International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 18 (2018) pp. 13762-13766

Stereoscopic Vision System for reconstruction of 3D objects

Robinson Jimenez-Moreno
Professor, Department of Mechatronics Engineering, Nueva Granada Military University, Bogotá, Colombia.

Javier O. Pinzón-Arenas
Research Assistant, Department of Mechatronics Engineering, Nueva Granada Military University, Bogotá, Colombia.

César G. Pachón-Suescún
Research Assistant, Department of Mechatronics Engineering, Nueva Granada Military University, Bogotá, Colombia.

Abstract

The present work details the implementation of a stereoscopic 3D vision system, by means of two digital cameras of similar characteristics. This system is based on the calibration of the two cameras to find the intrinsic and extrinsic parameters of the same in order to use projective geometry to find the disparity between the images taken by each camera with respect to the same scene. From the disparity, it is obtained the depth map that will allow to find the 3D points that are part of the reconstruction of the scene and to recognize an object clearly in it. By Delaunay triangulation, the points obtained are joined in order to have the three-dimensional, reconstructed perspective of the object, obtaining 90% of recognition by this method.

Keywords. 3D vision, camera calibration, depth map, stereoscopic vision.

INTRODUCTION

A three-dimensional vision system allows to obtain the spatial location of an object in a scene. Said spatial location allows to know at what distance the object is located and in what direction, for which it is essential to have a device for capturing information, in the simplest case a camera.

Several developments in 3D reconstruction systems have been presented in the state of the art. In [1], it presents a robust development for systems that can count on the capacity of a graphic card. The applications of 3D reconstruction have room in various disciplines, in [2], a representative case is presented at the cellular level. In [3], an application case for reconstruction focused on dynamic obstacles as vehicles is presented, using laser sensors for this case.

3D reconstruction systems can be based on a single camera, as proposed in [4], however, in this case, the camera must move around the object. In [5], a reconstruction system using a single camera based on a Kinect sensor is presented, which consists of a laser system for shaping the point cloud of 3D space.

Current reconstruction systems employ improvements in the reconstruction system based on optimization systems [6] [7], or techniques such as optical flow algorithms [8]. But nowadays, applications of mobile robotics, for autonomous anthropomorphic agents [9] or not, require reducing costs and using free software for their development. Which, by means of laser systems or mono-camera, does not allow to obtain a reduction in price or analysis of adequate depth.

The system proposed in this article presents a system of reconstruction with stereoscopic pair, i.e. two cameras that will capture the scene from their own perspective allowing, by means of projective geometry, to establish point correspondences in the scene to find the calibration matrices of each one (camera matrices). It is of particular interest to find the fundamental matrix and the essential matrix that allow obtaining the intrinsic and extrinsic parameters of each one and under which to relate the 2D spatial location on the X and Y axes with the depth given by the Z axis.

The development of this work is supported by OpenCV libraries, which allow finding the point cloud that makes up the 3D object in space by obtaining a depth map of the scene, representing the distances from the camera to the objects by intensities of gray, as shown in Figure 1.

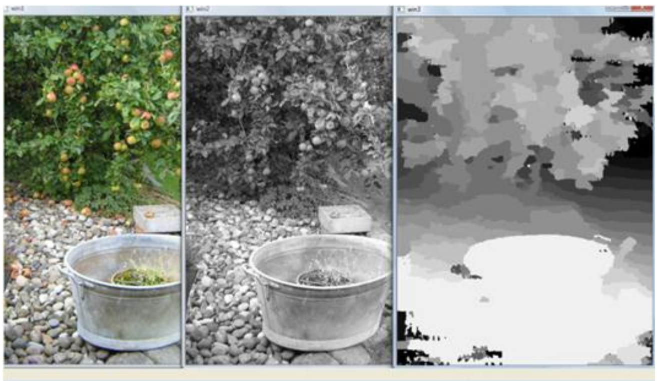

Figure 1. Depth map of a 3D scene.

Next, the camera calibration technique used is presented in Section 2, in Section 3, the 3D recognition of an object in the scene, in Section 4, an analysis of results is presented and finally, in Section 5, the conclusions obtained.

WORKSPACE AND CAMERA CALIBRATION

For the calculation of depth information in a stereoscopic system, the spatial and environmental relationship of the scene must be precisely defined. For this, a tripod base was implemented (Figure 2a), where the pair of cameras are fixed with a fixed and leveled distance and angle with respect to the ground (Figure 2b).

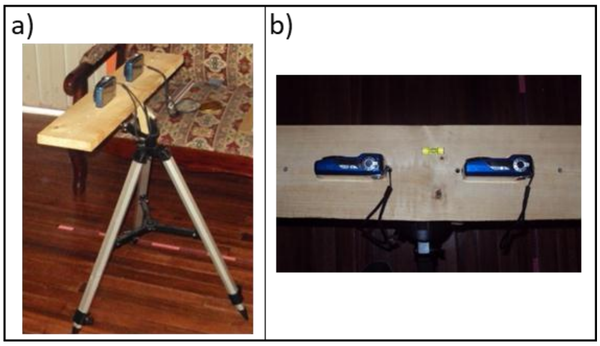

Figure 2. Stereoscopic arrangement.

The background in the scene may present noise, in this case, it is sought to centralize the detection of a particular object which has to be located in both cameras, so a uniform background is established that is easy to remove from both images, as illustrated in Figure 3.

Figure 3. Background scenario.

Establishing these operating characteristics of the stereoscopic system, determines the first step to characterize the stereoscopic system for depth measurement, such as the coordinates of each camera and the focal distance between them, aspects required to generate a camera model and the calibration thereof. Both aspects are described below.

CAMERA MODEL

Camera calibration requires knowing both intrinsic and extrinsic aspects of the camera, including the number of pixels per unit of distance defined as mx and my in the x and y directions, respectively. From these parameters, it is gotten that the calibration matrix for a CCD camera, like those used, is given by equation 1.

$$A = \begin{pmatrix} fx & 0 & Cx \\ 0 & fy & Cy \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (1)$$

Where fx and fy represent the focal length in terms of pixels in each direction. In turn, $Cx=mx*x_0$ and $Cy=my*y_0$, represent the origin of the coordinates in the plane of the image. Such parameters allow to identify the projection form of the different elements in the scene, within the projection plane of the camera.

In such a way that the 3D projection within the plane of the image can be calculated by means of the projective transformation (s) given by equation 2, where u, v represent the projection points in the plane of the camera. The first matrix after the equality is the matrix that relates the intrinsic parameters and the following matrix relates the extrinsic parameters, which includes the translation and rotation of the same (see equation 2).

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} fx & 0 & Cx \\ 0 & fy & Cy \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} r11 & r12 & r13 & t1 \\ r21 & r22 & r23 & t2 \\ r31 & r32 & r33 & t3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (2)$$

CAMERA CALIBRATION

The first aspect to take into account for the calibration of the pair of cameras that make up the stereo vision system is to use a calibration pattern, for this case it was done using a chessboard type mobile pattern as seen in Figure 4.

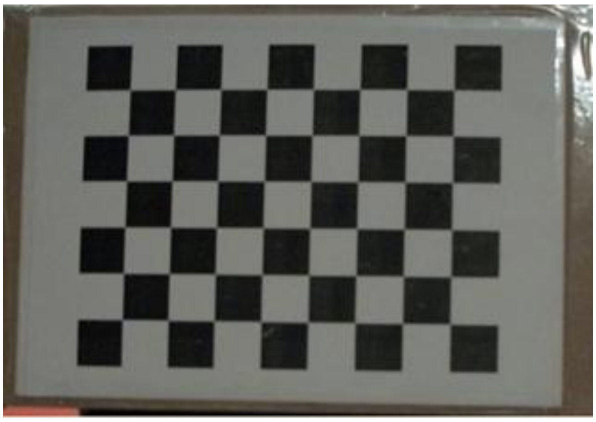

Figure 4. Calibration pattern.

The calibration process involves having an image for each camera in the same instant of time and running an algorithm to find the internal corners of the chess grid. This process must be performed several times changing the location of the calibration pattern as shown in Figure 5, in which these corners have already been identified by means of the OpenCV machine vision library, developed by INTEL and freely distributed.

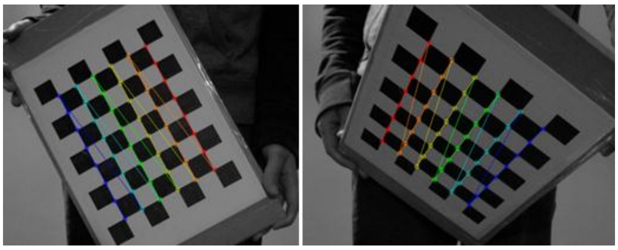

Figure 5. Variations of the calibration standard.

The OpenCV function that allows this task is *cvFindChessboardCorners*, from here it is possible to determine the point correspondences between the images of both cameras by projective geometry relationships, making use of epipolar lines, as seen in Figure 6 and by means of the *cvComputeCorrespondEpilines* function.

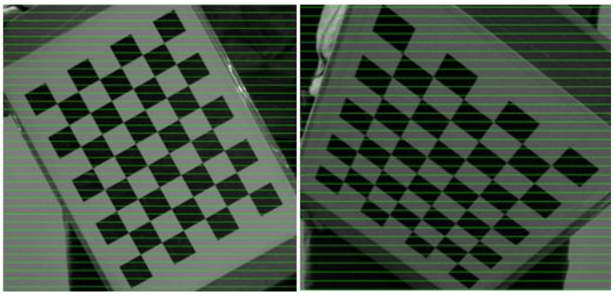

Figure 6. Epipolar lines.

The use of these functions allows to establish the necessary parameters to be able to find the fundamental matrix (*F*), which is related to the points of both images by means of equation 3.

$$x_i F x' = 0 \quad (3)$$

The epipolar lines are fundamental for the matching of points in the calibration pattern, in order to find coincidences in both images. The mathematical relationship between the epipolar lines of the left (*l*) and right (*l'*) image is given by equation 4, and determined by the epipoles (*e, e'*).

$$l' = F[e]_x l \quad \& \quad l = F^T [e']_x l' \quad (4)$$

The next step is to identify the projection matrix *P* whose size is 3x4, such that $x = PX$ (see equation 5):

$$\begin{bmatrix} x' \\ y' \\ w \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (5)$$

Being *x* coordinates in the image and *X* coordinates of the world, the matrix *P* is composed of the intrinsic parameters (focal length, camera center) and the extrinsic parameters of the camera (rotation and translation matrix), as well as the values of the distortion. In order to find these parameters, the *cvStereoCalibrate* function is used, which as a result gives the required matrices.

Having the calibration parameters, it can be performed recalibration of the stereo system that will make correction of the images through the parameters of each camera obtained from *F*, the function of openCV that implement this task is *cvStereoRectify*. This function generates the camera re-projection matrix *Q*, necessary to obtain the 3D points of the scene.

RECOGNITION

Once the calibration process has been completed, it is possible to start working with the 3D scene by means of the rectification parameters obtained. Based on these, it must be found the new correspondence between points of the two scenes, the difference between each pair of corresponding points is called disparity and the resulting image of the disparity between the left and right image is called the depth map.

From the disparity (*d*) and by means of the re-projection matrix *Q* it is possible to use the function *cvReprojectImageTo3D* to obtain the point cloud, i.e. the correspondence of the object in the scene with its coordinates x, y, z, based on equation 6.

$$l' = X_i - X_d = \frac{f * b}{Z} \quad (6)$$

Figure 7 shows how is the geometry of the system from which it is easy to infer the previous equation and therefore, the way to obtain Z in relation to a stereoscopic system of parallel optical axes through disparity.

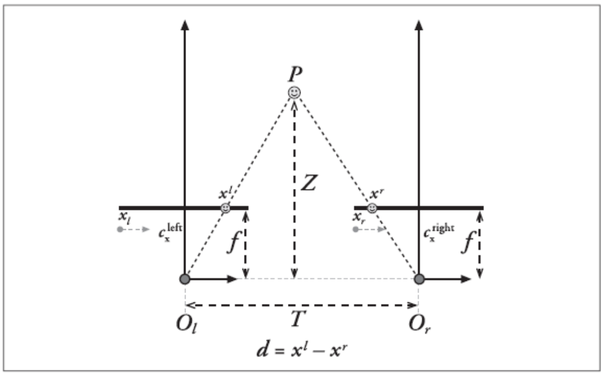

Figure 7. Depth ratio.

In this case, it is the re-projection matrix Q that relates the parameters of the camera with the 3D coordinates by means of equation 7.

$$Q \begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} \quad (7)$$

$$(X/W, Y/W, Z/W)$$

The disparity is obtained by implementing the function of matching points in the scene given by *cvFindStereoCorrespondenceBM*. Figures 8, 9 and 10 show some results of the disparities obtained for three different objects, based on the depth maps found.

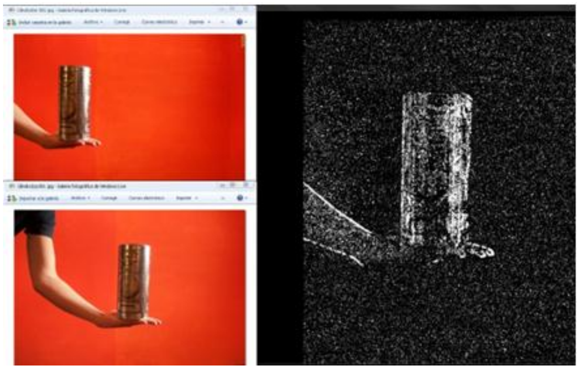

Figure 8. Disparity object 1.

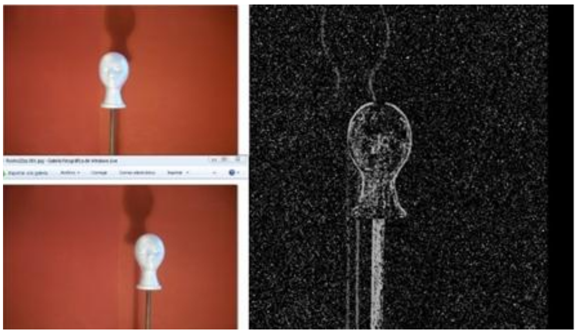

Figure 9. Disparity object 2.

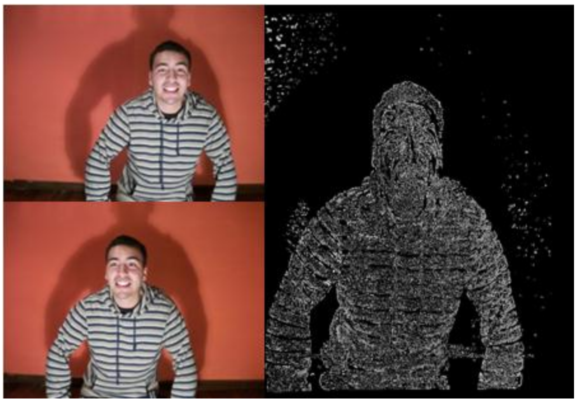

Figure 10. Disparity object 3.

Obtained the point cloud, i.e. the different distances that represent the object, it is possible to use different types of software for reconstruction such as OpenGL or MATLAB, in this particular case it was chosen the latter as a known environment. For this case, from OpenCV, it was obtained the list of points x, y, z that represents the object in the scene, entering them in MATLAB by coordinate and plotting with the plot3 command.

RESULTS ANALYSIS

In Figure 11, it can be seen the result of plotting the point cloud obtained in MATLAB, there is a rotation of the view from the top to appreciate how the point cloud concentrates in the face and fades back following the shadow that generated the source of light in the scene. Similarly, the concavity generated by only having frontal views of the scene can be appreciated towards the back of the concentration of the point cloud. By counting the number of points concentrated in the face area, a total of 90% of these are obtained, the rest are dispersed backwards from the concentration, allowing to maintain a sufficient grouping for the reconstruction.

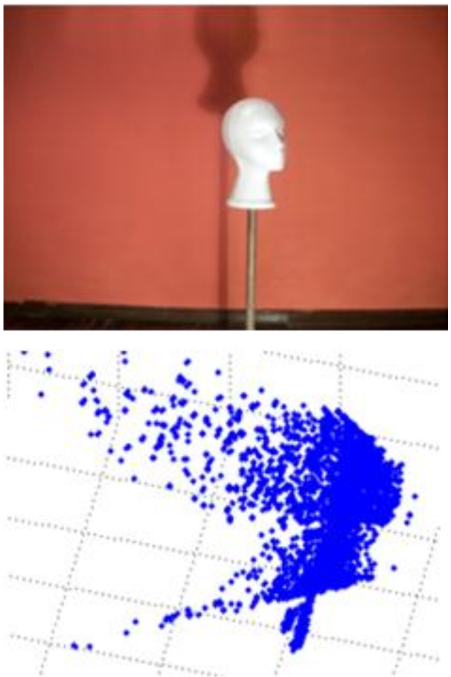

Figure 11. Point cloud of Disparity object 2.

Figure 12 shows the result of applying triangulation to the previous image after randomly reducing the number of points in order to apply Delaunay. This has been rotated since the union of all the points by the function makes it difficult to finally appreciate the relief of the image according to the perspective from which it is seen.

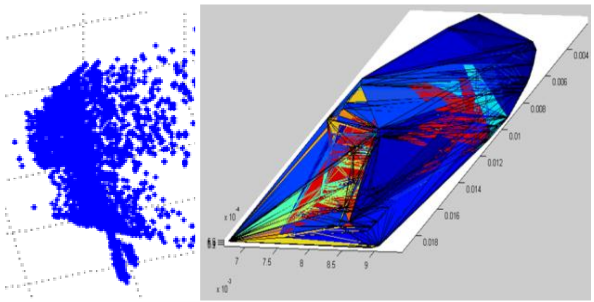

Figure 12. Delaunay triangulation for object 2.

In such a way that it can be appreciated that from the amount of points used for visualization, it is possible to infer the shape of the object. From here, it is feasible to reduce the number of points to be used for reconstruction by other triangulation techniques. The result of these depend on the final location of the points to be operated.

Given the disparity map, which graphically allows inferring results with respect to the resulting point cloud, it is clear that if a disparity is not obtained that clearly identifies depth levels, it will be difficult to discern the figure of interest from the rest of the scene.

CONCLUSIONS

The calibration pattern plays a fundamental role in the calibration of the camera, it requires that the points of the corners are well defined, as well as the contrast or intensity between the black boxes and the white boxes. The size of the tables also influences the accuracy of the calibration, the greater the better results.

Eliminating the background of the scene is definitive since it introduces many points that, for the purpose of reconstruction, are noise and generate too much computational load.

To facilitate the identification of corresponding points, it is recommended that the object of interest in the scene does not have too uniform colors, if necessary, the scene should be adapted with shadows and lights that highlight the main features.

When performing the 3D reconstruction process, by Delaunay triangulation method, it can be seen the outline of the object of interest according to the perspective used for the visualization, this is because the technique used triangulates all the points, which can generate misinterpretations in the representation.

REFERENCES

[1] N. Heide, C. Frese, T. Emter and J. Petereit, "Real- Time Hyperspectral Stereo Processing for the Generation of 3D Depth Information," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 2018, pp. 3299-3303. doi: 10.1109/ICIP.2018.8451194.

[2] V. On, A. Zahedi and B. Bhanu, "3D Reconstruction of Phase Contrast Images Using Focus Measures," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 2018, pp. 101-105. doi: 10.1109/ICIP.2018.8451163.

[3] D. Christie, Cansen Jiang, D. Paudel and C. Demonceaux, "3D reconstruction of dynamic vehicles using sparse 3D-laser-scanner and 2D image fusion," 2016 International Conference on Informatics and Computing (ICIC), Mataram, 2016, pp. 61-65. doi: 10.1109/IAC.2016.7905690.

[4] S. El Hazzat, A. Saaidi and K. Satori, "Structure from motion for 3D object reconstruction based on local and global bundle adjustment," 2015 Third World Conference on Complex Systems (WCCS), Marrakech, 2015, pp. 1-6. doi: 10.1109/ICoCS.2015.7483225.

[5] S. Zhang, H. Yu, J. Dong, T. Wang, Z. Ju and H. Liu, "Automatic Reconstruction of Dense 3D Face Point Cloud with a Single Depth Image," 2015 IEEE International Conference on Systems, Man, and Cybernetics, Kowloon, 2015, pp. 1439-1444. doi: 10.1109/SMC.2015.255.

[6] E. Lilienblum and A. Al-Hamadi, "Pattern Optimization for 3D Surface Reconstruction with an Active Line Scan Camera System," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 2018, pp. 3159-3163. doi: 10.1109/ICIP.2018.8451807

[7] Ningqing Qian and Chao-Yang Lo, "Optimizing camera positions for multi-view 3D reconstruction," 2015 International Conference on 3D Imaging (IC3D), Liege, 2015, pp. 1-8. doi: 10.1109/IC3D.2015.7391816.

[8] S. Wang, X. Shen and J. Liu, "Dense Optical Flow Variation Based 3D Face Reconstruction from Monocular Video," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 2018, pp. 2665-2669. doi: 10.1109/ICIP.2018.8451742.

[9] N. O. Valencia, L. F. V. Londoño, M. A. Jinete and R. Jiménez, "Movement detection for object tracking applied to the InMoov robot head," 2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA), Bucaramanga, 2016, pp. 1-6. doi: 10.1109/STSIVA.2016.7743328