

Visual-inertial odometry algorithms on the base of thermal camera

A P Alekseev^{1,2}, E V Goshin^{1,2}, N S Davydov^{1,2}, N A Ivliev^{1,2} and A V Nikonorov^{1,2}

¹Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

²Samara National Research University, Moskovskoe Shosse, 34A, Samara, Russia, 443086

e-mail: goshine@yandex.ru, artniko@gmail.com

Abstract. A trajectory building based on a camera data is one of the most popular tasks in the field of machine vision. In particular, this task appears when it is necessary to navigate in the absence of signals from global navigation systems such as GLONASS and GPS. In this work, study of existing methods of visual odometry for the flight trajectory restoration by shooting an infrared camera of the thermal range were conducted. To improve the accuracy, it is proposed to use the data from inertial sensors. As a result, it is shown that the proposed solution allows to successfully solve the problem of trajectory reconstruction.

1. Introduction

Nowadays, there is a need to create methods for building navigation systems based on the usage of video data, which would be based on cheap off-the-shelf hardware and software. One of the fundamental tasks in the field of mobile robots and unmanned vehicles is the localization of the object or vehicle. Along with existing systems, such as pulsed laser systems (e.g. LiDAR [2, 7], IMU [5], GPS, radar [6]), visual odometry methods are of great interest. These methods use the video stream of the camera installed on the object. Due to its low cost in comparison to most other technical means and the availability of algorithms capable of qualitatively converting photometric information into location information, this approach turns out to be quite promising. Of course, this method has disadvantages. Poor illumination of the scene can adversely affect the assessment of movement, besides, it is necessary to dominate the environment of static objects for better correct comparison.

Visual odometry in the case of infrared (IR) shooting has additional difficulties. The low contrast of the image from the infrared camera make it necessary to do an additional processing, which affects its speed. Sharpness of objects in the IR range is low. In addition, there are fundamental geometric constraints for determining the precise rotation and movement of a camera through images. For many systems, however, this is the most promising approach, its combinations with the use of other additional sensors (LiDAR, IMU, etc.) are not uncommon. Modern researches are far advanced in easing restrictions for the applicability of this method. Such researches are based on several 3D reconstruction techniques, which are widely useful in other applications e.g. [8], [9]. This paper shows the result of testing modern approaches of visual odometry in relation to data from a monocular IR camera.

In this article, a comparison of the existing approaches to the trajectory building with the use of a video sequence from an IR camera is made, the optimal approach has been chosen and a conclusion

has been made about its applicability for solving the problem of visual odometry. An approach to improving the accuracy of IR odometry through the use of inertial sensor readings is also proposed.

2. Monocular visual inertial odometry methods

Most monocular visual odometry algorithms for small UAVs are based on the technology of parallel tracking and mapping of PTAM (Parallel Tracking and Mapping) [3]. The PTAM technology is based on the method of simultaneous localization and mapping of the SLAM (simultaneous localization and mapping) map. SLAM provides reliability by tracking and displaying hundreds of control points and it works in real time. A special feature is the simultaneous execution of the task of displaying destinations and estimating the displacement, taking into account a rather effective correction based on processing images from different viewing angles. PTAM technology was developed for augmented reality applications in small spaces. There are several modifications of it, for example, with a limited number of personnel, which ensures its full-fledged work in the conditions of spaces with urban-type buildings [1].

According to research [4, 8]:

Nowadays the following methods are applied:

- single chamber (monoVO)
- multichamber (stereoVO, RGBD-vo)

Methods using data from one camera:

Direct:

- direct distributed (DSO - direct sparse odometry)
- semidirect (SVO - semi-direct odometry)

Indirect:

- method of key points (feature-base visual odometry)
- half-line

A comparison of the approaches is presented in Table 1, the details are analyzed in detail in [4, 8].

Table 1. The comparison of approaches to the construction of the trajectory.

Feature-based	Direct
Use only features (for example, angular)	Use the whole image
Faster	Slower (but can be well parallelized)
Easy to remove noise (outliers)	Not easy to remove noise
Resistant to rolling shutter	Not resistant to rolling shutter
Use a small amount of information from the images	Use more complete information
Do not require complex initialization	Require complex initialization
Sensitive to image intensity distortions	Less sensitive to image intensity distortions
Over 20 years of intensive development	About 4 years of research

All monocular algorithms have a set of similar requirements and limitations:

1. The need for accurate camera calibration. Less critical for feature-based algorithms.
2. The inability to determine the scale without the help of external sensors or user.
3. Camera requirements: high shooting speed and wide viewing angle. These parameters are associated with each other, and with the maximum speed of movement of the camera.

For the solvable problem of odometry using IR video streams, some of the drawbacks of indirect methods are critical, for example, when IR shooting, the sensitivity of the method to changes in intensity is essential. That is why the choice was made in favor of the algorithm DSO.

2.1. Description of the direct visual odometry method

The semi-direct odometry model is associated with a dense model, however, it will be shown later that sparse depth information is sufficient to obtain a rough estimate of the motion and the search function of control points [4]. As soon as the control points and the initial position of the camera are found, the

algorithm will use only control points. This explains its name - “semidirect”. This technique allows you to quickly set the processing frame in the new image. The Bayesian filter, which explicitly removes erroneous measurements, also evaluates the depth and positions of the control points. That is, a point is plotted on a three-dimensional map only when the accompanying depth filter has given convergence, which requires multiple measurements. The result of the work is a three-dimensional map with control points, whose reliability has been verified

The speed of the DSO algorithm significantly increases if the control points are not extracted from all frames, but only the key ones. Accuracy is also increased when using the sub-pixel checkpoint extraction function. Unlike direct methods, many small sections are used here, rather than several large flat sections, which positively affects the speed of the algorithm and its reliability.

As part of the project, the task was to form the trajectory of the infrared camera with reference to the objects of the terrain. The first stage is the assessment of existing approaches for solving this problem in the visible range. Despite the various methods of counting, common to all methods is the conversion of the processed image into shades of gray. Further actions take place with a monochromatic image.

Direct sparse odometry is based on continuous optimization of the photometric error over the last frame window, taking into account the photometrically calibrated imaging model. Unlike existing direct methods, all the involved parameters are jointly optimized here (built-in cameras, external cameras and inverse depth values), effectively performing the photometric equivalent of adjusting a window sparse beam. The geometric representation used by other direct approaches is preserved, that is, the points are represented as the inverse depth in the frame of reference (and, thus, have the same degree of freedom).

2.2. Visual inertial odometry

This paper proposes a modification of the DSO, which allows to take into account the data of inertial sensors when restoring the trajectory. According to [4], the key procedure in the DSO algorithm is to minimize the photometric error E_{pj} between the neighborhood of a point \mathbf{p} in the reference frame I_i and the neighborhood of the corresponding point \mathbf{p}' in the target frame I_j :

$$E_{pj} = \sum_{p \in N_p} w_p \left\| \left(I_j(\mathbf{p}') - b_j \right) - \frac{t_j e^{a_j}}{t_i e^{a_i}} \left(I_i(\mathbf{p}) - b_i \right) \right\|_{\gamma}. \quad (1)$$

Here, N_p neighborhood of the point where minimization is performed, $t_j e^{a_j}$ and $t_i e^{a_i}$ describe the exposure of the relevant frames, $\| \cdot \|_{\gamma}$ Huber rate, w_p keyframe weight.

In the original DSO point \mathbf{p}' is found as a back projection of a point \mathbf{p} in the reference frame:

$$\mathbf{p}' = \Pi_c \left(\mathbf{R} \Pi_c^{-1}(\mathbf{p}, d_p) + \mathbf{t} \right), \quad (2)$$

where d_p inverse point depth, \mathbf{R} and \mathbf{t} rotation matrix and translation vector.

In this work, it is proposed to replace the rotation matrix and translation vector in formula (2) with their estimates obtained using inertial sensors. The results of experiments showed, that a visual-inertial modification of the DSO algorithm allowed to significantly increase the accuracy of visual odometry.

3. Results

3.1. IR Camera Calibration

For successful calculations and trajectory construction, it is necessary to calibrate the camera. For cameras of the visual range there are special scales (Figure 1).

The infrared camera requires the same calibration board with contrast fields. After a series of experiments, the substrate material was selected, the printing method and the ink composition for the

areas absorbing infrared radiation, and a coating with maximum contrast was found by selecting materials. Figure 1 on the right shows the finished layout of the board for calibrating the IR camera.

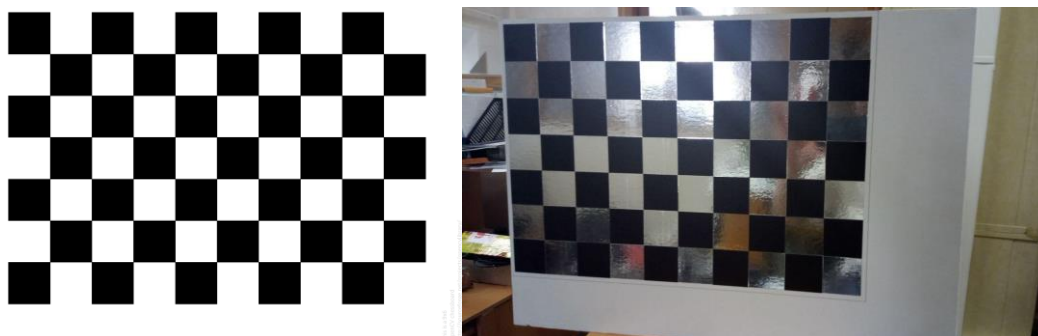


Figure 1. Layout of the calibration scale.

3.2. Experiment

To check the quality of the developed algorithms for visual inertial odometry, a full-scale survey was conducted from a hydroplane using two cameras, an IR and a visible range. As the infrared camera used camera thermal range COX 1000 with a resolution of 1024 * 768. A suspension was made for the field survey (Figure 2). For the control shooting used camera GoPro3. The pitch of both cameras is the same. The platform was fixed on the front wing of the wing to eliminate the structural elements of the aircraft from falling into the frame (Figure 2).

To control the accuracy of the constructed trajectory during the flight, the GPS coordinates were recorded using a mobile navigator. The trajectory of the flight experiment built on GPS is shown in Figure 3. The inertial sensor readings were taken from a mobile phone.



Figure 2. COX-1000 IR camera and its mounting on the same platform with GoPro box.

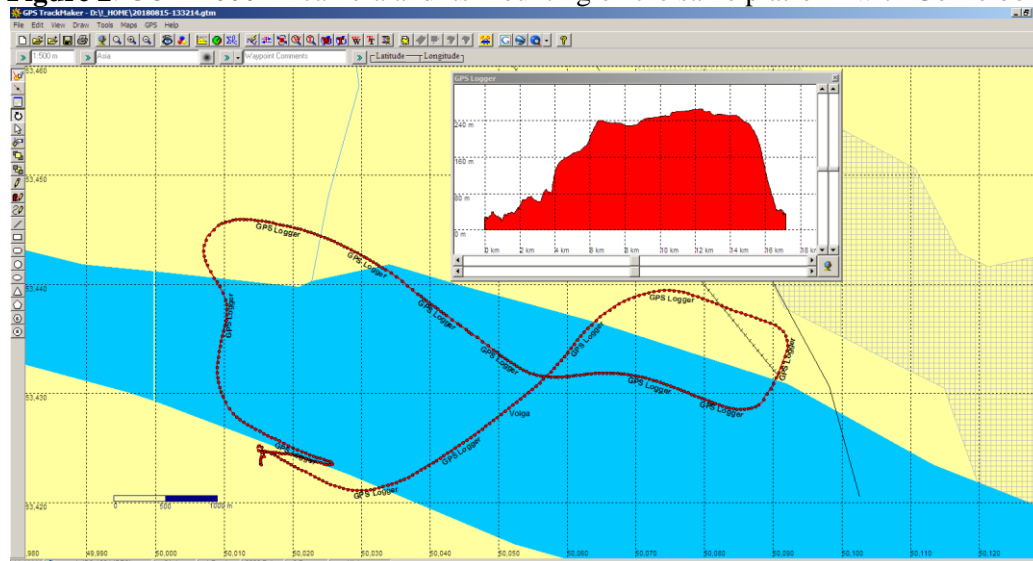


Figure 3. The trajectory and altitude map of the flight experiment recorded by GPS.

3.3. Accuracy of trajectory recovery using visual inertial odometry methods

In some areas of the flight, it was possible to build a fairly accurate flight trajectory. This selectivity is due to the characteristics of the infrared camera matrix, as well as the presence in the frame of fairly large areas with a uniform surface (water, forest). On such sites, the restoration of the trajectory is difficult. Examples of frames where trajectory recovery was completed successfully are shown in Figure 4.

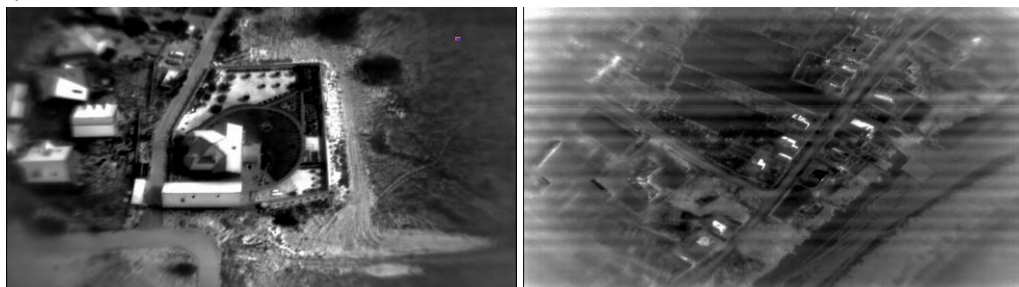


Figure 4. Examples of frames in video stream.

The proposed adjustment of the visual odometry algorithm through the use of inertial sensor data has significantly improved the accuracy of trajectory recovery. Figure 5 shows a comparison of the obtained trajectory sections with GPS data. The average deviation in the studied areas did not exceed 20 meters, which confirms the efficiency of the proposed approach.

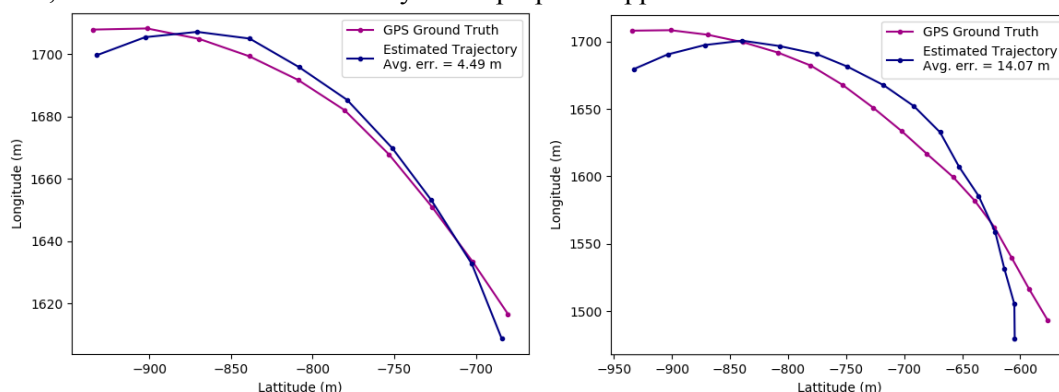


Figure 5. Comparison the trajectory reconstructed by visual odometry with trajectory recorded by GPS.

4. Conclusion

In this work, direct and indirect visual odometry methods for the problem of trajectory recovery from a sequence of frames taken in the thermal infrared range are investigated. For the experimental study, the IR camera was calibrated, a test bench was built, a test laboratory was assembled, full-scale video recordings were carried out with a camera installed on a small aircraft, under various lighting conditions and in different temperature conditions. The methods of direct visual odometry showed acceptable results on the quality of the hydroplane trajectory recovery in comparison with the GPS data. The modification proposed in this paper, which makes it possible to use inertial sensor data to clarify visual odometry, has provided a significant increase in the trajectory recovery accuracy.

5. References

- [1] Baker S, Matthews I 2004 Lucas-kanade 20 years on: A unifying framework *International journal of computer vision* **56(3)** 221-255
- [2] Cole D M, Newman P M 2006 Using laser range data for 3D SLAM in outdoor environments *Robotics and Automation, ICRA Proceedings IEEE International Conference* **1** 1556 DOI 10.1109/ROBOT.2006.1641929

- [3] Engel J, Schöps T and Cremers D 2014 LSD-SLAM: Large-scale direct monocular SLAM *Computer Vision – ECCV* **1** 834-849
- [4] Engel J, Koltun V and Cremers D 2018 Direct Sparse Odometry *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40(3)** 611 DOI 10.1109/TPAMI.2017.2658577
- [5] Yi J, Zhang J, Song D and Jayasuriya S 2007 IMU-based localization and slip estimation for skid-steered mobile robots *Intelligent Robots and Systems, IROS IEEE/RSJ International Conference* **6** DOI: 10.1109/IROS.2007.4399477
- [6] Checchin P, Gérossier F, Blanc Ch, Chapuis R and Trassoudaine L 2010 Radar scan matching SLAM using the Fourier-Mellin transform *Field and Service Robotics* **1** 151-161 DOI: 10.1007/978-3-642-13408-1_14
- [7] Hess W, Kohler D, Rapp H and Andor D 2016 Real-time loop closure in 2D LIDAR SLAM *Robotics and Automation (ICRA) IEEE International Conference* **1** 1271 DOI: 10.1109/ICRA.2016.7487258
- [8] Fursov V A, Goshin Ye V and Kotov A P 2016 The hybrid CPU/GPU implementation of the computational procedure for digital terrain models generation from satellite images *Computer Optics* **40(5)** 721-728 DOI: 10.18287/2412-6179-2016-40-5-721-728
- [9] Smelkina N A, Kosarev R N, Nikonorov A V, Bairikov I M, Ryabov K N, Avdeev A V and Kazanskiy N L 2017 Reconstruction of anatomical structures using statistical shape modeling *Computer Optics* **41(6)** 897-904 DOI: 10.18287/2412-6179-2017-41-6-897-904

Acknowledgment

Methods and algorithms were designed and developed with the support of RFBR grants (projects 19-29-01235-mk, 16-29-11744-ofi_m, № 16-29-09528- ofi_m, № 17-29-03112- ofi_m, № 18-07-01390-A, № 18-37-00457-mol_a), optics and experimental studies - in the framework of the state assignment of the IPSI RAS - a branch of the Federal Scientific-Research Center "Crystallography and Photonics" of the RAS (agreement № 007-Г3/Ч3363/26).