COORDINATED USE OF VISUAL ODOMETRY AND LANDMARKS FOR NAVIGATION OF MOBILE GROUND VEHICLES.

S. M. Sokolov [1]*, N. D. Beklemishev [1], A. A. Boguslavsky [1]

[1] KIAM RAS, Keldysh Institute of Applied Mathematics, 125047 Moscow, Russia – sokolsm@list.ru, nbekl@mail.ru, anbg74@mail.ru

KEY WORDS: Autonomous Vehicles, Navigation Challenge, Visual Odometry, Landmark Movement, Interpretive Navigation, Omnidirectional Camera, Stereo.

ABSTRACT:

The paper considers two directions in the use of visual data for information support of purposeful movements of ground vehicles. This is optical odometry and navigation by landmarks in the environment. Optical odometry builds the trajectory of movement of the vehicle based on the determination of displacements based on selective visual data from different fields of view. The choice and indication of landmarks at the described stage of research remains with the operator. The vision system (VS) monitors the specified landmarks and determines the position of the vehicle relative to them. The experiments used such fields of view as monocular forward looking, panoramic (fisheye type) and forward looking stereo system. When combining the data of the visual channel with each other and with the data of other navigation systems, the specificity of visual sensors is taken into account – a significant effect of the reliability and accuracy of the results from the observation conditions. Experimental verification of the VS layout showed the achievability of high accuracy in solving the navigation problem using the visual channel. All the components of the described process of organizing purposeful movements based on the use of the visual channel continue to be improved.

1. INTRODUCTION

The main direction in the information support of autonomous mobile vehicles is the integration of data from various sensor systems, which provides stable and reliable information support for control systems. In a number of sensor systems, vision systems (VS) are becoming one of the main elements. VS have a number of properties that, on the one hand, favourably distinguish them from other sensor systems, and, on the other, require special approaches to data collection, processing and integration.

The report examines one of the important components of information support for purposeful motion – the solution of the navigation problem by collecting and processing visual data in two ways: visual odometry and tracking landmarks in the surrounding space. These methods complement each other and form an independent information channel. Combining such a channel with other navigation tools further increases the accuracy and reliability of the solution of the navigation problem. As a means of hardware support, an omnidirectional camera and a forward – looking stereo system are used in part of the recording units, and a unit based on a universal processor is used in part of the computing and control units.

2. TRADITIONAL APPROACH

When combining the readings of several sensors, it is necessary to have data on the variances of the measurement results of these sensors. The traditional approach is either to measure the dispersion of the sensor at a time before the start of movement, or to measure the average spread of its readings over a certain last time interval in the current driving conditions. Such methods are not applicable for processing visual data, the accuracy of the readings can vary significantly from frame to frame, depending on both the traffic conditions in general and the currently observed scene. In real-world conditions, images from video cameras may be distorted, blurred, dark, and contain no objects of interest. The visual channel data contains non-stationary noise, so its variance cannot be estimated as an average spread of readings, but must be estimated taking into account specific observation conditions.

Taking into account the conditions of functioning requires the inclusion in the algorithms of collecting and processing visual data of a special part for controlling external conditions and parrying, if necessary, those changes that are beyond the capabilities of the algorithms of the normal mode. This includes algorithms for controlling image histograms, including geometrized histograms (Kiy, 2018), determining the properties of textures of regions of interest (Howarth and Rueger, 2004; Kolodnikova, 2004) in the fields of view of VS. The description of these algorithms is beyond the scope of this article, but these parts are included in the algorithms used in the experiments.

3. OUR RESEARCH

In our research, we focused on collecting and processing visual data from the omnidirectional field of view and the stereo system. The omnidirectional field of view allows viewing a large area of space and choosing long-term landmarks. Stereo systems allow getting additional information about the spatial location of objects in the field of view. If necessary, the component parts of the stereo can be used as separate fields of view. Modern computer technology allows you to rectify stereo systems and use stereo data almost at the pace of a normal television scan.

At the described stage of research, the information support of ground mobile vehicles with an increased degree of autonomy

is considered. The task of purposeful movement is set by the operator and controlled in supervisory mode. The purpose of the considered module of the onboard visual channel is to provide the control system with data on the position and orientation of the mobile vehicle either in a given coordinate system or in relation to the specified landmarks. The choice of the route of movement and ensuring the safety of movements are solved in other modules of the information system.

3.1 General scheme for acquisition visual data

The general scheme for acquisition and processing visual data when solving a navigation problem is as follows. Before starting the movement, the operator indicates the starting and ending points of the route and objects that can be used as landmarks, with or without the coordinates of these objects, using images of the fields of view or on the map of the area of operation of the mobile vehicle. The VS finds the specified reference points and confirms acceptance of the task. In the process of movement, the navigation module with a certain frequency outputs to the control system data on the position and orientation of the mobile vehicle in a given coordinate system or relative to the specified landmarks.

In the process of movement, visual odometry is performed based on visual data collected from the selected areas of the video system's visual fields. Zones are selected based on general traffic conditions (urban, natural, and mixed). The most common use case is the area with the image of the underlying surface in front of the vehicle. Tracking landmarks is performed in two fields of view: omnidirectional and stereo. In the omnidirectional field of view, the VS operates with long-term (relative to the duration of movement) landmarks, which are usually located at large distances from the moving vehicle. In the stereo field of view, in its middle and upper parts, landmarks are tracked over the entire range of ranges.

As already, noted, visual data is highly variable due to external conditions. In different fields of vision, these conditions are different; the common link of visual data is a mobile platform on which the recording units are installed. Mechanical (kinematic and dynamic) characteristics of which appear in the conditions of coordination of information and motor actions (Ionova et al., 1988) and allow you to additionally filter the results of measurements of the visual channel.

3.2 Visual Odometer

The visual odometer (Scaramuzza et Fraundorfer, 2011) determines the relative displacement of the robot from frame to frame and allows you to estimate its current position by counting the coordinates; the objects of the scene are not remembered. This allows you to use it in a changing environment. Selecting and tracing landmarks, on the contrary, allows you to remember the external environment, build a map of the functioning area in accordance with the concept of interpretive navigation (Kirilchenko et al., 2008), which is one of the ways to implement the SLAM approach. The use of landmarks eliminates errors that accumulate in other navigation systems and allows you to determine the relative position of the vehicle without using traditional coordinates. Visual odometry is based on working with optical flow (Natesh, 2010).

There are two approaches to determining the optical flow: dense and sparse. In a dense optical stream, they try to track the displacement of each pixel of the scene. Sparse optical flow can

be considered as a transitional link between "pure" odometry and landmark tracking. Here, landmarks are not specifically highlighted, but the features of the scene are not formed pixel-by-pixel, but as some more stable formations consisting of a group of pixels. The application of one or another approach is chosen based on the analysis of the texture of the underlying surface (the surface of motion).

3.3 Video Channel Features

The studies were selected and analysed the main factors affecting the quality of navigation using optic canal.

For the optical odometry part:
- the place of installation of the camera on the vehicle;
- camera tilt in relation to the plane of movement;
- calibration (quality of converting pixel measurements into real camera offsets);
- selection of a zone for determining the optical flow (OP);
- frame rate / frequency of determining the displacement in relation to the speed of movement (the amount of displacement in the elements of the raster;
- taking into account 2D or 3D coordinates of features in the surrounding space;
- image contrast of the zone for calculating the optical flow;
- selection of the algorithm for calculating the optical flow (dense or sparse);
- textural characteristics of the underlying surface;
- the number and quality (in terms of highlighting) of features used by the visual odometer.

For the landmarks navigation part:
- quality rectification of stereo systems;
- the size of the stereo base and its hardness;
- distance to the landmark;
- synchronization of visual fields;
- the accuracy of fixing the landmark image;
- the number of elements of a digital image;
- viewing angle;
- the number of landmarks, taking into account the quality of detection of each;
- conditionality of the solved system of linear algebraic equations.

Another limiting factor in the use of the visual channel is the requirement to work in real time. The scale of real time is determined by the conditions of coordination of information and motor actions (IDD), which determine the time intervals allocated for the collection and processing of visual data. In a number of navigation tasks, some relaxation of requirements is possible here, due to a controlled maneuver by the speed of movement, but in general, the requirements for the necessary accuracy and speed of solving the navigation problem impose additional restrictions on processing algorithms.

3.4 Combining visual channel data with each other and with data from other navigation sensors

The algorithm of information support of the moving means, there are several conditions of approval, binding the function parameters of the visual sensor with the motion parameters of the robot (speed, parameters of plot path, etc.). In particular, the condition of continuity of the review is to require some overlap

of the beginning of the next region review end of the previous field review. This condition determines the maximum speed of movement at a given speed of inspection of the terrain ahead of the vehicle and the period of inspection.

When combining the results of measurements based on visual data processing and combining these measurements with data from other sensory systems, these features must be taken into account. The paper uses an assessment of the accuracy of each individual measurement of the visual channel reading based on the analysis of errors that occur at all stages of visual data collection and processing. After obtaining such estimates, which replace the values of the variances of traditional meters, for a set of measurements of the visual channel, the weighted average is calculated, where the readings with the greatest reliability and accuracy have the greatest weight.
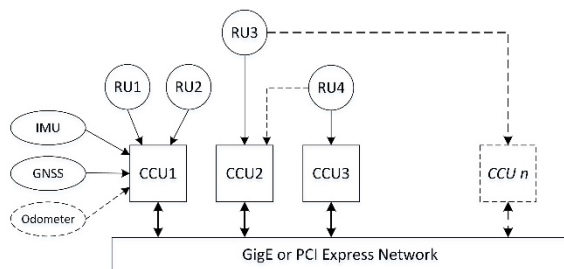

Figure 1. The general scheme of VS hardware and traditional information systems combination for mobile robots information support. RU – recording unit (video camera) with a certain field of view; CCU – computing control unit; GNSS – global navigation satellite system.

We use one more technique to improve (improve the quality and reliability) of the results in determining the trajectory of the vehicle on which the visual sensors are installed. This technique is as follows. As a unifying principle for all data, we take the trajectory of the moving RTK. When constructing this trajectory, we take into account such factors as the mechanical characteristics of the mobile RTK and the likely need to build a return path (solving the problem of returning to the starting point in an autonomous mode). The algorithm for describing the trajectory of a mobile robot is based on the approximation of the data of the integrated navigation systems in the form of a piecewise polynomial function (spline) of the 3rd order by the least squares method and checking the resulting spline for the absence of loops (Sprunk, 2008).

The hardware part of the architecture of the onboard VS is formed by reconfigurable combinations (network) of recording (RU) and computational and control units (VUB), and the software part is a large-scale software framework VS RT (Boguslavsky et al., 2019). In fig. 1 shows the general layout of the VS.

4. EXPERIMENTS RESULTS

The described approaches and corresponding algorithms were implemented within the framework of the real-time VS software framework (Sokolov and Boguslavsky, 2016) and tested experimentally at the test site, using the movement of the VS on a mobile platform along various routes. In fig. 2 shows a general view of a car with VS units installed on it. In fig. 3 examples of fields of view of the recording units.

Figure 2. Mono, stereo forward-looking and omnidirectional recording units installed on the roof of the Niva car.

(a)

(b)

(c)

Figure 3. Examples of fields of view of the recording units: (a) mono camera looking into a given area of space; (b) an omnidirectional camera with a fisheye lens; (c) forward looking stereo system.

As areas of images for the optical flow estimation the regions of interest in the VS fields of view of the recording units were selected (fig. 4-6).
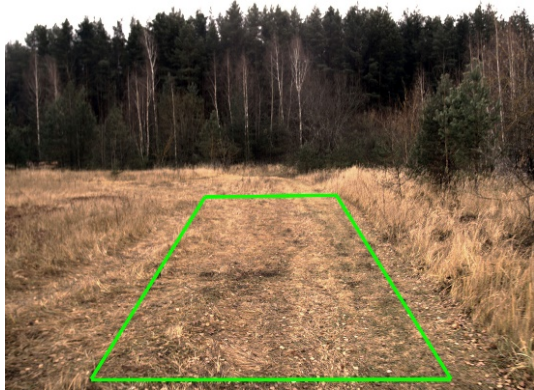

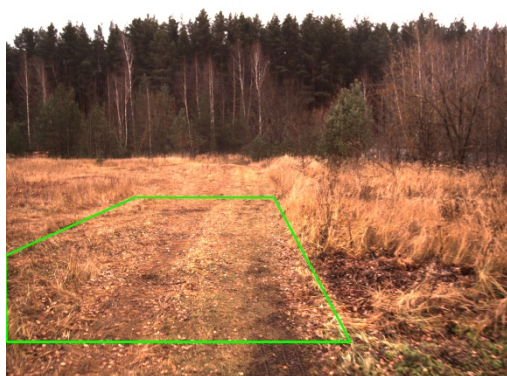**Figure 4**. Area of interest for calculating the optical flow of a monocular block.

Figure 5. of the stereo unit (one frame of the stereo system is shown).
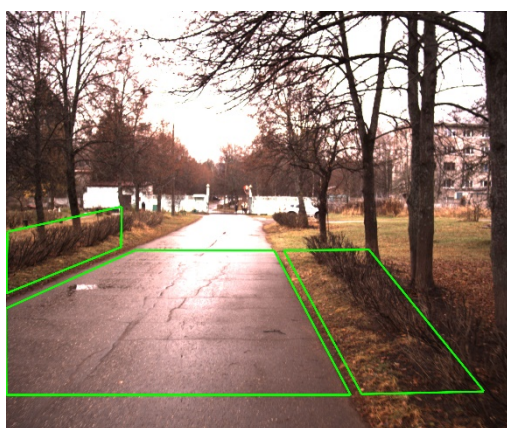

Figure 6. Additional areas of interest for calculating the optical flow in the case of a homogeneous underlying surface (a single frame of the stereo system is shown).

As part of the VS, an omnidirectional camera with a 5 Mp color matrix, a fisheye lens with a 360° viewing angle around the vertical axis and 90° relative to the horizon, and a stereo system consisting of two synchronized 5 Mp color cameras with 60° lenses were used in the experiments. An integrated GPS receiver combining satellite and inertial navigation systems was used as an additional means of arbitrarily recording movements (SBG Systems, 2021).

At the current stage of research, the choice of benchmarks remains with the operator performing supervisory control. VS provides tracking of the specified landmarks and the solution of the navigation problem. Building a map of the area and determining its position on it (the so-called SLAM method (Mur-Artal and Tardós, 2014)) is performed based on the concept of interpretive navigation (IN) (Kirilchenko et al., 2008). Fig. 7 shows the objects selected as landmarks for determining the position of the vehicle on the ground.
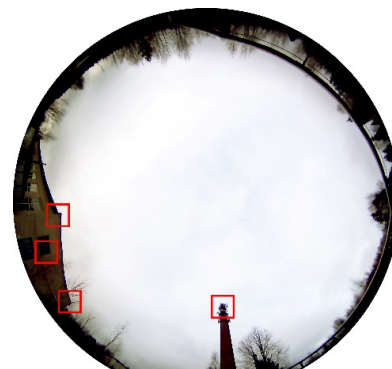
(a)
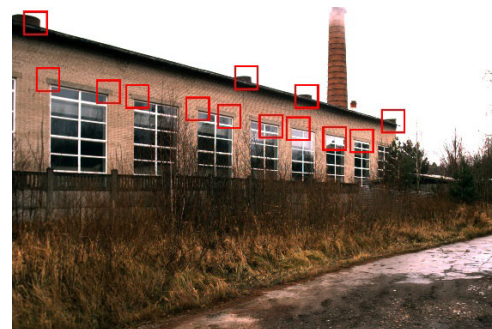
(b)

Figure 7. Examples of images of the registering blocks of the VS layout. (a) Omnidirectional / panoramic cameras; (b) Stereo system (a frame from one of the cameras is shown). Objects that can be used as landmarks are marked with red squares.

Experiments have shown the effectiveness of the described approach, high accuracy of determining the position (up to 1% of the distance traveled), orientation (up to 0.3 angular degrees) and other parameters that characterize the relative movement of the vehicle was obtained. Naturally, the described combined visual channel can be combined with other means of navigation: inertial, satellite, wheel odometry, which makes it possible to increase the reliability of solving the navigation problem in a wider class of conditions.

5. CONCLUSIONS

The considered organization of information support for purposeful movements of mobile vehicles using the visual

channel in model assessments and experiments has shown high efficiency. The algorithm for determining the trajectory of a moving vehicle based on the optical flow takes into account a large number of factors that affect the reliability of the result and uses different areas of several fields of view.

All the components of the described process of organizing purposeful movements based on the use of the visual channel continue to be improved. The database of objects that can be selected as landmarks is being expanded, and algorithms for automatically selecting such objects are being developed. A notable feature of the described algorithms is their implementation in the format of a unified modular software framework for real-time vision systems. This implementation allows you to quickly build software to solve new problems of information support based on the use of VS.

REFERENCES

Boguslavsky, A.A. et al. 2019: *Models and algorithms for intelligent control systems*. KIAM RAS, Moscow.

Howarth, P., Rueger S., 2004: Evaluating Texture Characteristics for Content-Based Image Search. *Proc. of the International Conference on Image and Video Search*, 326-334.

Ionova, J.N., Kiril'chenko, A.A., Pavlovsky, V.E., Platonov, A.K., Pryanichnikov, V.E., 1988: Conditions for coordination of the information and motion activities of mobile robots. *Preprints of the 3rd IFAC symposium Intelligent Autonomous Vehicles*, Madrid, Spain, March 25-27.

Kirilchenko, A.A., Platonov, A.K., Sokolov, S.M., 2008: Theoretical aspects of the organization of interpretive the navigation of a mobile robot. *KIAM RAS Preprint No. 19*.

Kiy, K.I., 2018: Image Understanding Systems Based on the Geometrized Histograms Method. *Extreme Robotics*, 1(1), 417-423.

Kolodnikova, N.V., 2004 : An overview of texture features for image recognition tasks. TUSUR Reports, Automated systems of information processing, control and design, 113-125.

Mur-Artal, R., Tardós, J., 2014. ORB-SLAM: Tracking and mapping recognizable features. *MVIGRO Workshop at Robotics Science and Systems (RSS)*, Berkeley, USA.

Natesh, S., 2010: Feature Based Landmark Extraction for Real Time Visual SLAM. *Proceedings of the 2010 International Conference on Advances in Recent Technologies in Communication and Computing*. 16-17 Oct.

SBG Systems, 2021. Ellipse Series. Miniature inertial sensors. https://www.sbg-systems.com/products/ellipse-series/

Scaramuzza, D., Fraundorfer, F., 2011: Visual Odometry Part I: The first 30 years. *Robotics & Automation Magazine*, 18(4), 80-92.

Sokolov, S.M., Boguslavsky, A.A., 2016: Methodological aspects for the development of information systems of unmanned mobile vehicles. *Proc. ICINCO16*, vol. 2, 492-498.

Sprunk, C., 2008. Planning Motion Trajectories for Mobile Robots Using Splines. Report, Faculty of Applied Sciences, Albert-Ludwigs-Universität Freiburg, Germany.