

Development of the technical vision algorithm

Pavel Elugachev^{1,*}, and *Boris Shumilov*¹

¹Tomsk State University of Architecture and Building, 634003 Tomsk, Russia

Abstract. The purpose of the work is creation of a set of technical means (switches, sensors) for recognition of transport infrastructure facilities, including development of algorithms for autonomous operation of technical facilities under changing environmental factors. In this work, we used methods for determining the volume of a three-dimensional facilities from the data of photo and video recording of the surrounding situation. The algorithm of technical vision was obtained, which is implemented as a program on a mobile device for recognition by means of stereometry of transport infrastructure facilities and their defects and storage of transport infrastructure defects. The novelty of the research is building of decision algorithms based on devices and sensors that recognize changing road conditions, namely, defects in coverage. The data obtained can be used in the planning of road repairs, in the analysis of traffic accidents by road police, road users for processing complaints, etc.

Introduction

The key component in implementation of a fully autonomous robot capable of performing tasks of collecting road data is the technical vision system [1-2]. Research in this area is performed by Google (USA), in Russia, the research is performed by Kamaz concern and others. The rapid growth in the number of automotive motor vehicles (AMV) in Russia and the world is many times ahead of the pace of road construction. As a result, the road network operates in stressed conditions. The condition of highways, the quality of the road bed surface, the visibility, the width of the roadway, the arrangement of the relevant signs have a significant impact on road safety and define the concept of “road conditions” in their totality. In investigation of road traffic accidents, in most cases it is believed that their main causes are the negligence or mistakes of the driver - the person (incorrect assessment of road conditions), as well as the AMV fault. According to experts, the real impact of road conditions on occurrence of road accidents is from 60 to 80% of cases [3].

The trajectory and speed of the automobile depends on the “road conditions”. If a person drives the car, then he or she makes a decision about the driving regimens based on their experience and psychological and physiological state, if the robot is driving, it is orientated on the basis of a set of sensors that allow navigation in space [4]. In the investigation of road accidents involving robots, it is believed that their main reasons are the lack of data on the engineering infrastructure due to the limited action of sensors and

* Corresponding author: Elugachev@mail.ru

the dynamically changing situation on the road. Creating a database and accessing it in road conditions is not effective, since it slows down the process of data processing and search for necessary solutions, and the dynamically changing road situation causes erroneous decisions in operation of control algorithms. Development of automatic diagnostic systems for evaluation of defects and detection of destructions of the road surface is possible with the use of video cameras and image processing algorithms. However, the problem of the operation of systems based on pure video is their inability to filter out such shading in pictures as road marking elements, smudges from oil and lubricant spills, brake traces, shadows of passing cars, and fresh patches [5].

One of the classical methods of extracting information about the depth of a three-dimensional (3D) object is the method of photogrammetric processing of stereoscopic images of a 3D object obtained from different angles [6-7]. The drawback of traditional digital photogrammetry is the significant involvement of the human operator, as well as the need to pre-scan analog images using high-precision scanners. The use of modern inexpensive photogrammetric methods [8-10] to determine the damage and unevenness of the road surface and structural elements of the road gives a new impetus to the development of digital technologies for the design of road repairs, increasing mobility and lowering the cost of the works.

Research methods

Due to affordability and prevalence of mobile devices (smartphones, unmanned aerial vehicles (drones), etc.), it is proposed to use mobile measurements on a series of successive video shots. The most accessible method can be to determine the amount of damage using markers from video recording [6]. To determine the volume of a three-dimensional object from the photo and video data, it is permissible to apply the method of reconstructing a three-dimensional coordinate vector from two perspective projections forming a stereo pair [11] or a simpler method using orthographic projections. This allows acting without accurate information about the transformation, relying on additional information about the object, which is provided with the presence of markers. From the graphical marks, the elements of the geometric transformations matrix are formed and the three-dimensional coordinates of the object points are determined. Then, the task of determining the volume of the object from its video image is reduced to transformation of the image into a digital volumetric model.

The main stages of the solution are as follows:

- identification of stable features of the video series;
- determination of reference frames;
- localization and definition of typical points of the video frame;
- solution of the inverse problem of photogrammetry;
- obtaining the three-dimensional mathematical model of the object. Further use of this model for expert evaluation and the formation of design estimates.

With the designation of homogeneous geometric coordinates, the transformation of the linear perspective (Fig.1) is represented as a 4x4 matrix:

$$[x \ y \ z \ 1][T'] = [x' \ y' \ z' \ h], \tag{1}$$

where x,y,z are point coordinates in the three-dimensional space, x',y',z' are homogeneous geometrical coordinates of the point, h is the scale factor, T' is the four-dimensional matrix,

$$[T'] = \begin{bmatrix} T'_{11} & T'_{12} & T'_{13} & T'_{14} \\ T'_{21} & T'_{22} & T'_{23} & T'_{24} \\ T'_{31} & T'_{32} & T'_{33} & T'_{34} \\ T'_{41} & T'_{42} & T'_{43} & T'_{44} \end{bmatrix}. \tag{2}$$

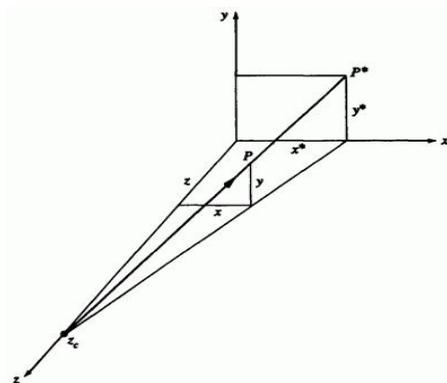


Fig. 1. Perspective projection to the point.

In the course of photographing, the results are projected onto a two-dimensional plane, in this case, on a $z = 0$ 0×1 plane by means of the projection transformation

$$[T''] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

The composition of these two linear transformations gives the following

$$[T] = [T''] [T'] = \begin{bmatrix} T'_{11} & T'_{12} & 0 & T'_{14} \\ T'_{21} & T'_{22} & 0 & T'_{24} \\ T'_{31} & T'_{32} & 0 & T'_{34} \\ T'_{41} & T'_{42} & 0 & T'_{44} \end{bmatrix} \tag{4}$$

As a result, let us put down the transformation in the following form

$$[x \ y \ z \ 1] [T] = [x' \ y' \ 0 \ h] = h [x^* \ y^* \ 0 \ 1]. \tag{5}$$

Where x^* and y^* are the coordinates of perspective projection on the picture plane of the photo image $z = 0$. After excluding scale factor h , we obtain two scalar equations:

$$(T'_{11} - T'_{14}x^*)x + (T'_{21} - T'_{24}x^*)y + (T'_{31} - T'_{34}x^*)z + (T'_{41} - T'_{44}x^*) = 0, \tag{6}$$

$$(T'_{11} - T'_{14}x^*)x + (T'_{21} - T'_{24}y^*)y + (T'_{31} - T'_{34}y^*)z + (T'_{41} - T'_{44}y^*) = 0, \tag{7}$$

Under the assumption of the known $[T], x, y, z$, these equations can be used for direct simulation of the photographing process. If x^*, y^*, x, y, z , are known, then equations (6) and (7) are two equations with 12 unknown elements T'_{ij} . By applying these equations to $n \geq 6$ noncoplanar points in the object space and to their images on the perspective projection, we obtain a homogeneous system of $2n$ equations with 12 unknown values.

To solve the resulting system, let us transfer the terms containing normalizing coefficient T'_{44} to the right-hand side and set value $T'_{44} = 1$. Thus, to find the solution of $[T]$,

we obtain the overdetermined system of equations, the matrix of which cannot be inverted, since it is not square. As it is known from the theory of the method of least squares, the best averaged solution can be calculated by multiplying both sides of the matrix equation by the transposed matrix of the system. Then we obtain a system of 11 linear equations with respect to 11 unknown values with a symmetric square matrix. This equation can be solved with the square root method. Thus, the known coordinates are used to determine the transformation generating this perspective projection, for example, a photo.

Finally, the last approach [12] assumes that $[T], x^*, y^*$ is known. In this case, two equations are obtained from three unknown spatial coordinates x, y, z . This is an underdetermined system of equations, so it is impossible to solve it. However, if two perspective projections are known, say, two photographs obtained from different angles, then equations (6 and 7) can be written for both projections. Then we get the following:

$$\begin{aligned}
 (T_{11}^1 - T_{14}^1 x^{*1})x + (T_{21}^1 - T_{24}^1 x^{*1})y + (T_{31}^1 - T_{34}^1 x^{*1})z &= x^{*1} - T_{41}^1, \\
 (T_{12}^1 - T_{14}^1 y^{*1})x + (T_{22}^1 - T_{24}^1 y^{*1})y + (T_{32}^1 - T_{34}^1 y^{*1})z &= y^{*1} - T_{42}^1, \\
 (T_{11}^2 - T_{14}^2 x^{*2})x + (T_{21}^2 - T_{24}^2 x^{*2})y + (T_{31}^2 - T_{34}^2 x^{*2})z &= x^{*2} - T_{41}^2, \\
 (T_{12}^2 - T_{14}^2 y^{*2})x + (T_{22}^2 - T_{24}^2 y^{*2})y + (T_{32}^2 - T_{34}^2 y^{*2})z &= y^{*2} - T_{42}^2,
 \end{aligned} \tag{8}$$

where upper indices 1 and 2 denote the first and the second perspective projection. These equations represent four equations from three unknown spatial coordinates x, y, z . Thus, an overdetermined system of equations is again obtained, and one can apply the methods of least squares and the square root to find the solution.

As a result, the elements of the geometric transformations matrix are formed from the graphical marks and the three-dimensional coordinates of the object points are determined. Thus, the problem is reduced to solving a system of linear equations from which the elements of the transformation matrix are determined. It is enough for the road foreman to make two photos, and send them for further processing.

Isolation of characteristic points of the image and determination of the contour of damage can be performed by one of the known methods of Sobel, Laplace, Kani [13-14]. Availability of local inhomogeneities on defects of natural origin (potholes, rills, holes) makes it expedient to use blob detectors based on the Laplace method [15], for correct identification of interpolation points of a three-dimensional object. With that, availability of a rectangular image table makes it possible to effectively apply the modern theory of shearlets to solve the problems posed [16].

Shearlets were first identified in 2006 as a structure that allows efficient working with multidimensional data. These structures are widely used to suppress noise on images in order to improve visual perception or increase clarity. Within the scope of the task set, shearlets can serve as a pre-processing tool for stereo pair images to facilitate automatic detection of local features of the damage object.

Results

The practical application of the claimed algorithm is quite wide for recording and determination of the actual dimensions of potholes in the road surface. Let us apply the method to a real stereo pair (Fig. 2).



Fig. 2. Stereo pair: road cone against the background of damage to the road surface.

The coordinates of the vertices of the 3D object were measured with a ruler. The coordinates of the corresponding points on the images were recorded in a graphic editor using the mouse. Attempt to restore the 3D coordinates of the vertices of the object gave a good match with the original values. Moreover, it was possible to correct the measurements on the photo and the typing errors when taking samples during the calculations. Given the coordinates of the road surface point on the left and right images, one can estimate the accuracy of the presented technical vision algorithm. In our case, the distance from the tip of the cone to the asphalt coating calculated by the Pythagorean theorem was 31.975 cm, which is 0.078% different from the 32 cm value on the technical data sheet.

Given the features of the image of roadway defects (lack of clear boundaries, presence of foreign objects, insignificance of certain defects), it should be possible to allow “manual” editing. The road foreman must be provided with an interface that allows highlighting typical points by moving the cursor. The algorithm for creating such interface on a smartphone running Android OS is described in [17] (Fig. 3). The algorithm of the program on a smartphone for a road foreman can be as follows: the left and right image of the stereopair is alternately dropped onto the screen of the smartphone, the image of the wire model of the 3D object is superimposed on the image, the characteristic points of the object flash on the model, and the user moves the cursor to select the corresponding point on the photo, after processing both images of the stereo pair in the smartphone memory, linear perspective transformation matrices are formed, then the command is given to the user to select an arbitrary point on each stereo pair image from the corresponding test mode menu (if it is a typical point of the object), or set the mode of verification of the Pythagorean Theorem (if it is a point on the road base), or go to continue working; then a voice command is given to the user to select the next point on each stereo pair image, the distance between points is calculated, then the user is asked to select the third and so on points in each stereo pair picture.



Fig. 3. Image of the program running on the smartphone screen.

As the array of typical image points is replenished, it becomes possible to construct a three-dimensional mathematical model for defects of the road surface. Then, based on the obtained model, the area and the volume of the geometric figure characterizing this particular damage are calculated using the triangulation technology.

Conclusion

The result obtained corresponds to the stated goal of the study. An algorithm for detecting road surface defects based on the photogrammetry method was obtained as a result of the work. The algorithm is implemented in the program for mobile devices, which can be used by road services, traffic police, emergency surveyors in the cases where it is required to establish the size of defects and their location. Further directions of the study may include the analysis of the ways of automating the process of searching for graphic markers and reference points on photo-images, including damaged road surfaces, evaluating ways of managing autonomous mobile robots, analyzing the constantly changing traffic situation and its effect on the percentage of erroneously calculated transport infrastructure defects, as well as further optimization of the proposed stereo matching algorithm.

The research was performed with the financial support of the Russian Foundation for Basic Research and Tomsk Region Administration (project code 16-41-700400 p_a).

References

1. A.A. Boreyko, S.A. Mun, A.F. Shcherbatyuk, *Mechatronics, Automation, Control*, **8**, Attachment 2-8 (2008). (in Russian)
2. A.V. Vazuev, V.P. Noskov, I.V. Rubtsov, S.G. Carichenko, *Izvestiya SFedU. Engineering Sciences*, **186**, 121-132 (2016). (in Russian)
3. E.R. Domke, *Rassledovanie i jekspertiza dorozhno-transportnyh proisshestvij* [Investigation and examination of road accidents] (Akademija, Moscow, 2009). (in Russian)
4. A. Pronobis, Ph.D thesis, KTH Royal Institute of Technology, Stockholm, 2011.

5. G. Bao, Ph.D thesis, The University of Toledo, 2010.
6. A.V. Mezhenin, *Metody i sredstva raspoznavanija obrazov i vizualizacii* [Methods and means of pattern recognition and visualization] (ITMO, St. Petersburg, 2012). (in Russian)
7. A.S. Nazarov, *Fotogrammetrija* [Photogrammetry] (TetraSystems, Minsk, 2006). (in Russian)
8. A. Chiuso, P. Favaro, H. Jin, S. Soatto, *IEEE Trans. Pattern Anal. Machine Intell*, **24**, 523 (2002).
9. A.J. Davison, I.D. Reid, N.D. Molton, O. Stasse, *IEEE Trans. Pattern Anal. Machine Intell*, **29**, 1052 (2007).
10. R. Tsai, *IEEE J. of Robotics and Automation*, **3**, 323 (1987).
11. D.F. Rogers, J.A. Adams, *Mathematical elements for computer graphics* (McGraw-Hill, New York, 1990).
12. C. Slama, (ed.), *Manual of Photogrammetry* (American Society of Photogrammetry, Falls Church, Va., 1980).
13. M. Sonka, V. Hlavac, R. Boyle, *Image Processing, Analysis and Machine Vision* (Thomson, Toronto, 2008).
14. Harris affine region detector, Available at: https://en.wikipedia.org/wiki/Harris_affine_region_detector (June 11, 2018)
15. T. Tuytelaars, K. Mikolajczyk, *Foundations and Trends in Computer Graphics and Vision*, **3**, 177-280 (2007).
16. G. Gao, Y. Liu, D. Labate, *Journal of Visual Communication and Image Representation*, **32**, 83-94 (2015).
17. S.A. Prokhanov, B.M. Shumilov, A.Zh. Kuduev, K.A. Abdykalykov, *Proceedings of Ninth Siberian Conference on Parallel and High Performance Computing (PVV-2017)* (October 10–12, 2017, Tomsk, Russia). (in Russian)