Mobile robot position assessment in the room with the use of on-board vision system

N V Udalova
Moscow Aviation Institute (National Research University), 4, Volokolamskoe shosse, Moscow, 125993, Russia
E-mail: vice1dean7@mai.ru

Abstract. The contemporary method of visual navigation of a mobile robot is presented. The visual navigation is an effective way to increase an accuracy of estimation for a current position of a mobile robot and its displacement. Using different specific features of a ceiling image along with photogrammetry theory permits to get a relative estimate for a current position of the mobile robot and its displacement in a room. Conventional approaches are ineffective because of presence of movable and immovable objects on the recognition scene. The ceiling image as the recognition scene is used in the proposed method. The recognition scene of ceiling is a stable and reliable image in the problem of finding reference points. Moreover there are no image distortions on the scene. All the mentioned above is a significant benefit of the proposed approach in terms of accuracy characteristics of the method of visual navigation of the mobile robot in the room.

1. Introduction
The navigation system of the mobile robot under consideration, is designed to assess the coordinates and orientation of the robot in the room [1, 2]. In such works, an inertial measurement module, including accelerometers and angular velocity sensors, is usually used as the core of the navigation system. Correction of errors accumulated by the inertial system is carried out by additional, for example, ultrasonic, infrared, etc. systems [3, 4].

This work is proposed to implement error correction using the robot technical vision system (STZ), using visual navigation methods [5, 6]. In this case, the observed scenes are room ceiling images. The characteristic elements of the ceiling image (sensors, lines, angles, design elements, light sources) are used by the STZ as visual reference points. To obtain images, a camcorder is installed on the robot, the axis of sight of which is directed vertically upwards. The advantage of this approach in comparison with the horizontal location of the camera is the use of an observation area that is not blocked by extraneous non-stationary objects, including people.

It should be emphasized that in this work, visual navigation is used as an auxiliary correction subsystem and does not provide a continuous assessment of the measured coordinates and orientation of the mobile robot in the room.

2. Formulation of the problem
The ceiling navigation system allows you to determine the position and orientation of the mobile robot based on the position of the observation system assessment relatively to certain visual landmarks with known coordinates. Various unique features of the ceiling images are used as reference points.
Let there be 2 images of the ceiling obtained by the mobile robot at 2 points of the trajectory of movement. In this case, some of the images capture the same section of the ceiling, the overlap of the two images must be at least 50% of the total image size. Since the optical axis of the video camera is directed vertically upwards, with a high probability it can be argued that there are geometric distortions between the obtained images in the form of linear displacement and rotation relatively to the optical center of the image. These distortions can be described using equations describing affine transformations [7]:

\[
\begin{align*}
    u_2 &= k \cdot u_1 \cdot \cos(\omega) - k \cdot v_1 \cdot \sin(\omega) + c_u; \\
    v_2 &= -k \cdot u_1 \cdot \sin(\omega) + k \cdot v_1 \cdot \cos(\omega) + c_v,
\end{align*}
\]

where \( u_1, v_1 \) are the coordinates of key points in the first image; \( u_2, v_2 \) are coordinates of key points in the second image; \( c_u, c_v \) — parameters of mutual linear displacement between images; \( \omega \) — rotation angle; \( k \) — scale factor.

In our case, \( k = 1 \) is known, since we believe that the ceiling height within one room is constant.

The stationary coordinate system of the room \( OXYZ \) is associated with the starting point of the robot. The exact value of the coordinates and the orientation of the initial position of the robot is irrelevant. Of practical importance are the relative distances and orientation from the starting point to the current position of the robot. If the coordinates and orientation of the robot for the first image are known, then the parameters of the movement of the robot at the point where the 2nd image was obtained can be calculated using formula (1). To estimate the parameters of the relative movement of the robot, it is necessary to determine the unknown values: the parameters of the mutual linear displacement between the images \( c_u, c_v \) and \( \omega \) — rotation angle.

The key points used as visual reference points on the ceiling are those points of the image that stand out against the background, are unique in a certain neighborhood and are resistant to noise and distortion [8].

The solution of the system of equations (1) makes it possible to obtain a linear equation with respect to unknown parameters:

\[
\frac{v_2 \cdot u_1 - u_2 \cdot v_1}{u_1} = -\sin(\omega) \cdot \left( \frac{v_1^2 + u_1^2}{u_1} \right) - \frac{v_1}{u_1} \cdot c_u + c_v.
\]

The number of key points in the image depends on the information content of the image and contains from ten to hundreds of such points. To solve the overdetermined system of equations, the least squares method is widely used; in this case, this method allows to find some “optimal” vector of unknown parameters from the point of view of the maximum proximity of the vector of deviations to zero.

Solving equation (2) using the least squares method allows to get parameter estimates \( c_u, c_v \) in the pixel coordinate system of the image. For the transition from the pixel coordinate system to the three-dimensional coordinates of the observed scene, a perspective geometric model of a camcorder is used [9], which is determined by the following relations:

\[
\begin{align*}
    x &= \frac{c_u \cdot Z}{f}; \\
    y &= \frac{c_v \cdot Z}{f},
\end{align*}
\]

where \( Z \) is the height of the ceiling relative to the mobile robot; \( f \) is the distance from the optical center to the image plane, which is determined at the stage of camcorder calibration.

3. Algorithm to assess a mobile robot position in a room via a computer vision system

The proposed algorithm to assess a mobile robot position is based on the use of various unique features of the ceiling image, identified by the methods of identifying visual features that are used
in computer vision systems. In particular, this approach can be implemented on the well-known correlation-extreme methods of visual navigation.

The algorithm to assess a mobile robot position can be divided into four stages:

- stage of camcorder calibration;
- stage of searching for key points on the image;
- stage of comparison of key points in two images;
- the stage of calculating the value of displacement and angle of rotation.

3.1. Calibrating the camcorder

The simplest from the point of view of material costs, namely the absence of the need for special equipment to install a camcorder and form the observed scene, is the method of calibrating a camcorder using a template. In this case, it is enough to take several photos of the template with known dimensions from different observation points. Usually, a ready-made template is used for these purposes, for example, in the form of a checkerboard. Key points are automatically selected on the resulting image, for which their relative position in space is known.

As a result of the calibration, we obtain the matrix of the optical parameters of the video camera $A$, the radial distortion coefficients $k_1, k_2, k_3$ and the tangential distortion coefficients $p_1, p_2$. The matrix of the optical parameters of a camcorder is presented:

$$
A = \begin{bmatrix}
\alpha_x & 0 & u_0 \\
0 & \alpha_y & v_0 \\
0 & 0 & 1
\end{bmatrix},
$$

(4)

where $\alpha_x, \alpha_y$ are the focal lengths along the $X$ and $Y$ axes; $u_0, v_0$ are coordinates of the optical center in the image plane.

Since distortion leads to distortions of the camera’s field of view (field of view), this effect is not acceptable in optics designed for photogrammetric measurements.

To eliminate distortion in the original images, the coordinates of the key points can be recalculated using the following equation:

$$
\begin{align*}
    u_{corr} &= u \cdot (1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) + 2 \cdot p_1 \cdot u \cdot v + p_2 \cdot (r^2 + 2 \cdot u^2); \\
    v_{corr} &= v \cdot (1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6) + 2 \cdot p_2 \cdot u \cdot v + p_1 \cdot (r^2 + 2 \cdot v^2),
\end{align*}
$$

(5)

where $(u, v)$ are the coordinates of the pixels in the original image, $(u_{corr}, v_{corr})$ are the coordinates of the pixel after removing geometric distortions.

3.2. Search for key points in the image

Today there is a large number of algorithms to find key points in the image (SURF, SIFT, BRIEF, ORB, etc.) [10–12]. The comparative analysis of key point detectors was carried out in terms of the quality of finding key points and the time of the detector operation. By “good” quality of finding special points we will consider such a detector, which is able to find special points both in areas with high contrast and in areas with low contrast. This need arises because on the observed scene, namely on the ceiling, there are artificial light sources that can be turned on or off at different times. When enabled, light sources become high-contrast objects. The detected key points on artificial light sources have a low positioning accuracy. Table 1 shows the results of using different detectors to find key points on one of the test images.

Based on the analysis of the main key point detectors functioning, the SURF (Speeded Up Robust Features) algorithm was chosen as the reference detection algorithm. First, this method is invariant to image scaling and rotation. For each point, the direction of maximum change
| Detector | Operation Time (ms) | Number of Key Points |
|----------|---------------------|----------------------|
| FAST     | 25                  | 786                  |
| BRISK    | 483                 | 325                  |
| ORB      | 40                  | 402                  |
| SIFT     | 127                 | 217                  |
| SURF     | 61                  | 508                  |

in brightness (gradient) and scale are calculated using the coefficients of the Hessian matrix. Secondly, this algorithm is optimal in terms of speed and efficient in terms of key points being detected.

### 3.3. Comparison of key points in two images

Since there are similar objects in the ceiling image, the matching algorithm sometimes does not correctly form a pair of key points. Figure 1 shows two sequential images, the dots show the location of key points that were found using the SURF detector. Lines connect pairs of corresponding key points. The image visually clearly shows that some pairs of key points were found incorrectly.

![Figure 1. Algorithm for matching key points without correction.](image)

To isolate incorrectly found pairs of key points from the sample, it is supposed to be used in the least squares method of element-wise removal of those pairs of key points that introduce the greatest error in the estimation of the model parameters (2). The idea of this approach is similar to the well-known RANSAC algorithm. The running time of the RANSAC method directly depends on the number of key points and the required accuracy to estimate the model parameters. This circumstance is essential for systems operating in real time. Before use the least squares method, the coordinates of the key points must be recalculated according to equation (5). Figures 2–4 show the result of removing incorrect pairs of key points from the initial sample.

The proposed approach allows you to correct the resulting matrix of matching key points, and, consequently, to increase the accuracy of finding unknown parameters $c_u$, $c_v$ and $\omega$.  

---

**Table 1. Detector results**

| Detector | Operation Time (ms) | Number of Key Points |
|----------|---------------------|----------------------|
| FAST     | 25                  | 786                  |
| BRISK    | 483                 | 325                  |
| ORB      | 40                  | 402                  |
| SIFT     | 127                 | 217                  |
| SURF     | 61                  | 508                  |
Figure 2. Complete set of key points.

Figure 3. The set of key points that make the most mistakes.

Figure 4. A set of key points that satisfy the requirements.

3.4. *Calculation of the offset and the angle of rotation*

The accuracy of calculating the displacement and angle of rotation depends on many factors: the quality and characteristics of images, technological errors in the production of camcorders,

![Graph showing offset error and rotation angle error](image)

Figure 5. Accuracy of determining the displacement and angle of rotation of the mobile robot.
the accuracy of determining the location of key points, identification of key points, distance to the observed scene, as well as image processing methods. It is impossible to determine the impact of each source of errors separately. In addition, the nature of the action of each source varies depending on many factors. The action of an unknown source is also possible.

The formulas for transforming coordinates between the coordinate systems of the image and the object are derived provided that the image of the object in the image is obtained in an ideal central projection (3).

The algorithm was tested using a camcorder with an image size of $1920 \times 1080$. Figure 5 shows one of the results with the accuracy characteristics depending on the height of the room.

Thus, for an image with a size of $1920 \times 1080$, the error in measuring the displacement along the $OX$ and $OY$ axes is less than 8 cm, and the error in measuring the angle of rotation of the mobile robot is less than 0.8 degrees.

4. Conclusion
Theoretical and experimental studies were carried out to confirm the efficiency of the proposed approach. In the process of research, new, practically significant results were obtained. It is shown that the use of a ceiling with a tiled structure leads to the appearance of false pairs of key points, which makes it difficult to correctly assess the position of the robot. To eliminate such uncertainties, it is proposed to use a hierarchical construction of the search algorithm.

In this paper, a promising method of visual navigation of a mobile robot in rooms is considered. The uniqueness of the method lies in the use of the ceiling image as the observed scene. The use of ceiling navigation eliminates the effect of obstructing visual reference points by people and mobile objects, which increases the reliability and accuracy of the robot navigation system. The experiments carried out have confirmed the efficiency of the proposed solutions.

Acknowledgments
The work was supported by RFBR grant 19-08-00613/a.

References
[1] Krasil’shchikov M N and Sebryakov G G 2003 Control and guidance of unmanned maneuverable aircraft based on modern information technology (Moscow: FIZMATTLit)
[2] Zenkevich S L and Nazarova A V 2006 Control system of a mobile wheeled robot Herald of the Bauman Moscow State Technical University 3 31–51
[3] Cui W, Wu C, Zhang Y, Li B and Fu W 2015 Indoor robot localization based on multidimensional scaling International Journal of Distributed Sensor Networks 11 8
[4] Noskov V P and Rubtsov I V 2007 Formation of a unified model of the external environment based on information from a camcorder and a rangefinder Mechatronics, automation, control 8 2–5
[5] Pershina Zh S, Kazdorf S Ya and Lopot A V 2019 Methods of visual navigation of a mobile robot and construction of cartographic models of the external environment Avtometria 55 92–102
[6] Miasnikov V A, Ignatiev M B and Pokrovskii A M 1984 Robot Visual Systems (Leningrad: Mashinostroenie) pp 354–81
[7] Kim N V 2001 Image processing and analysis in technical vision systems (Moscow: Moscow Aviation Institute Press)
[8] Vizilter Yu V, Gorbatevich V S and Moiseenko A S 2020 Single-shot face and landmarks detector Computer Optics 44 589–95
[9] Kudinov I A, Pavlov O V and Kholopov I S 2015 Implementation of an algorithm for determining the spatial coordinates and the angular orientation of an object based on reference marks, using information from a single camera Computer Optics 39 413–9
[10] Bay H, Ess A, Tuytelaars T and Van Gool L 2008 Speeded-up robust features (SURF) International Journal on Computer Vision and Image Understanding 110 346–59
[11] Rublee E, Rabaud V, Konolige K and Bradski G 2011 ORB: an efficient alternative to SIFT or SURF International Conference on Computer Vision (Barcelona: IEEE) pp 2564–71
[12] Leutenegger S, Chli M and Siegwart R 2011 BRISK: Binary Robust Invariant Scalable Keypoints International Conference on Computer Vision (Barcelona: IEEE) pp 2548–55