Comparative Analysis of Monocular Visual Odometry Methods for Indoor Navigation

E.O. Trubakov[1], O.R. Trubakova[1]
trubakoveo@gmail.com| trubakovaor@gmail.com
[1]Bryansk State Technical University, Bryansk, Russia;

The monocular visual odometry algorithm involves several basic steps and for each of them there is a number of methods. The purpose of this work is to conduct practical research of methods for key point detection and the optical flow calculation in the problem of determining the unmanned ground vehicle proper motion. For detection method research conduction the image panel containing various distortions typical for follow robot shot was made. It has been educed that among the accounted methods FAST finds the largest number of points within the minimum elapsed time. At the same time image distortions strongly affect the results of the method, which is negative for the positioning of the robot. Therefore the Shi-Tomasi method was chosen for key points detection within a short period of time, because its results are less dependent on distortion. For research undertake a number of video clips by means of the follow robot shot was made in a confined space at a later scale of the odometry algorithm. From experimental observations the conclusions concerning the application of Lucas-Kanade optical flow method tracking the identified points on the video sequence have been made. Based on the error in the results obtained it was implication that monocular odometry cannot be the main method of an unmanned vehicle positioning in confined spaces, but in combination with probe data given by assistive sensors it is quite possible to use it for determining the robotic system position.

Keywords: *positioning, monocular visual odometry, image key points, detector, FAST, Harris, ORB, Shi-Tomasi, optical flow, Lucas-Kanade, Farneback.*

1. Introduction

One of the fundamental tasks in the field of mobile robots and unmanned vehicles is the localization of the object and the construction of the surrounding area map. There are many approaches to solve this problem using various technical means, for example, laser systems such as LiDAR [3, 7], IMU [12], GPS, radar [2].

All these systems have various disadvantages (for example, high cost of equipment in LiDAR or data precision in GPS). This article will consider the method of visual odometry as a method of a moving object positioning. A noteworthy detail is that the main evils of this method is the dependence upon external factors, for example, the illumination of the studied room directly affects the level of the obtained data precision.

And for the correct objects matching in progressive photo images the dominating in the midst of static objects is required. In addition, there are other problems, for example, geometric limitations when specifying the exact rotation and camera movement through the images (without additional information from other sensors it is impossible to determine the displacement scale) [8].

Despite all this, for many mobile systems, this method suits.

The main point of visual odometry is to analyze the progressiveness of photos taken by the robot's camera. Through the objects position change in images, the repositioning of the robot over a distance is determined.

The monocular visual odometry algorithm consists of several steps. After receiving the image, the first step is to key points finding in the image. To implement this stage of odometry there is a number of methods. This paper analyzes some of the commonest methods, such as Harris [6], Shi-Tomasi [11], FAST [4], ORB [10].

The next stage of the algorithm is the optical flow achieving. Ad hoc studies were conducted to choose between the method of optical flow Lucase-Kanade [9] and the method of dense optical flow Farneback [5] (as one of the most effective and well-proven).

The final stage of the algorithm is to determine the motion of the camera based on the data obtained from the optical flow, i.e. the rotation matrix and displacement vector calculation. The paper investigates the methods of key points identification and the optical flow computation in the problem of proper motion detection.

2. Key points finding

For the detection of object movement it is necessary to identify changes in the sequential image set. To fulfill this requirement it's reasonable to find those points on the image that are steadily different from other points (key points) and check their displacement in subsequent images.

One of the methods of identification (detection) of key points in the image is the Harris method. It is based on considering the derivatives of image brightness variation to make the tracking of brightness in all directions possible. The principle of the method is that for the image I a window W is offered, which is dependent on the size of the image (the most commonly used window size 5x5) centered at (x, y), having its shift to (u, v).

The sum of the squared difference between the initial and shifted window is:

$$E(u,v) = \sum_{(x,y)\in W} w(x,y)\left(I(x+u,y+v) - I(x,y)\right)^2 \approx$$

$$\approx \sum_{(x,y)\in W} w(x,y)\left(I_x(x,y)u + I_y(x,y)v\right)^2 \approx (x\ y)M\begin{pmatrix}x\\y\end{pmatrix},$$

where $w(x,y)$ is a weight function (usually Gauss function or binary window is used); M is autocorrelation matrix:

$$M = \sum_{(u,v)\in W} w(u,v)\begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix}.$$

On change of the function $E(x,y)$ at a large scale in the line of (x,y) modulo large eigenvalues of the matrix are obtained M, which is quite time-consuming. Therefore a response measure parameter was created:

$$R = detM - k(trM)^2 > k,$$

where k is the empirical constant ($k \in [0{,}04; 0{,}06]$).

In this case the value R is positive for angle points. Then the points having the value R less than the prescribed threshold value are removed from the set of points. Further the local maxima of the function R are calculated in the neighborhood of the given radius and the obtained points are chosen as angle key points.

There is another detector that is similar in algorithm to the Harris detector. It is called the angle detector of Shi-Tomasi. The difference between the Shi-Tomasi detector and the Harris detector is in the calculation of the response measure:

$$R = min(\lambda_1, \lambda_2),$$

where λ_1 and λ_2 are the eigenvalues of M.

This method calculates the eigenvalues directly, as the search for angles will be more stable. Essentially, it is necessary to set a threshold value and if the calculated value is higher than the threshold, the point is considered as an angle, in other words, an interest point.

The described methods identify key points by applying their algorithms directly to the pixels of the initial image. There is an alternative approach, which involves using of computer-aided learning algorithms training a point classifier within a set of images. The FAST method makes decision trees for pixel classification.

For each pixel p, a circle of radius 4 inscribed in a square area with a side of 7 pixels is introduced. On the basis of this set of points the conclusion about whether the starting point is the key point or not is made.

The circle passes through 16 pixels. Each of the pixels of the circle referring to the pixel p can be in one of three states:

$$S_{p\to x} = \begin{cases} d, & I_x \le I_p - t \text{ (darker color)} \\ s, & I_p - t \le I_x \le I_p + t \text{ (same)} \\ b, & I_p + t \le I_x \text{ (lighter color)} \end{cases}$$

For each x and obtained $S_{p\to x}$ for each $p \in P$ (P is the set of all pixels of the training set of images) the set P is divided into 3 subsets of points P_d, P_s, P_b that are darker, same, or lighter than the point x respectively. Then the decision tree is built. According to the results of the tree the angles on the test images are determined.

The main disadvantage of the FAST method is the order of selecting points and this affects the efficiency. It is also worth taking into account the fact that in the environment of the starting point may occur other key points and in this case the method can give erroneous results.

Another method of finding key points is the ORB method [10]. Its algorithm is as follows:

1. 1. Key points are detected using a fast tree-like FAST in the initial image and in several images from the thumbai pyramid.
2. 2. The Harris measure is calculated for the detected points. Try outs with a low Harris measure value are neglected.
3. 3. The orientation angle of the key point is calculated. In this regard the luminance elements for the key point neighborhood are calculated:

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y),$$

where x, y are pixel coordinates, I is brightness.

And then the orientation angle of the key point is calculated:

$$\theta = \text{atan2}(m_{01}, m_{10}).$$

The result is a particular direction for the neighborhood of the key point.

4. 4. Having the orientation angle of the key point, the sequence of points for binary comparisons in the BRIEF [1] descriptor rotates according to this angle.

Mathematically the new positions for the points of binary tests are calculated as follows:

$$\begin{pmatrix} x_i' \\ y_i' \end{pmatrix} = R(\theta) * \begin{pmatrix} x_i \\ y_i \end{pmatrix}.$$

5. 5. The binary descriptor BRIEF is calculated from the obtained points.

3. Case Studies

All studies were conducted on the basis of a robotic device, which is a four-wheeled unmanned vehicle. The system is implemented on the basis of the Raspberry Pi 3 microcomputer. On this microcomputer a four BCM2837 core processor having 1200 MHz and 1 GB SDRAM is installed. The methods have been implemented in the C++ programming language under the Ubuntu operating system.

3.1. Analysis of key points detection methods

To study the methods of detecting key points, a number of photos were taken with a wide-angle camera installed on the Waveshare robot. The photos were presented in three main resolutions: 400x250, 800x600, 1920x1080. The images were taken in different resolutions to check the dependence of the methods on scaling. Besides it pays in both: determining the best resolution for the assembled device and choosing the detection method upon the criterion of speed.

The photos were edited to test the methods ' susceptibility to various distortions (darkening, rotating, blurring and with added noise). A darkened image may occur due to changes in the illumination of the space during the robot's movement. The rotated image can be caused by the device movement in the process of shooting when it travels along uneven surfaces.

A blurred image may appear because the photo was taken while moving and the boundaries of the objects in the photo may be blurred. Noise can happen due to changes in light, image compression, or natural features such as the appearance of a dust cloud in the area. In studies with distorted images the identified key points have been counted and the percentage of point losses on images without modification has been determined. These experiments have revealed the methods, which are least susceptible to image distortion.

This FAST method was chosen as the first method to be studied was implemented in two forms. The first did not have the maximum suppression of the average shift algorithm, and the second was absolutely devoid of it. For the first alternative of the FAST method about 1900 key points were found in the image with a resolution of 400x250 pixels within 0.14 seconds. For photos with 800x600 pixels about 8800 points within 0.66 seconds were detected. In the image with a resolution of 1920x1080 pixels about 27400 key points were found within 2.32 seconds. The time taken to revelation of key points is directly proportional to the number of these points.

The second implementation of the method finds the number of key points several times more, while the time spent on their search increases less than twice. For example, in an image with a resolution of 400x250 pixels, about 7700 key points were found within 0.3 seconds. That means that the first implementation of FAST method recognized the number of key points 4 times less than the second one, spending time half as much.

The next implemented detector was Harris method.In the image with a resolution of 400x250 pixels, about 250 key points were found within 0.11 seconds. Since the Harris detector is an angle detector, it detected key points neither on the circumference nor on the edges of objects. In the photo with a resolution of 800x600 pixels there were about 600 points detected within 0.56 seconds. On the image with a resolution of 1920x1080 pixels, about 1080 key points were found within 2.25 seconds. It can be noted that unlike previous detectors, this method finds fewer key points within a shorter period of time and that resulted in the increasing of accuracy of this method to be used in the future.

In the ORB algorithm the maximum number of key points by default cannot be more than 500, if it is, then Harris angle detector is applied for excluding the least significant ones. In this regard, the algorithm gave the following results: in the image with a resolution of 400x250 pixels about 470 key points were found within 0.24 seconds; in the image with a resolution of 800x600 pixels 500 key points were found within 0.83 seconds; in the image with a resolution of 1920x1080 pixels 500 key points were found within 2.98 seconds. Experiments have shown that this detector is not the fastest among the described above.

The Shi-Tomasi detector is based on the Harris detector. But in spite of this fact, the Shi-Tomasi algorithm detected only 25 key points in images with different resolutions within the same period of time, which is 10 times less than the Harris detector did in these images.

The second type of experiments was carried out to analyze algorithms connected with image distortion. As a result the percentage of point reduction comparing to the initial image having a resolution of 1920x1080 pixels became apparent.

FAST detectors turned out to be particularly instable in darkening. These algorithms lose about 65-70% of key points when the image is distorted. With a small rotation of the image (20 degrees) these detectors lose about 48-53% of points. FAST detectors proved to be also instable in image blurring. In this case the algorithm having not maximum suppression loses more key points (89%) than the second implementation of this method (58%). The FAST method flounders in case of noisy images and is noninvariant to the appearance of noise (these algorithms find 82-88% more key points, which is erroneous).

Harris detector showed invariance to distortions such as darkening, blurring and noise, but it was instable in case of image rotation (this method loses about 36% of the points). It can't typify the Shi-Tomasi method, which is based on the Harris detector, but unlike Harris it is independent on rotating like the Orb detector as well. During the experiments it was found out that the main advantage of Shi-Tomasi and ORB detectors is invariance with respect to noise, blurring and darkening.

3.2. Analysis of optical flow construction methods

The following type of research was carried out with a group of video sequences having spans of 288 frames and 504 frames. Each group was divided into three options depending on the video resolution: 800x480, 640x360 and 320x240. These studies were conducted to find the method of the optical flow constructing for its further use in the algorithm of robot's visual odometry. The Lucas-Kanade method and the Farneback method of dense optical flow were chosen to be studied.

The Lucas-Kanade optical flow method was chosen first for that purpose. When the video resolution is 320x240, the working time of the method is approximately 0.07 seconds per frame. When the video resolution is increased, the operation time increases too. So when the resolution is 640x360 the frame is processed within about 0.11 seconds, and in case of 800x480 resolution the frame processing time is already about 0.16 seconds. Studies have shown that the method does not depend on the span of the video sequences, but depends on its extension.

The next implemented method connected with optical flow was the Farneback dense optical flow method. This method like the Lucas-Kanade method showed the best working time at a lower resolution (320x240). This behavior is reasoned because the method deals with a smaller area of the image. When implementing this method for video with a resolution of 320x240, the frame processing time was about 0.42 seconds per frame. With a frame resolution of 640x360, the runtime is approximately 1.1 seconds per frame. And in case of 800x480 resolution this speed is reduced to 1.9 seconds per frame.

3.3. Experimental proof of detection method effect on the determination of the proper motion

The final studies were conducted on two groups of video sequences. The first type of video lies in the equable movement of the robot forward and then back having the final point of the movement coinciding with the initial one. The second type of motion is a sustainable closed circular motion in which the final point of motion coincides with the initial point.

The trajectory calculated on the basis of visual odometry should be linear forward and backward in the first case and it is to describe a closed trajectory in the second case. Due to the fact that the initial points coincide with the final ones, it is no need taking into consideration the displacement scale received from other sensors, but it's quite sufficient to use the data received from the camera.

The first method chosen for these studies was Shi-Tomasi. When moving forward by 2.5 meters, and then moving back by 2.5 meters in different situations, the error of reaching the initial point in the motion diagram is about 6.9%. In a circular motion with a radius of 0.4 m and a robot speed of 0.2 m / s, the trajectory calculated on the basis of visual odometry has an error of about 20.7%.

Studies have proved both: the sensitivity of visual odometry calculation method and the err of calculations when using the Shi-Tomasi method of key points detection.

The next method to be investigated was the Harris method. When moving back and forth using the same videos that in situations with the Shi-Tomasi method, the trajectory calculations based on odometry were in error of about 8.2%. When conducting the study on the second type of video records, i.e. in case of the circular motion, the error of calculations was 24.8%. Studies have shown that the Harris method is subject to computational errors not unlike Shi-Tomasi method.

4. Results

Comparing the implemented methods of key point detection minding the robot's speed in relation to the number of detected points the FAST method turned out to be the fastest detector among all others considered in these studies, since this method finds several times more points than the others within a shorter period of time. Also, this method will be the best solution with the tasks when the number of obtained points really matters. However, it is highly fallible with image distortions.

Both Harris and Shi-Tomasi detectors exhibited the best results in terms of speed (fig. 1). But the number of found points in these methods differs exponentially (10 times).

For a more accurate result when working with detectors it is better to neglect a small number of key points. Therefore, the Harris detector is better when we mean the operating time regarding to the number of points.

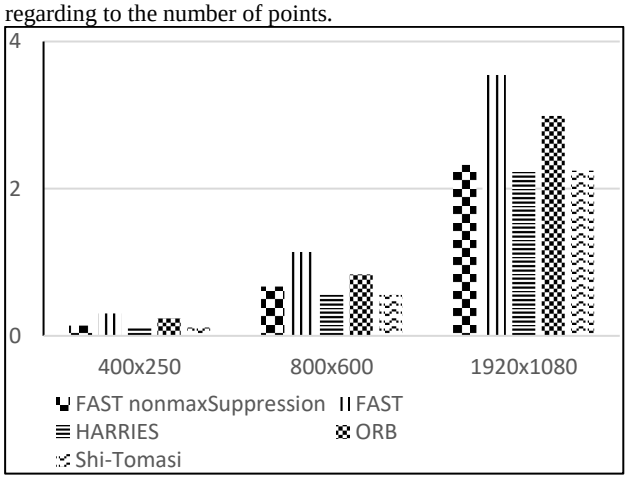

Fig. 1. The operating time of detectors depending from the image size

Maximum independence from all types of distortions was presented by both: ORB and Shi-Tomasi detectors (fig. 2).

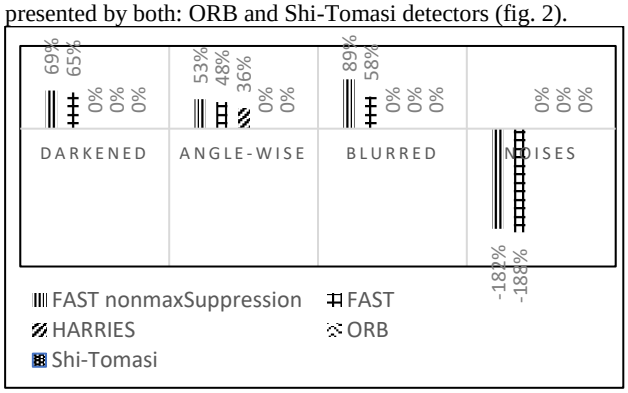

Fig. 2. Percentage of reduction of found points depending on image distortion

Thus, on the basis of the conducted experiments it can be concluded that the Harris detector is preferable in terms of speed

being independent from image distortions. But at the same time it is fallible to distortion of the image in case of rotation. The Shi-Tomasi method has the same execution time and is not fraught with image distortions.

However, under similar conditions this method found an order less key points. This fact can have negative consequences on the operation of the robot positioning algorithm as a whole. This may occur due to the fact that pictures are taken in the dynamics and at high speed of the robot, so the key points can go out of frame. That is why follow-up studies were conducted in regard to both methods : Harris and Shi-Tomasi

The results of optical flow methods research show the time dependence in percentage terms upon the video sequence duration. Figure 3 shows that the Lucas-Kanade optical flow method functions much faster than the Farneback dense optical flow method. When the resolution is 320x240, this difference reaches 5.8 times, if it is 640x360, the difference is 9.8 times, and in case of the 800x480 resolution it is 11 times faster. Thus, it can be concluded that the Lucas-Kanade method of the optical flow calculating is much more preferable comparing to the Farneback method having in mind the speed of execution in relation to a robotic vehicle and regardless of the video resolution.

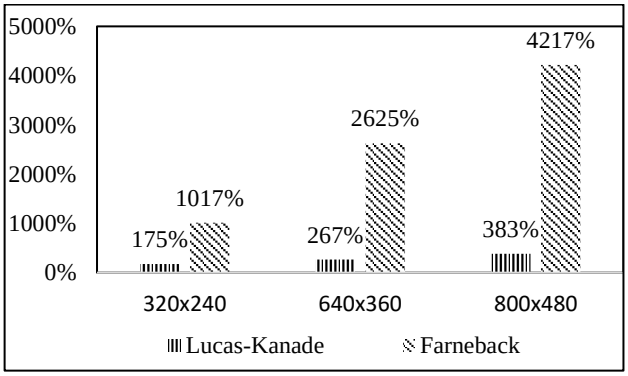

Fig. 3. Time taken by method in percentage terms regarding to video duration

The final stage of the study showed that the previously selected methods for identifying key points (Shi-Tomasi method and Harris method) do not provide 100% accuracy in the calculations of camera displacement. However, both of them can be applied in the algorithm of monocular visual odometry bearing in mind method errors. It is also worth noting that the Shi-Tomasi method has better accuracy comparing to the Harris method

5. Conclusion

The article analyzes the algorithm of visual odometry using a single camera (monocular visual odometry) for mobile object positioning. Empirical studies have been conducted for various methods aimed at implementing of algorithm steps.

According to the results of studies it has been concluded that using of the Shi-Tomasi method for an unmanned ground-based robotic vehicle operating when detecting key points of the image is the most preferable in comparison with the others considered here. This is due to the fact that it is the least prone to various image distortions such as darkening, rotation, blurring and noise.

To track the identified key points of the image in the video, it is more preferential to use the Lucas-Kanade optical flow method. This choice is based on the speed of its operating if it is compared or contrasted with some other considered methods. However, the experiments proved that this method is not able to manage each frame in real-time mode. So, when implementing monocular visual odometry for a moving unmanned vehicle, it is necessary to track key points only on subframes of video stream. Besides it, finding the traffic speed, at which the loss of information about its movement isn't significant should be indispensible.

The final studies have been fulfilled in experiments conducted to make the use of visual odometry as an option to identify the robot displacement relatively to the initial point possible. It should be noted that the methods for calculating the rotation and displacement matrices of the camera are quite sensitive to external parameters influencing their calculations and giving errors. It is also worth mentioning that errors in the calculations of the camera position displacement can be caused by the environment. For example, the low illumination of the room can result in the number of key points, which is insufficient for the algorithm work.

On the basis of the study it can be concluded that the monocular visual odometry having low accuracy in the calculations of motion indicators can not be the main method for determining the whereness of a moving system. However, in conjunction with data obtained from other sensors, it can serve as a method for the robotic ground vehicle positioning. Additional studies connected with assistive sensors application are essential for the conclusion confirmation.

6. References

[1] Calonder, M. BRIEF: Binary Robust Independent Elementary Features / M. Calonder, V. Lepetit, Chr. Strecha, P. Fua, // 11th European Conference on Computer Vision (ECCV). – 2010. – pp. 778 – 792.

[2] Checchin, P. Radar scan matching SLAM using the Fourier-Mellin transform / P. Checchin, Fr. Gerossier, Chr. Blanc // Field and Service Robotics / Springer. – 2010. – pp. 151–161.

[3] Cole, D.M. Using laser range data for 3D SLAM in outdoor environments / D.M. Cole, P.M. Newman // Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on / IEEE. – 2006. – pp. 1556–1563.

[4] Drummond, E.R.a.T. Fusing Points and Lines for High Performance Tracking, 2005.

[5] Farneback, G. Two-frame motion estimation based on polynomial expansion / G. Farneback // Lecture Notes in Computer Science. – 2003. – pp. 363-370.

[6] Harris, C. A combined corner and edge detector / C. Harris, M. Stephens // In Fourth Alvey Vision Conference, Manchester, UK. – 1988. – pp. 147-151.

[7] Hess, W. Real-time loop closure in 2D LIDAR SLAM / W. Hess, D. Kohler, H. Rapp, D. Andor // Robotics and Automation (ICRA), 2016 IEEE International Conference on / IEEE. – 2016. – pp. 1271–1278.

[8] Kitt, B.M. Monocular visual odometry using a planar road model to solve scale ambiguity / B. M. Kitt, J. Rehder, A.D. Chambers, M. Schonbein, H. Lategahn, S. Singh // In Proc. European Conference on Mobile Robots, September 2011– 2011.

[9] Lucas, B.D. An Iterative Image Registration Technique with an Application to Stereo Vision / B.D. Lucas, T. Kanade // Proceedings of the 7th international joint conference on Artificial intelligence. – 1981. – pp. 674-679.

[10] Rublee, E. ORB: an efficient alternative to SIFT or SURF / E. Rublee, V. Rabaud, K. Konolige, G. Bradski // Computer Vision (ICCV), IEEE International Conference on. IEEE. – 2011. – pp. 2564-2571.

[11] Shi, J. Good features to track / J. Shi, C. Tomasi // TR 93-1399, Cornell U., 1993

[12] Yi, J. IMU-based localization and slip estimation for skid-steered mobile robots / J. Yi, J. Zhang, D. Song, S. Jayasuriya // Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on / IEEE. – 2007. – pp. 2845–2850.