Available online at:https://ijact.in

Date of Submission	18/02/2021
Date of Acceptance	07/04/2021
Date of Publication	03/05/2021
Page numbers	3969-3972 (4 Pages)

ISSN:2320-0790

COMPUSOFT
An International Journal of Advanced Computer Technology

SOLVING LOCAL OPTIMA PROBLEM IN OBJECT MOVEMENT COMPUTATION

Jihun Park[1]

[1]Department of Computer Engineering, Hongik University, 94 Wowsanro , Mapo. Seoul, S. Korea

Abstract: In this paper, we have proposed that local optima exist for object motion calculation. While using the optimization method for object motion computation, we present a method to calculate the initial value necessary to overcome the local optima. The object motion was calculated for scenes photographed with a moving camera using the proposed method, and the result is presented using three-dimensional reconstruction.

Keywords: camera calibration; joint movement computation; kinematics;

I. INTRODUCTION

This paper proposed the existence of local optima for calculating all kinds of joint motion taken using a single camera. Recently, research on the movement computation of objects with joints is active, with the purpose of improving the performance of athletes or making movies from augmented reality. The purposes of calculating human motion include motion capture, surveillance, and motion analysis. Motion captures motions of the character. Surveillance monitors special movements tied to certain behaviors. Motion analysis has several purposes. It can analyze the athletic performance of athletes. It can analyze the force applied to joints of patients. Much effort has been devoted to motion analysis in a single image. The computation presented by this paper results in the relative orientation and position of the joints.

When 3D information is captured as an image, it is projected onto 2D image, during which a loss of information occurs. This loss of information should be recovered to find the movement of the joint. Recovering this information has been a difficulty, and this paper uses the following approaches. First, we fix a reference coordinate on an environmental object. Using multiple views of a scene as an input, we compute the orientation and position of the camera at the time of which scenes were captured. The joint movement is calculated using the equation of coordinate transformation as shown in this paper. Because camera formulas and kinematic body equations for joints are nonlinear, 3D joint reconstruction is not possible if the formulae for joint motion calculation are incorrect. There are several reasons that accurate joint body reconstruction is not possible using the method presented in this paper: an unsatisfactory initial value for parameter optimization, an error in the input data, or an erroneous computation for either of the camera orientation or position. If 3D reconstruction of the articulated body was possible, we can infer that the equation for joint angle calculations is correct.

II. RELATED WORKS

In this paper, we calculate the movement distance of the joint. The method of using the sensor is not covered in this paper. We use images to calculate the distance of joint movement. In order to calculate the motion of the joint using images, we can either use geometric analysis or deep-learning based training. Our approach is based on geometry.

The old method of extracting three-dimensional joint information from 2D image information is devised by Lee and Chen [1]. Lee and Chen's method calculates head motion. The requirement of their method is that minimum six feature points must be known in advance as well as the length of the human joint. In this method, when a computation error occurs in the calculation of coordinate system transformation using human head feature points, it affects not only the camera parameter calculation but also the whole calculation. Park and Sheikh [2] attempted using a sphere with a radius of bones. When the length of the bone is determined, their method can be regarded as a sphere having a radius of the bone, and two intersections between the sphere and a ray from the camera are possible. In terms of three-dimensional reconstruction, previous studies such as Hartley's book [3] and the Moons paper [4] are useful references. Many methods of three-dimensional reconstruction [3,5] are based on feature points.

III. PROOF OF LOCAL OPTIMA

A. *Simple Example*

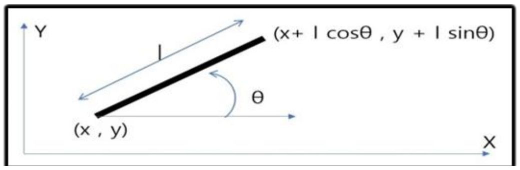


Figure: 1 A simple example of an articulated body.

In the following special case, an object consists of a single line segment and is only capable of rotation and translation on the xy-plane in a special form of three-dimensional motion as shown in Figure 1. The camera is only capable of parallel projection onto the xy-plane. The inputs are both endpoints of a single line segment. Our goal is to determine the translational and rotational angle of the object. Left tip is set as (x_1, y_1), while right tip is set as (x_2, y_2). There are two scenes 1 and 2. We find the variable values that make the two feature points (I_{x1f1},I_{y1f1}),(I_{x2f1},I_{y2f1}) of scene 1 coincide with the two feature points (x_{1f2},y_{1f2}),(x_{2f2},y_{2f2}) of scene 2 respectively. We define scene 1 the reference scene. To obtain coordinate information for scene 1, you must set the variable to the first feature point in scene 1. Generally, we do not know the three-dimensional location of the feature points of each scene. In this example, we know the projected two-dimensional point because the ray is projected onto a two-dimensional plane. When two rays point to the two feature points of scene 1, the algorithm presented in this paper transform the ray formulas, which are translated and rotated, resulting in the rays that pass through the feature points of scene 2. The length of the object, l, is unknown. The variables of this optimization task are the distance to the origin of the coordinate system of scene 1, the distance to the origin of the coordinate system of scene 2, and the relative rotation angle. If the variables are (x_{1f1},y_{1f1}) to the origin of the coordinate system of scene 1, the variables are (x_{1f2},y_{1f2}) to the origin of the coordinate system of scene 2, and the relative rotation angle variable is θ, the optimization equation including the transformation equation is as follows. Since the values of (x_{1f1},y_{1f1}) are (I_{x1f1},I_{y1f1}), and it is easily known, so it is excluded from the variable.

$$\min_{x_{1f2},\, y_{1f2},\, \theta} \left\| \begin{pmatrix} I_{x_{1f2}} & I_{x_{2f2}} \\ I_{y_{1f2}} & I_{y_{2f2}} \\ 1 & 1 \end{pmatrix} - \begin{pmatrix} cos\theta & -sin\theta & x_{1f2} \\ sin\theta & cos\theta & y_{1f2} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -I_{x_{1f1}} \\ 0 & 1 & -I_{y_{1f1}} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} I_{x_{1f1}} & I_{x_{2f1}} \\ I_{y_{1f1}} & I_{y_{2f1}} \\ 1 & 1 \end{pmatrix} \right\|^2 \tag{1}$$

If the two feature points of scene 1 are $(0,0)$ and $(l,0)$, the optimization equation including the above transformation formula is as follows. In this case, there are three variables, (x,y) and θ.In this case, the length of the object is l and the input points are (I_{x1f2}, I_{y1f2}) and (I_{x2f2}, I_{y2f2}) , and the kinematic equations of both ends of the object are (x, y) and $(x + l\ cos\ \theta, y + l\ sin\ \theta)$.

$$\min_{x,\, y,\, \theta} \left\| \begin{pmatrix} I_{x_{1f2}} & I_{x_{2f2}} \\ I_{y_{1f2}} & I_{y_{2f2}} \\ 1 & 1 \end{pmatrix} - \begin{pmatrix} cos\theta & -sin\theta & x \\ sin\theta & cos\theta & y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & l \\ 0 & 0 \\ 1 & 1 \end{pmatrix} \right\|^2 \tag{2}$$

B. *Proof of Local Optima*

The translation distance of the joint and the angle of rotation must be calculated to describe the movement of a joint. However, it is not easy to calculate the kinematic parameters of the joint because there exists a local optimum. We prove that there is a local optimum for calculations involving lower degree of freedom. If there is a local optimum in calculations with lower degrees of freedom, we will prove that there is a local optimum for higher degrees of freedom. The existence of local optima for the simple example in section 3.1 is demonstrated below. If there is a local optimum in this 2D special case, the local optima necessarily exist in the general 3D case. In this case, the length of the object is l and the input points are (I_{x1f2}, I_{y1f2}) and (I_{x2f2},I_{y2f2}) ,and the kinematic equations of both ends of the object are (x, y) and $(x + l\ cos\ \theta, y + l\ sin\ \theta)$. The optimization equation, $f(x,y,\theta)$, for finding the rotation and translation values of the object in accordance with the input feature points (I_{x1f2}, I_{y1f2}) and (I_{x2f2}, I_{y2f2}) is as follows.

$$Min_{x,y,\theta}\ (x-Ix1f2)^2+(y-Iy1f2)^2+(x+l\ cos\theta-Ix2f2)^2+(y+l\ sin\theta-Iy2f2)^2 \tag{3}$$

If $f(x,y,\theta)$ is the minimum value, by definition $\partial f(x,y,\theta)/\partial x = 0$, $\partial f(x,y,\theta)/\partial y = 0$, $\partial f(x,y,\theta)/\partial \theta = 0$ and we obtain three equations. One can find several sets of x, y, θ that satisfy the three equations. Therefore, there is a local optimum in the special case of two-dimensional translation and rotation. In general, it follows that there exists a local optimum when calculating the general three-dimensional translation and rotation motion using an optimization method.

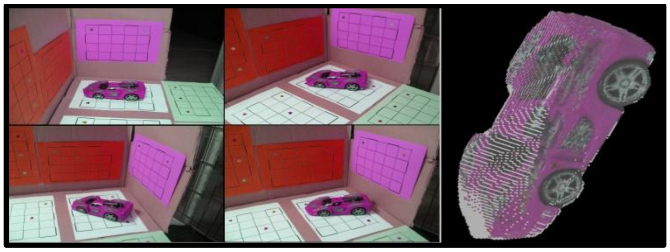
IV. COPING WITH LOCAL OPTIMA

A. *Fixed Scene Analysis and Camera Calibration*

Our 3D reconstruction method [6] is a variant of Zhang's method [7]. 3D scene analysis and camera calibration methods are not focus in this paper. The parameter optimization technique for calibrating cameras is similar to Zhang's method. In fact, this part can be replaced by any 3D reconstruction algorithm [3,5].

B. Initial Value Computation Method for Object Motion

As we have already proved, there is a local optimum for the function of calculating the motion of one joint segment. It is important to calculate the initial value of parameter optimization to calculate the joint motion while avoiding the local optima. For this purpose, we can calculate the approximate translational value and use it as the initial value for joints that include the translational function in contact with the surrounding environment. In this way, local optima could be avoided. There is only a small obstacle of local optima when the motion of an object is small. An example of an environmental object is a floor object in contact with the human foot, which is an articulated body. Plane equation such as a floor can be computed using 3D reconstruction as well as position and orientation of the camera using in photographing a scene. Since the orientation and position of the camera are calculated in advance, the plane equation of the floor object can be investigated by providing three or more independent feature points constituting the floor. By calculating the translational value of the articulated body using the provided information, the calculation of an object contact point with a plane object such as a floor can be found by calculating a point of intersection between a plane and a ray. The feature points used as the initial value must be close to the floor. In each scene, we cast a ray to one feature point of the first object and calculate the approximate translational distance of the first object using the point where it meets the floor plane. Initial value for a rotation angle is set to zero. This value is used as an initial value when calculating joint motion using the optimization method.



(a) four input files (b) reconstructed 3D volume of a moving car.

Figure: 2 Input images and corresponding three-dimensional volume reconstructions of a car.

V. EXPERIMENTAL RESULTS

Table 1: Number of feature points for Figure 2.

Fixed/moving	*Top left image*	*Top right image*	*Bottom left*	*Bottom right*
fixed part	19	25	28	30
moving part	6	6	6	6

This paper discusses on computing object movement using a moving camera. All input images are photographs of a body that contain different rotational angles and translational values. The assumption is that we know foreground and background parts. The method proposed in this paper is based on a single object computation, and three-dimensional reconstruction is done to prove our computation works.

The car in Figure 2 had two-dimensional translation and one-dimensional rotation. Six degrees of freedom were used to model relative translation and relative rotation. For this, (n−1) variables were used, where the extra n−1 constraints come from the quaternion.

In this paper, we provided background removed images and provided the correspondence between feature points as input. Feature extraction and background removal algorithms are not the focus of this research. More input images can produce more accurate three-dimensional articular reconstruction for all three-dimensional joint reconstructions. The experiment in this study always used only 4 input images for convenience.

Table 1 shows the number of input points used for 3D reconstructions with textures and the number of points used for each background part. It also has feature point information on the translational and rotational part for each input image. The correspondence relation between each feature point is entered by another algorithm or by a user. 3D position information on feature points on background objects can be calculated by applying any 3D reconstruction method. The origin of the reference coordinate of the three-dimensional reconstruction is fixed on the wall.

All the code was written in C programming language without using any library other than GRG2 [8], and the input image of this paper was created using one IXY810 portable digital camera.

VI. CONCLUSION

There are local optima in equation (3) during the calculation of object motion. The presence of local optima is demonstrated in Section 3.2. To overcome the local optimization, we use initial values of the parameters for calculating the object movement. If the initial value is close to the actual answer, one can easily find the corresponding object motion. Three-dimensional reconstruction is possible for each scene, making way for the calculation of correct angle of rotation and translation values. Using the three-dimensional position, an accurate value for the angle of rotation can be calculated. For more accurate calculations, more feature points are needed.

The problem that needs to be solved is the calculation of three-dimensional joint motion values for an articulated object. Specifically, as the camera moves, we need to find the three-dimensional motion values in a moving and rotating three-dimensional object. We use this result in three-dimensional moving object reconstruction. Many papers have attempted to calculate the correct joint angles, and their continuous efforts evidently show that this is a captivating topic open to diverse applications. It is very difficult to reconstruct three-dimensional motion without calculating an object movement. However, as shown by the

3D reconstructions presented, the method presented in this paper is accurate, supported by the correct 3D reconstructions.

An important constraint on the method presented in this paper is the presence of local optima. There is a local optimum as shown in Section 3. The initial value for optimization can be obtained by using the computation results of a correct calculation of adjacent scenes. In the case of multiple, consecutive input scenes with small changes in time, the object movement will be small. For input images with small object movement, it is much easier and faster to calculate the object movement using parameter optimization. It is convenient to use the optimization because the initial joint value and the correct joint value are similar.

The core of joint 3D reconstruction lies in the accurate calculation of the object movement. Because our 3D motion computation depends on parameter optimization, it is very important to minimize the number of variables used in the optimization. If the number of optimization variables increases, the computational time and complexity increase significantly. To reduce the number of variables used for optimization, we use quaternion.

VII. ACKNOWLEDGEMENT

This work was supported by 2019 Hongik University Research Fund.

VIII. REFERENCES

[1] Lee, H.J. and Chen, Z. 1985. Determination of 3D human body postures from a single view. *Computer Vision, Graphics and Image Processing , 30*, 148-168.

[2] Park, H.S.and Sheikh, Y. 2011. 3D reconstruction of a smooth articulated trajectory from a monocular image sequence. The IEEE International Conference on Computer Vision (ICCV).

[3] Hartley, R.and Zisserman, A. 2004. *Multiple View Geometry in Computer Vision*, 2 ed.; Cambridge University Press.

[4] Moons, T., van Gool, L.and Vergauwen, M. 2008. 3D Reconstruction from Multiple Images, Part I: Principles. *Foundation and Trends in Computer Graphics and Vision*, *4*, pp. 287–404.

[5] Moons, T., Vergauwen, M.and van Gool, L. 2008. 3D Reconstruction from Multiple Images. In *ICVSS08* ftp://ftp.esat.kuleuven.ac.bepsi/visics/konijn/ ICVSS08/vangool.pdf pp. 17–18.

[6] Park, J.and Park, S. 2013. Improvement on Zhang's Camera Calibration, *Applied Mechanics and Materials*, *479-480*, pp. 170–173.

[7] Zhang, Z. 2000. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, *22*, pp. 1330– 1334.

[8] Lasdon, L., Fox, R. and Ratner, M. 1973. Nonlinear Optimization using the Generalized Reduced Gradient Method. Department of Operations Research 325, Case Western Reserve University.