

Advanced Research in Computer Engineering

Sandip A. Kale Editor

Research Transcripts in Computer, Electrical and Electronics Engineering

6

The Method of Depth Map Calculating Based on Soft Operators in Multi-Agent Robotic Stereo Vision Systems

M. V. Bobyr^{a,*}, N. A. Milostnaya^a, S. V. Gorbachev^b, S. Bhattacharyya^c and J. Cao^d

^aDepartment of Computer Science, Southwest State University, Kursk, Russia ^bNational Research Tomsk State University, Tomsk, Russia ^c Rajnagar Mahavidyalaya, Birbhum, India ^dSoutheast University, Nanjing, China and Yonsei University, Seoul, Korea

*Corresponding author: maxbobyr@gmail.com

ABSTRACT

The fuzzy method of depth map calculating using stereo images obtained on the path of mobile robots (agents) is considered. The method is based on SAD (sum of absolute difference) algorithm composition and fuzzy inference. Its special feature is soft arithmetic operators with fuzzy implication usage. The accuracy of the constructing depth map method is estimated by RMSE (root mean square error). The bestest soft operator has minimum RMSE. The method of depth map calculating which has seven steps is presented. The proposed method has showed that the accuracy of the SAD algorithm increases by 20% when soft operators is used. This conclusion is confirmed by the simulation results presented in the chapter.

Keywords: Mobile Robots, Stereo Vision, Depth Map, Soft Computing, SAD, Fuzzy Logic

1. INTRODUCTION

The development and development of mobile robotic systems (MRS) and the system is going on very intensively in today's time. Navigation problems are one of the main problems encountered by mobile robots. Localization, mapping, and route planning are important parts of the autonomous device navigation task [1]. Localization is responsible for accurately determining the current position of the robot in the surrounding space. The maps are built by collecting the robots ' sensory data and storing it in a form that is convenient for subsequent processing. Using the data obtained during the construction of the map, it becomes possible to plan a route that allows you to reach the whole from the starting position, without allowing a collision with obstacles. The Simultaneous Localization and Mapping (SLAM) method combines the two processes to provide more efficient offline navigation. SLAM algorithms are able to cope with various hardware or external constraints, for example, SLAM methods for monocular systems [2], SLAM methods using timely lidar and visual sensors [3].

In many cases, the capabilities of an individual robot will not be sufficient to solve the tasks set [15]. There is a need to use mobile robotic groups. One of the important tasks that is performed in the management of such a multi-agent system is the formation and support of robots of a given order. Methods of group control of unmanned aerial vehicles, methods based on the use of computer vision systems, neural networks and fuzzy logic [16], and advanced robotics are being actively developed. Recklessness on a large number of existing solutions to problems of movement of groups of robots system, when implemented, problems can be with the limitation of available computing resources of the mobile robot, and each robot and groups in general.

The use of SLAM and autonomous navigation methods when controlling moving groups of MRS formations requires the addition of the motion control system of each robot with functionality related to obtaining and analyzing information from the surrounding space. In this case, the obstacle detection subsystem is of key importance, since obstacles play the role of "anchors" for linking the terrain map under construction. Solving the problems of detecting obstacles when navigating outdoors requires recalling up-to-date information about the surrounding space in the form of three-dimensional dynamic images.

A large research and publication concerns monocular vision and laser sensors. From where stereo vision technologies take on additional benefits. The hot range information preceding the stereo vision is more accurate than the information preceding the laser sensors, it can be used effectively to calculate the three-dimensional placement and configuration of the object at a lower cost. In addition, in a complex noisy environment, stereo vision can recognize the necessary volume, reducing such cheeks sliding with background elements. Secondly, it is expected that systems using stereo vision will be more reliable in real scenes, where sudden changes in awareness often occur.

Problems associated with determining the location of moving and / or stationary obstacles on the path of a mobile robot arise during the development of a control system for it [4, 5]. To determine the relative position of objects, it is necessary to process data about the depth of the video scene and obstacles located along the path of the robot. Various methods are used to construct such scenes, for example, obtaining data from scanning lidars or using stereo vision systems [6]. It is necessary to solve a number of problems associated with finding objects located on stereo images during of developed of the stereo vision systems [7]. A depth map is a black and white image with grayscale color, in which the brightness of individual objects indicates the proportional distance to them. The main algorithms for constructing depth maps are SGM (Semi-Global Matching Algorithm) and SAD (Sum of Absolute Differences algorithm) [8]. Several passes are required to calculate the depth map using the SGM algorithm, while the second algorithm uses only one pass along the image, which reduces the processing time required to calculate the disparity value. It should be noted that 1 frame of stereo information in HD format (720p resolution 1280×720 pixels) should be processed in 1/30 second, that is, in 33 ms. Therefore, it takes about 50 ns to process 1 pixel. Therefore, decisionmaking time is one of the main factors that must be considered when developing stereo systems for controlling mobile robots. Fuzzy logic is one of the ways to improve the accuracy of the stereo vision model [9]. However, fuzzy models have a number of system errors that reduce their accuracy [10]. It is recommended to use soft arithmetic operators during the fuzzy implication to eliminate these errors.

2. FOUNDATIONS

The obstacle detection subsystem for MRS navigation in a dynamically changing environment must meet stringent requirements. The need for a moving mobile robot to respond to changing environments in a timely manner, especially at high speeds, implies that the obstacle detection and classification system operates in real time. This implies the necessary maximum performance in conditions of operation on an on-board computer with limited computing power. Increased performance can be achieved by parallelizing data processing algorithms, so it is desirable to perform parallel execution of more resourceintensive processing steps to improve performance. Also, the dynamically changing environment imposes the required size of the guaranteed radius of detection of dangerous obstacles. This value must exceed the length of the braking distance of the mobile robot at the maximum possible speed, including in bad weather conditions. Terrain patency at short and medium distances can be estimated using existing solutions, to minimize trajectory planning errors, the priority duty is to be obstacles at a long distance. In addition to the coordinates of the detected obstacles to plan the motion path of the robot is important additional information about the dimensions, the characteristics of the object and its overcome (for example, you can pass through low bushes, but the attempt to overcome tree trunk can result in the release of MRS of the car). The obstacle detection subsystem should allow you to get an estimate of the specified parameters for the detected objects. The sensor system used to obtain three-dimensional information about the surrounding space should ensure the functioning of the obstacle detection subsystem at your favorite time of day and have the maximum effective range performance. The sensor system should transmit the resulting three-dimensional spatial information in the video of a three-dimensional image (point clouds) - the multiplication of the coordinates of the vertices, the spatial mood that characterizes the surrounding objects. Cloud points (vertices) can carry additional information about the power characteristics of an object, such as blooming or reflectivity. To process the incoming information, it is necessary to use the methods of three-dimensional computer vision.

Often, certain methods and algorithms cannot always be successfully applied in solving traditional problems of technical vision, there are no unified methods, which are universal. This is due to the fact that the technical vision systems of robot must satisfy hard trackers, which have a decently complex implementation, besides, no sensor appears ideal for solving a certain task.

From the point of view of the implementation of the methods, there are a number of additional requirements for the technical vision system. In contrast to the algorithms for analyzing time signals, processing and analysis of images are more complex, since they represent three coordinates and time.

In the process of analyzing the literature revealed that this analysis of images is often considered as unrelated processes. It is necessary to conduct a comprehensive analysis throughout the study, while loading 6the speed of the newly developed algorithms or modifying them. In offers segmentation methods, describes the possibility of direct access to the source code. But the methods described in ra6ot are not suitable for systems that operate in real time. Today, complex hardware and software tools (APS) are used in three-

dimensional technical vision systems, which affects the production processing methods for solving the problem in real time.

These problems can be solved by developing soft fuzzy algorithms and creating new methods of processing to extract such data as flexible in this respect. To create high-performance 3D technical vision systems, it is necessary to exclude the operation of matching stereo pairs when analyzing frames, to ensure the receipt of spatial video information by a single mobile video sensor, to increase the noise immunity and to develop a simple software of support – to synthesize a three-dimensional image, it is necessary to determine a strategy for images obtaining that lack projections of the object based on the results of images obtaining of the original object. To the basics of these confusions, it can be argued that the capabilities of the currently existing means of forming and image processing are limited, so when developing a video system, it is necessary to load both a trace to the completeness and quality of the information posted by them, and the possible wearing of their constituent elements.

In order to perform the task in real time, the algorithms used should be more simple and allow for efficient computational implementation. But the high range of tasks and the desire for the universality of the system force us to abandon simple algorithmic solutions and switch to complex methods of processing and analysis of video information. It is also worth noting that it is possible to organize the system's flexibility for calculating various different algorithms with a large number of variable parameters. It is also necessary to note that the operators of robots are often ordinary users, not highly qualified specialists, and often they do not fully control the mathematical methods of video information processing, as well as programming. The above-mentioned specifics of solving technical problems actualizes the development of completely new hardware and software tools, which will effectively supported all the necessary modes of operation of the robot technical vision systems.

Based on the above, we can draw the following conclusion: when developing multi-agent stereo vision systems in complex environmental conditions, the range of tasks to be solved with the help of robot is significantly complicated and, satisfactorily, the need for the development of a universal technical vision system, equipped with satisfying means for solving complex functional problems arising from the conditions of work, increases. At the same time, the defining qualities of the functional should be the speed, accuracy and flexibility of information processing methods when solving the problem in real time.

2.1. Statement of the problem

Let's set the task of improving the accuracy of the stereo vision system for constructing depth maps based on the estimate of the RMSE (root mean square error) coefficient:

$$RMSE = \sqrt{\frac{1}{w \times h} \sum_{i=1}^{h-1} \sum_{j=1}^{w-1} (I'_{i,j} - I_{i,j})^2} \to \min,$$
(1)

where w, h are width and height of the image; $I'_{i,j}$ is brightness level on the reference image; $I_{i,j}$ is brightness level in the output image.

For the study, we will use soft operators with the following formulas:

- Operator Formulas for fuzzy implication
- MIN $I = \min(a, b)$ (2)
- $PROD \qquad I = a \times b, \tag{3}$

MEAN
$$I = \frac{a+b}{2}$$
, (4)

Soft-min I
$$I = \frac{a+b+\delta^2 - \sqrt{(a-b)^2 + \delta^2}}{2},$$
 (5)

Soft-min II
$$I = \frac{\sum_{i=1}^{n} a_i e^{-ka_i}}{\sum_{i=1}^{n} e^{-ka_i}}, z \partial e k = -100.$$
 (6)

where a, b are operands; δ is coefficient of softness ($\delta = 0.05$); n is the number of operands.

The most accurate method is the one with the minimum RMSE.

3. RESEARCH METHODOLOGY

Fuzzy method for constructing a depth map from stereo images

The calculation of fuzzy method for constructing a depth map from stereo images is performed in several steps.

Step 1. The calculation of the difference in intensity levels on stereo images:

$$SAD(x, y, d) = \sum_{m=-1}^{m=1} \sum_{n=-1}^{n=1} R \sum_{m=-1}^{m=1} G \sum_{m=-1}^{m=-1} B \left| I_{l} \left(x + m, y + n \right) - I_{r} \left(x + m + d, y + n \right) \right|$$
(7)

where I_b I_r are color intensity on the left and right stereo images; x, y are coordinates on stereo images; R, G, B - components of the three-channel color model.

The calculation the disparity from array of the difference in intensity levels on stereo images:

$$D(x, y)_{SAD} = \underset{d \in [0..d_{max}]}{\operatorname{arg\,min}} SAD(x, y, d).$$
(8)

The SAD algorithm in graphically view is presented in Figure 1.



Depth Map

Fig. 1. Graphical interpretation of the SAD algorithm

Step 2. Calculation of the degrees of input membership functions (MF). In order to increase the speed of the fuzzy multi-criteria decision-making (FMCDM) system and implement it on the basis of Field-Programmable Gate Array

(FPGA) it is recommended to use the type of MF for calculating the degrees of input membership functions which fewer arithmetic operations are needed. For instance, to calculate triangular MF, only subtraction and division operations are required. The subtraction operation in FPGAs is simple. To implement the division operation, it is recommended to select the difference between the MF labels to be a multiple of a power of 2 (2, 4, 8, 16, 32 ...). Then the division operation is replaced by a right shift operation. Input MFs in FMCDM are shown in Figure 2, a.



Fig. 2. Membership functions: a – input MF; b – output MF

The degrees of the membership functions of the input variables are calculated using the following formulas:

$$A_{1i} = \begin{cases} (5000 - SAD(x, y, d)) / 5000 \, if \, SAD(x, y, d) < 5000 \\ 0, else \end{cases}$$

$$A_{2i} = \begin{cases} SAD(x, y, d) / 5000 \, if \, SAD(x, y, d) < 5000 \\ (10000 - SAD(x, y, d)) / 5000 \, if \, SAD(x, y, d) \ge 5000 \\ 0, else \end{cases}$$

$$A_{3i} = \begin{cases} (SAD(x, y, d) - 5000) / 5000 \, if \, SAD(x, y, d) > 5000 \\ 0, else \end{cases}$$
(9)

where *i* is the number of the input variable ($i = 1 \dots 3$).

Step 3. Calculation degree of conclusions of fuzzy rules. In general, the fuzzy rule has the form [12, 13]. The degree of conclusions will be calculated using the formulas shown in Table 1.

The conclusions of fuzzy rules
$Y_1 = \Theta(A_{31}, A_{32}, A_{33})$
$Y_2 = \max([\Theta(A_{31}, A_{32}, A_{23})], [\Theta(A_{31}, A_{22}, A_{33})])$
$Y_{3}=\max([\Theta(A_{31}, A_{32}, A_{13})], [\Theta(A_{31}, A_{22}, A_{23})], [\Theta(A_{31}, A_{12}, A_{33})])$
$Y_{4} = \max([\Theta(A_{31}, A_{12}, A_{13})], [\Theta(A_{31}, A_{12}, A_{23})], [\Theta(A_{31}, A_{22}, A_{13})])$
$Y_{5} = \max([\Theta(A_{21}, A_{22}, A_{33})], [\Theta(A_{21}, A_{32}, A_{23})], [\Theta(A_{21}, A_{32}, A_{33})])$
$Y_6 = \max([\Theta(A_{21}, A_{12}, A_{33})], [\Theta(A_{21}, A_{22}, A_{23})], [\Theta(A_{21}, A_{32}, A_{13})])$
$Y_{7} = \max([\Theta(A_{11}, A_{32}, A_{33})], [\Theta(A_{21}, A_{12}, A_{23})], [\Theta(A_{21}, A_{22}, A_{13})])$
$Y_8 = \max([\Theta(A_{11}, A_{22}, A_{33})], [\Theta(A_{11}, A_{32}, A_{23})], [\Theta(A_{21}, A_{12}, A_{13})])$
$Y_{9} = \max([\Theta(A_{11}, A_{12}, A_{33})], [\Theta(A_{11}, A_{22}, A_{23})], [\Theta(A_{11}, A_{32}, A_{13})])$
$Y_{10} = \max([\Theta(A_{11}, A_{12}, A_{23})], [\Theta(A_{11}, A_{22}, A_{13})])$
$Y_{11} = \Theta(A_{11}, A_{12}, A_{13})$

 Table 1. Degrees of conclusions of fuzzy rules

Note. The Θ sign means a fuzzy-logical operation of finding a hard or soft minimum using formulas (2)÷(6).

Step 4. Defuzzification.

$$D(x, y)_{d=1}^{d_{\max}} defuz = \frac{\sum_{k=1}^{11} Y_k \cdot M_k}{\sum_{k=1}^{11} Y_k}.$$
(10)

Step 5. Calculation of disparity value on the depth map.

$$D(x,y) = \arg\min_{d \in [0..d_{\text{max}}]} \left(D(x,y)_{d=1}^{d_{\text{max}}} defuz \right).$$
(11)

4. EXPERIMENTAL RESULTS

Stereo images with a size of $450\div375$ pixels (Fig. 3) were used for simulate the process of constructing a depth map [14]. This makes it possible to assess visually and quantitatively the effectiveness of the proposed FMCDM. The fuzzy inference system is formed on the basis of the MFs shown in Figure 2. The maximum depth value $d_{max} = 64$. The disparity values (see Eq. 11) were multiplied by 4 to correspond to 256 shades of grayscale. In the MS Visual Studio 2013 on the C # programming language was developed a software model that allows building depth maps and evaluating their accuracy using formula (1). Studies were carried out for the SAD method, fuzzy operators (2)÷(6) according to the method presented in [14], and using the FMCDM from stereo images. Table 2 summarizes the values of the RMSE coefficient.



Fig. 3. Stereo images: a – Teddy; b – Cones; c – Venus

Note: The figure shows the left and right stereo images and a reference depth map. It should be noted that in the methodology [11], only a hard operator was used to construct the depth map in the compositional rule of Zadeh.

Image	Method	Operator	MIN	PROD	MEAN	Soft-	Soft-
C						MIN I	MIN II
Teddy	SAD	RMSE	47.77				
	[11]	RMSE	44.91	46.74	40.98	41.09	38.42
	FMCDM	RMSE	38.65	38.86	38.09	37.76	38.32
Cones	SAD	RMSE	51.79				
	[11]	RMSE	50.81	52.76	47.12	39.77	39.7
	FMCDM	RMSE	32.43	32.83	31.89	31.64	32.45
Venus	SAD	RMSE	50.07	· · ·			
	[11]	RMSE	46.64	46.64	46.05	45.92	46.16
	FMCDM	RMSE	41.44	44.94	38.8	39.75	39.63

Table 2. Value of RMSE

Note. The minimum RMSE values are in bold italic type.

As shown by the data presented in Table 2, when using soft operators, the accuracy of the method [11] increases. Table 3 shows video images of depth maps obtained using the proposed FMCDM. The visualization of the depth map using soft operators (2) \div (6) is also shown.

Analysis of Table 4 showed that all images contain artifacts in the form of white dots. White dots on the depth map show objects that are closest to stereo cameras. For example, when a mobile robot avoids obstacles, white dots will give false objects that are in front of its path.

To compensate for artifacts, it is proposed to use the following method.

- It is necessary to evaluate the data that comes to the input of the fuzzy system in areas with artifacts. In Figure 4 (a) artifact has coordinates x = 254, y = -68. The value of the disparity at this point is equal 62. This value is predicted by analyzing the input data presented in Figure 4(b). Taking into account formula (11), the disparity is equal to the value of the fourth local minimum. Scaling 62 × 4 = 248 produces a light tone.
- 2. It is necessary to estimate the real value of disparity on the depth map. Visual assessment (Fig. 4, b) shows that the correct value of disparity on the depth map should correspond to the first minimum and be equal to 16.
- 3. Exclude from the calculation the false zones of fuzzy inference triggering. To ensure this task, it is necessary to introduce an additional condition for calculating the disparity value. Taking into account the data presented in Figure 4,b, such a condition is the rule:

IF (SAD(x, y, d) > 1200) AND $(D(x,y) \ge 32)$ THEN $A_i = 1000$



94

This compensating rule makes it possible to reduce the appearance of maximum values of the value D (x, y) determined by formula (11) (Fig. 4, c). Using the proposed rule a value equal to 16 is formed at the output of the fuzzy system.



(a)

SAD(x, y, d)SAD(x-3, y, d) SAD(x+3, y, d) dmax (b) D(x,y after correction D(x,y)D(x,y)before correction dmax (c)

Fig. 4. Depth map: a - Teddy; b - before correction D(x,y); c - after correction D(x,y)

Visualization of the depth map using the compensating rule is shown in Figure 5. Analysis of Figure 5(c) shows that there are no white dots in the resulting image.



Fig. 5. Correction: a – algorithm SAD; b – operator Soft-min I without correction rule; c – operator Soft-min I with correction rule

5. CONCLUSION

Analysis of modeling the process of constructing a depth map using the proposed method showed that:

1. The use of soft operators based on the analysis of RMSE increases the accuracy of the FMCDM. Thus, the fuzzy method for constructing a depth map for the three analyzed images proposed has an advantage over the SAD algorithm by 24%, compared with the prototype method by 11%.

2. When soft operators are used in the compositional rule, the accuracy of the fuzzy model increases. The best RMSE values were obtained using the following soft operators: MEAN, Soft-min I, and Soft-min II.

3. For eliminate artifacts, it is necessary to introduce a compensating rule in the structure of fuzzy inference, which limits the possible value of input of fuzzy model.

4. The smallest operating time of the fuzzy inference system was obtained using the soft MEAN operator.

5. The developed method will increase the efficiency of mobile robot control in multi-agent robotic stereo vision systems by increasing the accuracy of determining the distance to objects and reducing the decision-making time.

ACKNOWLEDGMENT

This studies was supported by the State Assignment (Agreement N_{0} 0851-2020-0032).

REFERENCES

- 1. Jorge Fuentes-Pacheco, José Ruiz-Ascencio, Juan Manuel Rendón-Mancha Visual simultaneous localization and mapping: a survey, Artificial Intelligence Review, – November 2015.
- A. Buyval, I. Afanasyev, and E. Magid Comparative analysis of ROSbased Monocular SLAM methods for indoor navigation, The 9th Int. Conf. on Machine Vision, 2016.
- 3. R. W. Wolcott, R. M. Eustice Visual Localization within LIDAR Maps for Automated Urban Driving, 2014.
- Algabri, M., Mathkour, H., Ramdane, H., and Alsulaiman, M., 2015. Comparative study of soft computing techniques for mobile robot navigation in an unknown environment. Computers in Human Behavior. 50: 42–56. doi:10.1016/j.chb.2015.03.062.
- Bobyr, M., Titov, V., Belyaev, A. 2016. Fuzzy System of Distribution of Braking Forces on the Engines of a Mobile Robot. MATEC Web of Conferences, 79, 01052. doi.org/10.1051/matecconf/20167901052.
- Delmerico, J.A., David, P., and Corso, J.J. 2013. Building facade detection, segmentation, and parameter estimation for mobile robot stereo vision. Image and Vision Computing. 31(11):841–852. doi.org/10.1016/j.imavis.2013.08.006.
- Lange, M.M., Ganebnyh, S.N., and Lange, A.M. 2017. On efficiency of the hierarchical algorithm for searching approximate nearest neighbor in a given set of images. Inform. Appl. 11(3):51-59. doi.org/10.14357/19922264170306.
- Zha, D., Jin, X., and Xiang, T. 2016. A real-time global stereomatching on FPGA. Microprocessors and Microsystems. 47: 419–428. doi.org/10.1016/j.micpro.2016.08.005.
- Pérez-Patricio, M., Aguilar-González, A., Arias-Estrada, M., Hernández-De León, H.R., Camas-Anzueto, J.L., and de Jesús Osuna-Coutiño, J.A. 2016. An FPGA stereo matching unit based on fuzzy logic. Microprocessors and Microsystems. 42:87–99. doi.org/10.1016/j.micpro.2015.10.011.
- Bobyr, M.V., Emelyanov, S.G. 2020. A nonlinear method of learning neuro-fuzzy models for dynamic control systems. Applied Soft Computing Journal. 2020, 88, 106030. doi.org/10.1016/j.asoc.2019.106030.
- 11. Georgoulas, C., and Andreadis, I. 2011. A real-time fuzzy hardware structure for disparity map computation. Journal of Real-Time Image Processing. 6(4):257-273. doi.org/10.1007/s11554-010-0157-6.
- Bobyr, M.V., Kulabukhov, S.A. 2017. Simulation of control of temperature mode in cutting area on the basis of fuzzy logic. Journal of Machinery Manufacture and Reliability, 2017, 46(3), p.288–295. doi.org/10.3103/S1052618817030049

- Bobyr, M.V., Yakushev, A.S., Dorodnykh, A.A. 2020. Fuzzy devices for cooling the cutting tool of the CNC machine implemented on FPGA. Measurement: Journal of the International Measurement Confederation, 2020, 152, 107378. doi.org/10.1016/j.measurement.2019.107378
- 14. Middlebury Stereo Datasets. Available at: http://vision.middlebury.edu/stereo/data
- 15. Gorbachev, S., Gorbacheva, N., Koynov S. 2016. A synergistic effect in the measurement of neuro-fuzzy system, MATEC Web of Conferences, 79, 01065. doi: 10.1051/matecconf/20167901065.
- 16. Gorbachev, S., Syryamkin, V. 2018. High-Performance Adaptive Neuro-Fuzzy Classifier with a Parametric Tuning, MATEC Web of Conferences, 155, 01037. doi: 10.1051/matecconf/201815501037.

Cite this article

M. V. Bobyr, N. A. Milostnaya, S. V. Gorbachev, S. Bhattacharyya and J. Cao, The Method of Depth Map Calculating Based on Soft Operators in Multi-Agent Robotic Stereo Vision Systems, In: Sandip A. Kale editor, Advanced Research in Computer Engineering, Pune: Grinrey Publications, 2021, pp. 83-98