observation. It is also accepted to consider the room only in 2D dimensions length "l" and width "w". This assumption is applicable, because in most the cases the robot movements are on the room floor.

The proposed visual design of 2D space of observation of the robot is presented in Fig.1 as an algorithm of steps, which can be used first for the initial states settings and then as an interactive work in the current simulation of the audio and visual situation in area or space of robot observation. In the Fig.1 are shown only initial steps of the algorithm.

The first step of the algorithm is for the input of the room dimensions length "l" and with "w", which here are present as variables "rooml" and "roomw" in the algorithm. Then, for the visual representation on a screen, it is necessary to calculate the resolutions Nx and Ny of display in direction x and y (related for example with room length and width respectively). For these calculation, one of these relative and discrete dimensions Nx or Ny can be chosen, for example Nx = = 1024 to guaranty a good and precise resolution in x direction. Then it is defined the coefficient Kx as a relation between metric and discrete room dimensions:

$$K_x = \frac{rooml}{Nx}. \qquad (1)$$

The equation (1) can be used in the calculation of discrete room dimension Ny in direction y:

$$N_y = \frac{roomw}{Kx}. \qquad (2)$$

The chosen and calculated Kx from equations (1) and (2) discrete room dimensions Nx and Ny can be used for room drawing and displaying as an rectangle. The next step in the algorithm shown in Fig.1 allows interactive to choose the position of the robot in the room space, i.e. in the rectangle, which is already drawn. The interactive chosen robot position is represented in rectangle room space as robot co-ordinates xr and yr, respectively.

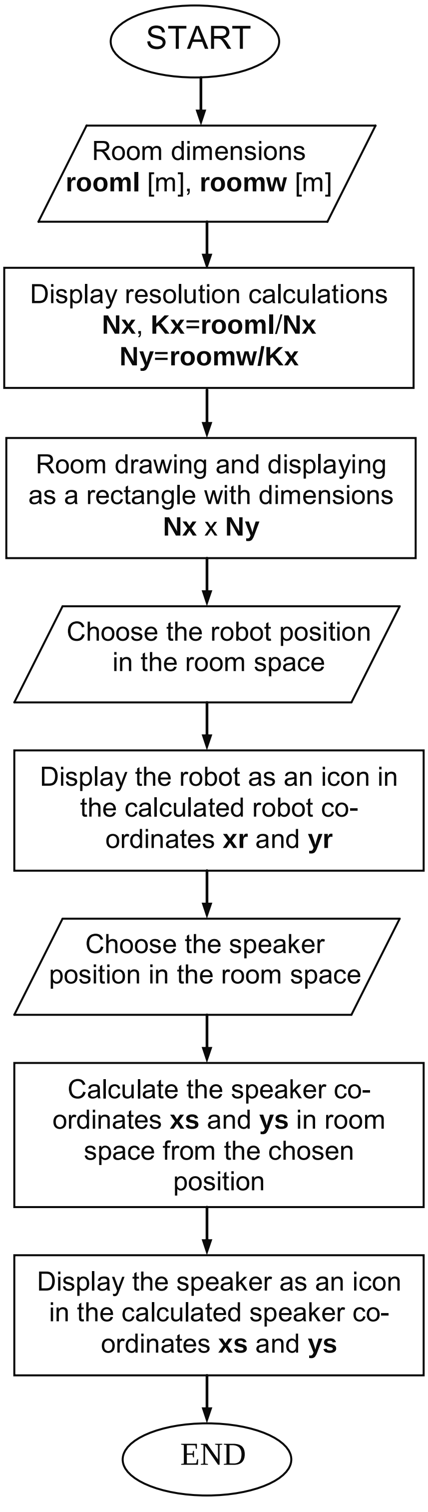

Fig.1. Initial state of algorithm

The visualization of the robot place in the area of room rectangle is made with an icon of an stylize and simplified image of a robot.

In the similar way the next steps of the algorithm shown in Fig. 1, are for an interactive choosing the speaker position in the space of rectangle of robot observation. Then this position is transformed after some calculations to the absolute speaker co-ordinates xs and ys in the space of robot observation. This is necessary to display the speaker as an initial place in the space of observation and on an appropriate distance from the robot position.

The results of the visual design algorithm in the steps of initialization are shown in Fig. 2.

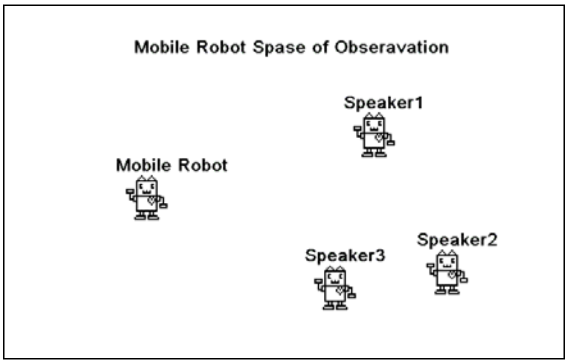

Fig. 2. Mobile robot space of observation

Of course, it is possible to extend the initial audio and visual situation in space of observation with more than one speaker (Speaker1, Speaker2, speaker 3 in Fig.2), choosing the additional positions for other speakers in the area robot observation. Also it is necessary to define the co-ordinates xm, ym of microphone array and the visual sensor xv, yv or video camera to be equal of the robot co-ordinates xr and yr in 2D space of observation representation:

$$xm=xr;\ xv=xr \quad ym=yr;\ yv=yr. \qquad (3)$$

When preparations of the initial stages are finished, the steps for preparing and calculating the audio and video system characteristics are performed.

For an audio robot system these characteristics are the number of microphones nmic in microphone array, the distance dm between microphones and length lmicar of microphone array, if it is chosen a linear array. The relations of these characteristics are expressed as:

$$Dm=lmicar/nmic. \qquad (4)$$

Also for a video robot system it is important to choose the type of video system: mono or stereo, the focus distance, the ability to change direction of observation, etc.

The initial step is followed by a concrete simulation or testing the work of the audio and video robot system. Each of two robot systems – audio and video execute at the same time different tasks. Audio system performs sound localization and calculates direction of talk to the robot. Video system, also process the images of the room to separate the place of the speaker. The results of the operation of audio and video systems are the direction of arrival and the co-ordinates of the speaker, which are calculated both from audio system [4]:

$$\theta_d = \frac{\sum_{\theta=1^0}^{360^0} R_{ij}[k(\theta)] x\theta}{\sum_{\theta=1^0}^{360^0} R_{ij}[k(\theta)]} \qquad (5)$$

and video system [5]

$$xs = \frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} a_i}; \quad ys = \frac{\sum_{i=1}^{n} a_i y_i}{\sum_{i=1}^{n} a_i}, \qquad (6)$$

where

Θ_d is the angle of direction of arrival;

$R_{ij}[k(\Theta)]$ – cross correlation in received microphone array speech signals in relation to angle $\Theta=1^o \div 360^o$;

xs and ys – speaker coordinates, calculated from visual system;

x_i and y_i – current image co-ordinates;

a_i – the brightness value of current image point.

The equation (5) and (6) gives the final results from the work of audio and video system. They serve as audio and video information for the proposed here visual design and representation of moving robot audio and video system.

The interactive part of the visual design with simulation of audio and visual system operation is shown in Fig. 3.

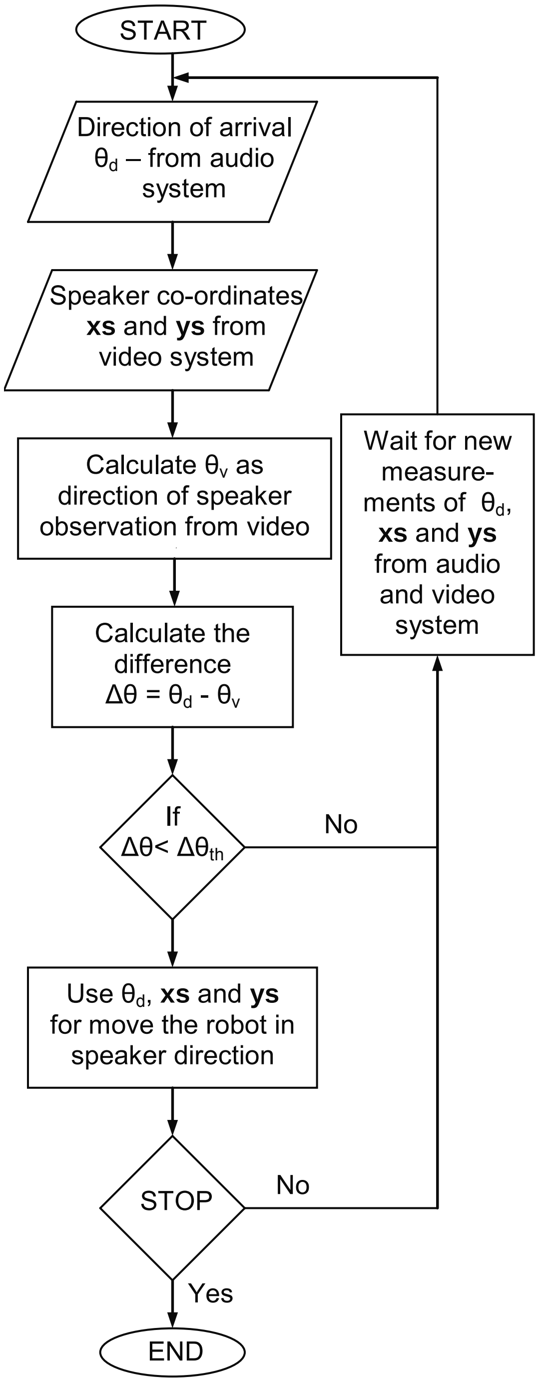

Fig. 3. Interactive state of algorithm

This part of algorithm start after the initial steps, described and shown in Fig.1.The current direction of arrival as angle θ_d is inputted as a result of work of audio robot system in the first step of algorithm in Fig. 3. Also the current speaker co-ordinates xs and ys are entered, which are derived form visual robot system. From these co-ordinates it is possible to calculate the speaker direction θ_v as:

$$\theta_v = \arcsin \frac{xs - xr}{ys - yr}. \qquad (7)$$

Two directions θ_d, calculated from audio system (equation (5)) and θ_v, calculated from video robot system (equation (6)), must be equal if both audio and video systems determine the exactly speaker position as a direction of robot observation. But in a real situation, each of these systems can calculate the direction with an error from the precise speaker position. Therefore, it is necessary to choose which of these two directions to use for move the robot or two directions are so much different, that it is not possible to use them.

This is shown in the next steps of algorithm in Fig. 3, when the difference $\Delta\theta$ between two angles θ_d and θ_v is calculated:

$$\Delta\theta = \theta_d - \theta_v. \qquad (8)$$

The decision of using or not the appropriate angles θ_d and θ_v in the action to move the robot in speaker direction is made after a comparison of difference $\Delta\theta$ with some practically chosen threshold $\Delta\theta_{th}$:

$$\Delta\theta < \Delta\theta_{th}. \qquad (9)$$

If the condition (9) is satisfied, then the step to move the robot in speaker direction is executed. This is accomplished first with an averaging to directions θ_d and θ_v:

$$\theta_{av} = \frac{1}{2}(\theta_d + \theta_v), \qquad (10)$$

to calculate an average direction $\Delta\theta_{av}$, and then to use this average direction in the action to move robot. This proposition can reduce or average the two possible existing errors of θ_d and θ_v determination from audio and video robot system, respectively.

If the condition (9) is not satisfied, then it is necessary to wait until a new measurement of

θ_d, xs and ys are performed from the audio and video robot system, respectively.

The iterative process to calculating the speaker direction and its co-ordinates from audio and visual system and moving the robot can be stop in a chosen moment for ending of the simulation.

3. RESULTS AND CONCLUSION

As an example of the work of the algorithm in Fig. 3 it is shown a situation in room with a robot and three speakers, when the Speaker1 is talking. The results of audio and video system actions are shown with different line styles for clearly separation of real direction (solid line), audio system direction of arrival determination (dashed line) and observation from video system the direction of Speaker1 (dotted line). Also in Fig. 4 are shown the values of angle (theta=60^o) calculated from equation (3) and Speaker1 co-ordinates xs1 and ys1, calculated from equation (4).

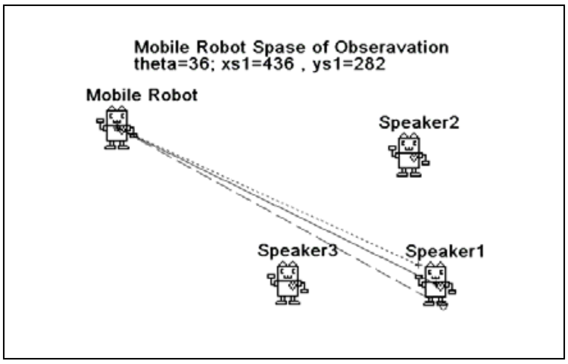

Fig. 4. Mobile robot space of observation after initial state

A full sequence of robot movement simulation in direction of the Speaker1 is shown in Fig. 5, where it is seen, as a sequence of points, some of the current robot positions, calculated from audio and video system in the movement of the robot to the Speaker1. These points are near as possible to the real direction to the Speaker1.

The presented work of the proposed visual design tool or interface for simulation and testing the ability of audio and video robot system shows, that this visual tool give good possibilities to express all steps: initial, interactive and final of the algorithm, in a very convenient form. Also this visual tool allows analysing the results for the working, capacity and efficient of the audio and video robot system. In the future works it is expected to extend the view and the capabilities of this tool for visual representation and interactivity in the real time working robot audio and visual system.

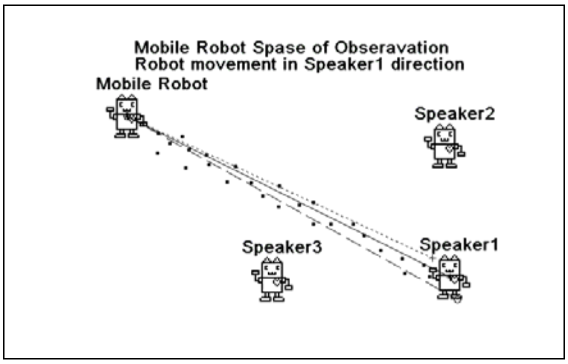

Fig. 5. Robot movement in Speaker1 direction

4. ACKNOWLEDGMENTS

This work was supported by National Ministry of Science and Education of Bulgaria under Contract BY-I-302/2007: "Audio-video information and communication system for active surveillance cooperating with a Mobile Security Robot".

References

[1] J. Fritsch, M. Kleinehagenbrock, S. Lang, G.A. Fink, and G. Sagerer. Audiovisual Person Tracking with a Mobile Robot. Proceedings of Biefeld University, Faculty of Technology, pp.207-216, Biefeld, Germany, 2007.

[2] D. Giuliani, M. Omologo, and P. Svaizer. Talker localization and speech recognition using a microphone array and a cross-power spectrum phase analysis. In Proc. Int. conf. on Spoken Language Processing, volume 3, pages 1243-1246, Yokohama, Japan, 1994.

[3] S. McKenna, Y Raja, and S. Gong. Tracking color objects using adaptive mixture models. Image and Vision Computing, 17, pp. 225-231, March 2006.

[4] Sn. Pleshkova, Al. Bekiarski. Sound Propagation Model for Sound Source Localization in Area of Observation of an Audio Robot. WSEAS.2008.

[5] M. Grau, Al. Bekiarski, P. Venkov, I. Nacheva. Robot Contol with Camera Eye. CEMA'07,Sofia, 2007.

[6] Milushev M., S. Hecht, T. Djamiykov Design of a Sensor for the Stepping Contact Force of a Walking Machine 53rd Internationales Wissenschaftliches Kolloquium - Technische Universität Ilmenau (53. IWK), 08 – 12 September, 2008 pp. 77-79.