Range Measurement by Computer Vision Systems Based on Invariant Moments

Mikhail Khachumov
Intelligent Control Laboratory
Ailamazyan Program Systems Institute
of Russian Academy of Sciences
Pereslavl-Zalessky, Russia
Federal Research Center "Computer
Science and Control" of RAS
Moscow, Russia
khmike@inbox.ru

Vyacheslav Khachumov
Information technologies department
RUDN University
Moscow, Russia
Federal Research Center "Computer
Science and Control" of RAS
Moscow, Russia
vmh48@mail.ru

Nikolay Abramov
Intelligent Control Laboratory
Ailamazyan Program Systems Institute
of Russian Academy of Sciences
Pereslavl-Zalessky, Russia
n-say@nsa.pereslavl.ru

Vitaly Fralenko
Intelligent Control Laboratory
Ailamazyan Program Systems Institute
of Russian Academy of Sciences
Pereslavl-Zalessky, Russia
alarmod85@hotmail.com

***Abstract* — We review range measurement methods based on various configurations of computer vision sensors. We discuss the mathematical model of a stereoscopic pair and the conditionality of its fundamental matrix, which is essential for correct recognition and range measurement. A lot of researchers study the measurement based on single video or photo cameras, they are often installed on small mobile flying vehicles. We focus on the methods based on invariant moments. It is demonstrated that range measurement can be performed by a single camera. A complex of experiments was conducted in order to study single camera range measurement based on invariant moments. The deviation of experimental values of distance measurement that we have done didn't exceed 9%. The error can be explained by the inaccuracy in measuring the distance from the camera lens to the object, unstable position of the camera, and low quality of the obtained image due to the camera restrictions.**

Keywords — invariants, image, computer vision, model, video camera, distance, measurement method.

I. Introduction

Range measurement, i.e. measuring the distance to objects under observation, is a part of many algorithms, including the ones used for recognition. For the solution of this problem, various devices and physical principles can be applied [1].

The advantages of ultrasound sensors include their ability to measure the distance to translucent objects, and relatively low cost [2]. Their disadvantages are the following: small range (20-25 m) due to sound wave dispersion, relatively low precision, and the requirement that the object is large enough and has a sound-reflecting surface. This method is suitable for range assessment or preliminary calculation.

Laser rangefinders compare well with other devices [3]. They offer wide range (30-250 m), high precision (about ± 1-5 mm) throughout the range, and the ability to measure distances to small objects. On the other hand, they fail when the illumination is too bright, and cannot measure distances to windows and mirrors. Infrared sensors have an infrared light source and a sensor of the reflected light. Their advantage is high-speed measurement of distances to hot objects. The shortcomings are the influence of adjacent hot bodies, loss of precision when the lenses get dirty, and a small range (from 20 cm to 100 m).

Under certain conditions, we can use computer vision systems (CVS) based on stereoscopic pairs or a single video/photo camera [4,5] transformed into a measuring device with incorporated models and algorithms. Range measurement by images from two cameras is a key CVS problem essential for mobile robot navigation. For example, during Mars Exploration Rover mission [6], the MER-B Opportunity rover used a stereo navigation camera to obtain 3D images of Cape Tribulation, one day before climbing its summit. Ingenuity, the first helicopter operating on Mars (NASA's Mars 2020 mission), uses visual odometry and positioning, processing the optical flow coming from an onboard camera [7].

Mathematical models play an important part in computer vision systems. For example, stereoscopic pairs use models represented in the form of matrices 4x4 or 3x3 [8-10]. When well conditioned, the mathematical models of stereoscopic pairs are stable with respect to the observation system characteristics, and can be considered as invariants to small fluctuations of the environment. Their parameters can be altered by changing the distance between the cameras.

A lot of researchers study the measurement based on single video or photo cameras [11-14]. They are often installed on small mobile flying vehicles.

The CVS enlisted here are an alternative to the conventional laser or ultrasound rangefinders, being used when those are lacking or along with them. The precision depends on the distance to the object, system calibration, and function parameters when generating the displacement map. As compared to laser or ultrasound rangefinders, CVS provide that the measurements are hidden, covering all distant objects within the field of view. Single camera measurements are usually performed by choosing different viewpoints. This hampers the measurement, especially when the object is moving. An interesting approach is to measure the geometrical sizes of a frontal projection applying invariant moments, without the use of the second camera. An example can be the invariant moments [15-17] whose application is efficient in the problems of detection and recognition of arbitrarily oriented objects on an image.

II. Range Measurement Based on a Stereoscopic Pair

In order to turn a stereoscopic pair into a measurement tool, you need to build its mathematical model as a left and right matrices of the form

$$A=\begin{bmatrix} T_{11}^1-T_{14}^1x^{1*} & T_{21}^1-T_{24}^1x^{1*} & T_{31}^1-T_{34}^1x^{1*} \\ T_{12}^1-T_{14}^1y^{1*} & T_{22}^1-T_{24}^1y^{1*} & T_{32}^1-T_{34}^1y^{1*} \\ T_{11}^2-T_{14}^2x^{2*} & T_{21}^2-T_{24}^2x^{2*} & T_{31}^2-T_{34}^2x^{2*} \\ T_{12}^2-T_{14}^2y^{2*} & T_{22}^2-T_{24}^2y^{2*} & T_{32}^2-T_{34}^2y^{2*} \end{bmatrix}, \quad (1)$$

here T_{ij}^k is an element of the mathematical model of the k -th camera represented by a fundamental matrix A .

The information necessary to build the matrix is collected with the technology of observing six points on the object surface, with known coordinates. Pairs of pixels should correspond each other in the following way: each pair is a projection of the same 3D point, but the first pixel belongs to the left image of the stereo camera, and the second one belongs to its right image.

The problem of the observed coordinate measurement (3D reconstruction) can be reduced to the solution of the system

$$AX=B, \quad (2)$$

where

$$B=\begin{bmatrix} T_{44}^1x^{1*}-T_{41}^1 \\ T_{44}^1y^{1*}-T_{42}^1 \\ T_{44}^2x^{2*}-T_{41}^2 \\ T_{44}^2y^{2*}-T_{42}^2 \end{bmatrix}, X=\begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

The system has no exact solution because it contains three unknowns and four equations. In order to estimate the maximum deviation of the system solution ξ, let us reduce the system (2) to the form

$$AX-B=E, \quad E=\begin{bmatrix} \xi \\ \xi \\ \xi \end{bmatrix}. \quad (3)$$

By solving the system (3), we can find the accuracy of ξ approximation. Note that the problem of restoring the coordinate X is practically equivalent to the problem of range measurement $r=\sqrt{x^2+y^2+z^2}$ if the focal length is neglected and the matrix (1) is well conditioned.

It is known that the system (2) can be approximately solved in the following way:

$$X=\left(A^T A\right)^{-1} A^T B. \quad (4)$$

To use this tool, we solve the problem of the fundamental matrix conditionality. In order to assess the quality of a mathematical model, the conditioning number should be calculated [18]. The conditioning number of matrix A is defined as $\operatorname{cond}(A)=Q/q$, where $Q=\max_{x\neq 0}\left(\|AX\|/\|X\|\right)$, $q=\min_{x\neq 0}\left(\|AX\|/\|X\|\right)$, where $\|X\|=\sum_{i=1}^{n}|x_i|$.

Let ε be a certain small disturbing parameter, and $\Delta=|X_\varepsilon - X|$ be the deviation of the solution. Let us now consider a disturbed system

$$(A+\varepsilon A_1)X_\varepsilon = B+\varepsilon B_1,$$

the deviation of the solution can be estimated as [19].

$$O(\varepsilon)=\left\|(A+\varepsilon A_1)^T(A+\varepsilon A_1)\right\|-\left\|A^T A\right\|.$$

If the fundamental matrix is well conditioned, small changes of the parameters lead to small changes in the accuracy of the problem solution.

Let $f(x,y,z)$ be an analog function describing the brightness of the points with the coordinates (x,y,z) in a three-dimensional space. For the discrete case (a digital image), the following central moments [20-22] are defined:

$$\mu_{lmn}=\sum_X\sum_Y\sum_Z (x-\overline{x})^l (y-\overline{y})^m (z-\overline{z})^n f(x,y,z),$$

where X,Y,Z is the domain of the image pixel coordinates, and $\overline{x},\overline{y},\overline{z}$ is the centroid.

From the works [20-22] several invariants were empirically chosen, forming a set of moments invariant to turns and shifts:

$$I_1=\mu_{200}+\mu_{020}+\mu_{002};$$
$$I_2=\mu_{200}\mu_{020}+\mu_{200}\mu_{002}+\mu_{020}\mu_{002}-\mu_{101}^2-\mu_{110}^2-\mu_{011}^2;$$
$$I_3=\mu_{200}\mu_{020}\mu_{002}-\mu_{002}\mu_{110}^2-\mu_{020}\mu_{101}^2-\mu_{200}\mu_{011}^2+$$
$$+2\mu_{110}\mu_{101}\mu_{011}-\mu_{011}^2-\mu_{101}^2;$$

$$F_1=\mu_{003}^2+6\mu_{012}^2+6\mu_{021}^2+6\mu_{030}^2+6\mu_{102}^2+15\mu_{111}^2-3\mu_{102}\mu_{120}+$$
$$+6\mu_{120}^2-3\mu_{021}\mu_{201}+6\mu_{201}^2-3\mu_{003}(\mu_{021}+\mu_{201})-3\mu_{030}\mu_{210}+$$
$$+6\mu_{210}^2-3\mu_{012}(\mu_{030}+\mu_{210})-3\mu_{102}\mu_{300}-3\mu_{120}\mu_{300}+\mu_{300}^2;$$
$$F_2=\mu_{200}^2+\mu_{020}^2+\mu_{002}^2+2\mu_{110}^2+2\mu_{101}^2+2\mu_{011}^2; \quad (5)$$
$$F_3=\mu_{200}^3+3\mu_{200}\mu_{110}^2+3\mu_{200}\mu_{101}^2+3\mu_{101}^2\mu_{020}+3\mu_{101}^2\mu_{002}+$$
$$+\mu_{020}^3+3\mu_{020}\mu_{011}^2+3\mu_{011}^2\mu_{002}+\mu_{002}^3+6\mu_{110}\mu_{101}\mu_{011};$$
$$F_4=\mu_{300}^2+\mu_{030}^2+\mu_{003}^2+3\mu_{210}^2+3\mu_{201}^2+3\mu_{120}^2+3\mu_{102}^2+3\mu_{021}^2+$$
$$+3\mu_{012}^2+6\mu_{111}^2;$$

$$F_5=\mu_{300}^2+2\mu_{300}\mu_{120}+2\mu_{300}\mu_{102}+2\mu_{210}\mu_{030}+2\mu_{210}\mu_{030}+$$
$$+2\mu_{201}\mu_{003}+\mu_{030}^2+2\mu_{030}\mu_{012}+2\mu_{021}\mu_{003}+\mu_{003}^2+\mu_{210}^2+$$
$$+2\mu_{210}\mu_{012}+2\mu_{201}\mu_{021}+\mu_{120}^2+2\mu_{120}\mu_{102}+\mu_{102}^2+\mu_{021}^2+\mu_{012}^2.$$

If the fundamental matrix (1) is well conditioned, the system (4) can be applied for 3D object recognition [23]. In the same time, the value $I_1 = \mu_{200} + \mu_{020} + \mu_{002}$ can serve as a rough estimation of the "three-dimensional" size of the observed image. This value, when the moments are not normalized, depends on the distance to the object and formally can be a base for the range measurement. The application of the system (4) is possible only when a stereoscopic pair is used and reliable 3D models of the objects are available, which involves difficulties. Most often, CVS are equipped by single video or photo cameras which observe the objects in the same plane; therefore, hereafter we will restrict our consideration by this case. The formulas for the 3D model can be simplified by removing one of the coordinates.

III. Range Measurement Based on a Single Video Camera

Let us consider the way of determining the distance from the video camera to the underlying surface based on invariant moments retrieved from halftone images.

For a digital 2D image, the central moments are defined as

$$\mu_{pq} = \sum_x \sum_y (\mathrm{x} - \bar{\mathrm{x}})^p (y - \bar{y})^q f(x, y),$$

where $p + q \le 3$; $(\bar{\mathrm{x}}, \bar{y})$ is the "center of gravity".

To build a measuring tool, it is enough to take only the first invariant moment (invariant to turns)

$$M_1 = \mu_{20} + \mu_{02} .$$

It is shown in [24] that the value $r = \sqrt{\mu_{20} + \mu_{02}} = \sqrt{M_1}$ characterizes the size of the image projected on the observation plane. The first moment characterizes the size of the observed (and known) object. This important fact is the base for the range measurement. If the value h is the distance to the object, from the geometric properties of projective transformations it follows that $r \cdot h = const$.

IV. Experiments on Range Measurement Based on Invariant Moments

Practical implementations of range measurement are presented on some online resources, for example, with stereoscopic pairs [9] and with single cameras [11].

The experimental object was an outline of an airplane. The image was moving relative to the camera, providing the change of scale. The distance from the initial position to the final one was split into 90 fixed phases corresponded by certain image frames and invariant moments. Each frame was processed using the OpenCV library, checking the statement $C = r \cdot h = const$.

In Fig. 1, the experimental dependence $r = \sqrt{M_1} = \varphi(h)$ is shown.

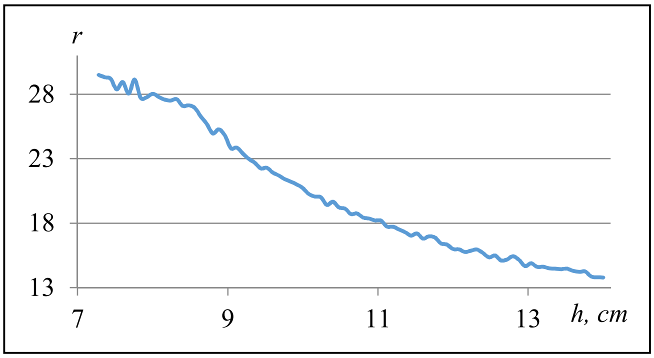

Fig. 1. Dependence $r = \varphi(h)$.

You can see that the function value r almost linearly decreases as the distance h grows. The deviations on the video frames are due to experimental errors.

In Fig. 2, you can see averaged values $C = \sqrt{M_1} h$ which theoretically should be constant for the observed object at different distances.

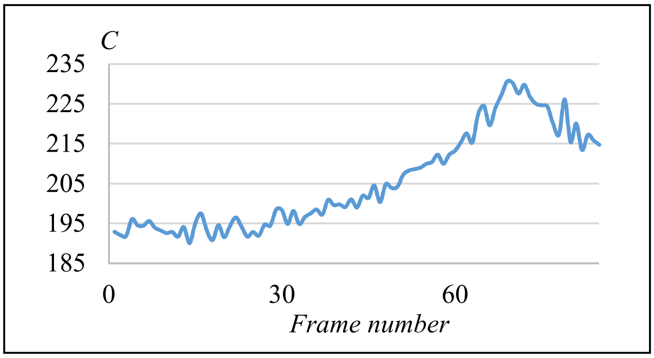

Fig. 2. Values $C = \sqrt{M_1} h$.

It will be correct to call the value C a pseudo constant because it is unstable but stays within certain limits.

The Fig. 3 and Fig. 4 demonstrate the change of the averaged estimation of C for various distances in another experiment.

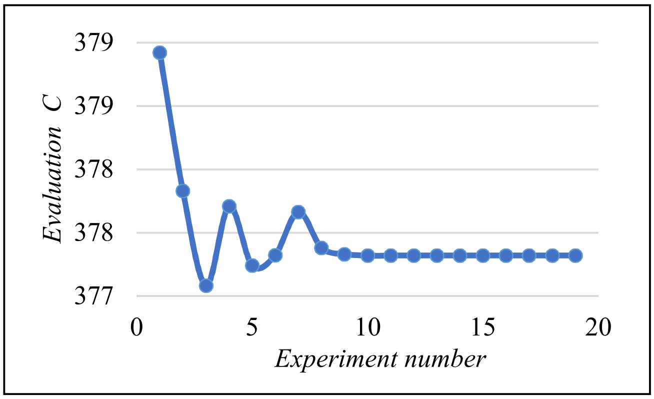

Fig. 3. Estimation of C on binary images for various distances.

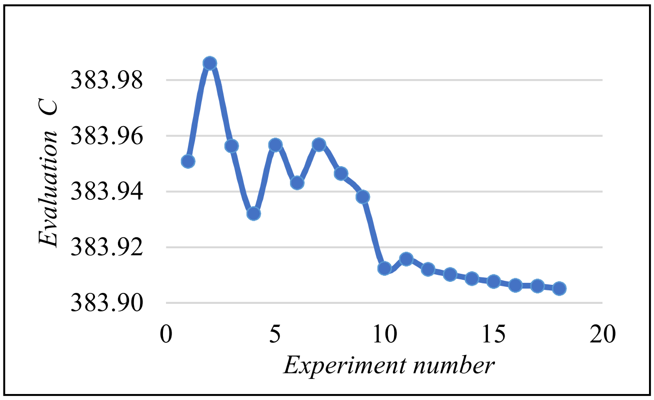

Fig. 4. Estimation of C on halftone images for various distances.

You can see that the deviation of experimental values of C did not exceed 9%. The error in the pseudo constant measurement can be explained by the inaccuracy in measuring the distance from the camera lens to the object, unstable position of the camera, and low quality of the obtained image due to the camera restrictions.

It follows from the experiments that the invariant moments are calculated with deviations which can be reduced by a better video equipment and post-processing. Having a pre-calculated pseudo constant C for each observed object, in each time moment you can approximately calculate the values r and h.

Conclusion

We have discussed the problems of range measurement using a stereoscopic pair or a single camera. In the case of a stereo image, the knowledge of the mathematical model of computer vision allows you to form a system of equations and find its approximate solution in the center of a tetrahedron. In the case of a single camera, it is appropriate to use invariant moments tolerant to turns of the image.

References

[1] G. Berkovic and E. Shafir, "Optical Methods for Distance and Displacement Measurements," Advances in Optics and Photonics, vol.4, pp.441–471 (2012), doi:10.1364/AOP.4.000441.

[2] Bongolan Riyadh Renzo S., Eden John Glendale M., Mendoza Elyzar Tim DT., Pancho Jim Kyle A., and Telan Elijah Carl M., "Ultrasonic Distance Measuring Device Study" (2019), doi:10.13140/RG.2.2.22528.17928/1.

[3] Fonseca Joao Antonio Santos Dias, Baptista António, Martins Ma Joao, and Torres João Paulo, "Distance Measurement Systems Using Lasers and Their Applications," Applied Physics Research, vol.9, №4, pp.33–43 (2017), doi:10.5539/apr.v9n4p33.

[4] F. Valocký, P. Drahoš, and O. Haffner, "Measure Distance Between Camera and Object using Camera Sensor," Cybernetics & Informatics, pp.1–4 (2020), doi:10.1109/KI48306.2020.9039879.

[5] A. Abdulov and A. Abramenkov, "Visual Odometry System Simulator," Proceedings of the International Siberian Conference on Control and Communications (SIBCON), 5 p. (2017), doi: 10.1109/SIBCON.2017.7998584.

[6] Approach to "Cape Tribulation" Summit (Stereo) (2015). URL: https://mars.nasa.gov/resources/6908/approach-to-cape-tribulation-summit-stereo/.

[7] B. Balaram, T. Canham, C. Duncan, M. Golombek, H.F. Grip, W. Johnson, J. Maki, A. Quon, R. Stern, and D. Zhu, "Mars Helicopter Technology Demonstrator," American Institute of Aeronautics and Astronautics (AIAA) SciTech Forum Conference, 8–12 January 2018, Kissimmee, Florida (2018), doi:10.2514/6.2018-0023.

[8] D.K. Patel, P.A. Bachani, and N.R. Shah, "Distance Measurement System Using Binocular Stereo Vision Approach," International Journal of Engineering Research & Technology (IJERT), vol.2, issue 12, pp.2461–2464 (2013).

[9] Calculating a Distance from an Object to a Camera Using Stereo Vision (2018). URL: https://stackoverflow.com/questions/51955394/calculating-a-distance-from-an-object-to-a-camera-using-stereo-vision.

[10] Depth Estimation from Stereo Video. URL: https://www.mathworks.com/help/vision/ug/depth-estimation-from-stereo-video.html.

[11] Measure Distance to Object with a Single Camera in a Static Scene (2016). URL: https://stackoverflow.com/questions/41102478/measure-distance-to-object-with-a-single-camera-in-a-static-scene.

[12] A. Peyman, "Object Distance Measurement Using a Single Camera for Robotic Applications," A thesis Submitted in partial fulfillment of the requirements for the degree of Master of Applied Sciences (MASc) in Natural Resources Engineering (2015). URL: https://zone.biblio.laurentian.ca/bitstream/10219/2458/1/Peyman%20Alizadeh%20MSc.%20Thesis%20Corrected_2_2.pdf.

[13] O. Shahar, A. Faktor, and M. Irani, "Space-Time Super-Resolution from a Single Video" (2011), doi:10.1109/CVPR.2011.5995360. URL: http://www.wisdom.weizmann.ac.il/~irani/PAPERS/singleVideoSR.pdf.

[14] J. Alarcon, "A Measure of Vision Distance for Optimization of Camera Networks," Electronic Theses and Dissertations (2014). URL: https://scholar.uwindsor.ca/etd/5199.

[15] J. Flusser, T. Suk, and B. Zitova, "Moments and Moment Invariants in Pattern Recognition," Wiley & Sons Ltd., London (2009). doi:10.1002/9780470684757.

[16] J. Flusser, B. Zitova, and T. Suk, "Moments and Moment Invariants in Pattern Recognition," John Wiley & Sons Limited, 314 p. (2018).

[17] M.-K. Hu, "Visual Pattern Recognition by Moment Invariants IRE," Transactions on Information Theory, vol.8, no.2, pp.179–187 (1962).

[18] M.H. Pesaran, "Time Series and Panel Data Econometrics," Oxford University Press, 1104 p. (2015).

[19] S.A. Amelkin, V.V. Trushkov, and V.M. Khachumov, "Assessment of the Accuracy of the Reconstruction of Three-Dimensional Coordinates in the Technical Vision System," Aerospace Instrument-Making, №10, pp.3–8 (2010).

[20] N.M. Thakare and V.M. Thakare, "A Supervised Hybrid Methodology for Pose and Illumination Invariant 3D Face Recognition," International Journal of Computer Applications (0975 – 8887), vol.47, no.25, pp.24–29 (2012), doi: 10.5120/7537-474.

[21] I. Sommer, O. Müller, F.S. Domingues, O. Sander, J. Weickert, and T. Lengauer, "Moment Invariants as Shape Recognition Technique for Comparing Protein Binding Sites," Bioinformatics, vol.23, no.23, pp. 3139–3146 (2007), doi: 10.1093/bioinformatics/btm503.

[22] T. Suk and J. Flusser, "3D Rotation Invariants," Department of Image processing (2012). URL: http://zoi.utia.cas.cz/3DRotationInvariants.

[23] Thanh Nguyen Duy, "Invariant in the Pattern Recognition," Discrete and Continuous Models and Applied Computational Science, №1, pp.76–85 (2016).

[24] S.A. Dudani, K.J. Breeding, and R.B. McGhee, "Aircraft Identification by Moment Invariants," IEEE Transactions on Computers, C-26, pp.39–45 (1977), doi:10.1109/TC.1977.5009272.