Jurnal Mantik, 7 (1) (2023), ISSN 2685-4236 (Online)

Published by: Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Measuring the distance of crowd movement objects to the camera using a stereo camera calibrated with object segmentation histogram of oriented gradient

Bintang Eka Putera[1], Singgih Jatmiko[2] , Ary Bima Kurniawan[3]

[1,2]Faculty of Computer Science and Information Technology, Gunadarma University, Indonesia

[3]Faculty Of Industrial Technology Gunadarma University, Indonesia

ARTICLE INFO

Article history:

Received Apr 24, 2023
Revised Apr 27, 2023
Accepted May 4, 2023

Keywords:

Calibrated Stereo Camera;
Crowd Movement Object;
Euclidian Distance camera;
Histogram of Oriented Gradient;
Measurement object;

ABSTRACT

Measuring the distance of objects to human objects is currently under development. In its development, a lot of research on measuring object distances was carried out in developing security systems and surveillance systems, one of which was in security in the environment of many human objects or crowds. This study uses the object segmentation method using the Histogram of Oriented Gradient feature to segment crowd objects. In determining the value of the distance based on information using a segmented object centroid. Calculations are performed using the Euclidian Distance calculation method to find the shortest distance between the centroid of the bounding box and the camera. The results of this study from object distance can distinguish human objects that have crowds with the best accuracy with a measurement error of 5.7%. The research have conclusion that the main findings produced can be used to produce an accurate human crowd object recognition system that is able to provide information on the value of the object's distance to the camera when the object.

Corresponding Author:

Bintang Eka Putera
Faculty of Computer Science and Information Technology
Gunadarma University
Jl. Margonda Raya Number. 100, Depok 16424, West Java
Email : bintang_ep@staff.gunadarma.ac.id

1. INTRODUCTION

Measuring object distances is information needed in several fields, such as surveillance and security systems, industry, navigation, robotics and in smart vehicle applications (Bailo et al., 2018). In general there are two ways of measuring the distance of objects, namely: active and passive. The active measurement method measures the object's distance by sending multiple signals to the object. The Previous research have Several methods that are often used include: radar, infrared sensors, radio signals, ultrasonic sensors, lasers and others (Ji, S et al., 2020). The research of sensors in the active object distance measurement method is very vulnerable to environmental conditions, such as: temperature, fog and noise such as wave disturbances in the room. The problem with this method is that it only produces information on object distances without being able to measure object geometry (Wang et al., 2018).

The another previous research have method is a passive measurement that only receives information on the position of objects without transmitting signals, usually with information on light intensity (Bian et al., 2017). This is one of the developments in computer vision systems that use visual image information of an object.

This research was conducted to calculate object distances using visual information with a pair of images from two cameras called stereo images. The correlation between the position of the object in the two right and left images provides the required distance information. Research develops stereo vision and triangulation concepts (Gil et al., 2018), and has a gap in using crowd objects as data, because crowd objects have changing values from video images every time, so they can know the ability of stereo vision concepts from objects that have changes against time. This research is expected to be able to measure the mass distance of objects with a stereo calibration camera so that they can detect objects with a small error value.

2. RESEARCH METHOD

2.1 Image

Image is visual data that represents the spatial distribution of physical quantities such as light intensity and spatial frequency of an object. This information is represented by components such as brightness, color and border (Bailo et al., 2018). Digital images are formed by a collection of dots called pixels (pixels or picture elements). Each pixel in an image is represented by binary data. This data can be a single value or a combination of several values (Gil et al., 2018). Each pixel has a position coordinate. The coordinate system used to represent digital images is shown in Figure 1

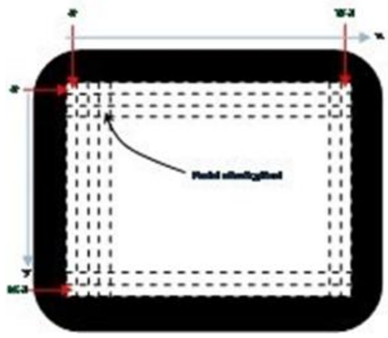

Figure 1. MxN sized image coordinate system

2.2 Stereo Vision And Stereo Geometry Of The Camera

Stereo vision system (stereo vision) is a field related to determining the three-dimensional structure of a scene from two or more digital images obtained from different angles (Chai et al., 2019). The definition of a stereo camera is a camera with the same type and specifications and is installed in a straight line both in the horizontal and vertical planes. Distance measurements can be made when the object is in a point of view that overlaps between the two cameras.

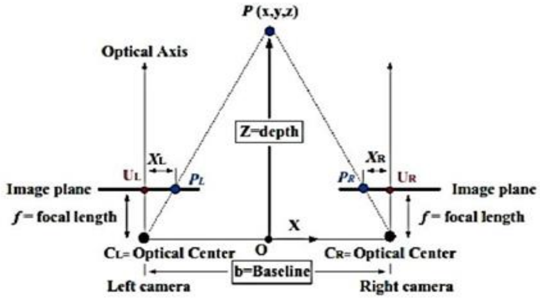

Figure 2. Epipolar geometry with parallel optical axes

Figure 2 is the camera's stereo coordinate system which is assumed to be in the midpoint between the coordinate systems of the right camera and the left camera

2.3 Camera Calibration

The camera calibration function in stereo vision is to determine the geometric parameters of the lens and evaluate both the performance and stability between the lens and the camera. To calibrate the camera, we have to extract all the corner points from each image and then calculate the intrinsic and extrinsic parameters of the camera. Intrinsic parameters show the internal characteristics of the camera, such as: focal length, image center point, tilt coefficient, and lens aberration parameters, while extrinsic parameters, such as: rotation and translation orientation that connect the two cameras (Rohac et al., 2015)

2.4 Image Rectification

Image rectification is a process for projecting two or more images into one image field. This process corrects lens aberrations by transforming the image into a standard coordinate system (Semeniuta, 2016). Because the two cameras capture images from different angles, the resulting images are in different planes. Rectification can be divided into two types, namely: rectification on calibrated cameras and rectification on image results from cameras that are not calibrated

On a calibrated stereo camera, image rectification projects two images onto one image plane so that the same features in each image appear on the same line. This image projection makes the image look like two cameras in a parallel position. In rectification with an uncalibrated camera, the intrinsic and extrinsic parameters of the camera are unknown, so this process requires several points that correspond between the two images. to generate these correspondence points, we have to collect similar pixel points from both images and then select the possible combinations between them. Image matching can use the SURF, SIFT, HARRIS, SAD features and others.

2.5 Histogram Of Oriented Gradient

HOG (Histogram of oriented gradient) aims for object recognition, information in locally distributed form of intensity gradient or edge direction even without precise information about the location of the edge. The image is divided into smaller parts into sub-images called cells. The flow of the detection process using HOG can be explained by the diagram in Figure 3.

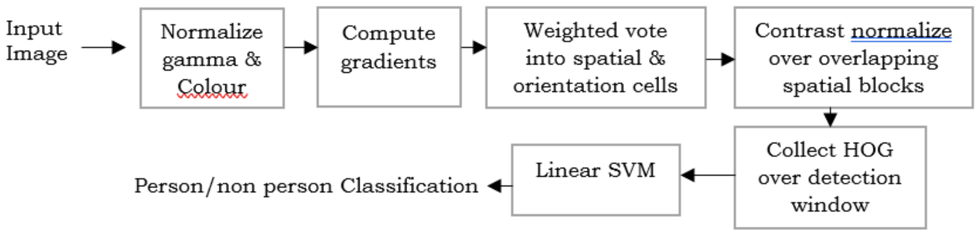

Figure 3. The HOG method

These cells can be rectangular (R-HOG) or circular (C-HOG). Histograms of edge orientation are collected in these cells. The combined input histograms are used as feature vectors that describe objects. To provide better invariant lighting (lighting, shadows, etc.) cells are normalized in a larger area by combining several cells into blocks

3. RESULTS AND DISCUSSIONS

This research begins with image acquisition functioning during the camera calibration treatment for taking crowd movement objects to identify the object distance. Acquisition parameters in Table 1.

Tabel 1. Data collection parameters

	Crowd Movemment Object	
Parameters	Right Camera	Left Camera
Resolution	640 x 480	640 x 480
Duration	16 second	16 second
Total *frame*	160	160
Frame rate	10 fps	10 fps
Distance Camera	12 cm	12 cm
Number of calibration images	19 pairs	19 pairs
Light intensity	340 lux	340 lux
Observed distance	The distance when the crowd of objects is moving	The distance when the crowd of objects is moving
Field conditions	Crowded object	Crowded object

Table 1 shows the parameters for data retrieval, using two cameras to determine the parameters for the data retrieval process. Segmentation on objects using the HOG feature which is used for classification of human density subregions. The classification model used for training is 128x64 pixels in size. The image used to train the model includes the background pixels around the person. Therefore, the actual size of the detected crowd is smaller than the size of the training image.

Tabel 2. HOG features in crowd subregion classification

HOG Feature	Value
Classification Model	[128 64]
Classification Threshold	2
Window Stride	[8 8]
Scaling Factor	1,05
Maximum Size Bounding Box	[384 192]

The classification threshold used is worth 2. This threshold is used to control whether a sub-region will be classified as human. The higher the threshold value, the greater the classification requirements. The influence of this threshold value has a value of 80% to 90% on the success of the existing detection process. The next parameter is windowstrides a, the windowstrides used are of size [length] = [8 8] which "slide" across the image to search for human crowd subregions. The smaller the windowstride, the more windows that need to be evaluated but requires a higher computational load.

The scaling factor used is 1.05. Decreasing the scale factor can improve detection accuracy. However, it increases computation time. Then the maximum size of the human subregion displayed in the form of a bounding box is set to 384 pixels long and 192 pixels wide. In the total number of trials, the 135th frame of the human crowd subregion was detected in a bounding box with position and size [x y width height] = [256 22 173 346]. By using the HOG features, to detect crowds get a segmentation accuracy of 95%

The next stage is the parameter for calculating the distance between the camera and the crowd object

Tabel 3. Parameters for calculating the object distance

HOG Feature	Value
Bboxes sizes	[256 22 173 346]
Centroid bboxes	[343 195]
Centroid idx	139047
Centroid 3D	[2,19e+02 -35,06 3,193+03]
Distance	3.200 mm

The parameter values for the distance determination stages in table 3 show that segmentation was successfully carried out on the value of bounding boxes [256 22 173 346]. The value of the center point of the bounding box is at the coordinates [x y] = [343 195] mm and the centroid index value is 139047. The coordinate value of the 3-dimensional centroid of the object in the 135th frame is [x y z] = [2.19e+02 -35.06 3 ,19e+03] mm. Then the distance value is 3.2 meters, while the actual distance value is 3 meters. In this 135th frame there is no measurement deviation that is too far because the position of the object is not too far from the camera. The results of measuring objects as follows

Table 4. The results of measuring Crowd Movement Object distances

No	Real Distance (m)	average measured distance (m)	distance difference (m)	Error (%)
1	6.37	6.22	-0.15	2.35479
2	6.01	5.93	-0.08	1.33111
3	5.55	5.11	-0.44	7.92793
4	5.33	5.23	-0.1	1.87617
5	5.00	4.75	-0.25	5
6	4.67	4.33	-0.34	7.28051
7	4.21	4.09	-0.12	2.85036
8	4.01	4	-0.01	0.24938
9	3.73	3.43	-0.3	8.0429
10	4.57	4.23	-0.34	7.43982
11	4.77	4.11	-0.66	13.8365
12	5.89	5.23	-0.66	11.2054
13	3.99	3.97	-0.02	0.50125
14	3.46	3.22	-0.24	6.93642
15	3.33	3	-0.33	9.90991
Error rate				5.7823

Table 4 shows the results of measuring Crowd Movement Object distances, by comparing distances using a calibrated camera with real distances, the total error of this comparison reaches 5.7 percent. This study describes the main factor causing the biggest influence on measurements on moving crowd data is due to the low processor which results in video acquisition, the camera does not start and end at the same point. So that requires manual cutting of frames which causes some frames not to be paired with the frame they should be. . The measurement error is influenced by several factors, namely, the data collection method, environmental conditions, lighting factors, and projection errors during camera calibration.

The method of data collection is an important factor in measurements using this system. In measurement data when moving objects cause measurement errors because the object dimensions are different, causing errors in centroid calculations, different pair coordinate positions, and differences in light intensity resulting in differences in pixel values. in both images.

Environmental conditions also affect the level of accuracy in this study. The conditions of the study site were not sterile, causing noise caused by the presence of foreign objects. The noise by these foreign objects affects the segmentation process and inaccuracies in centroid calculations because the pixels of these foreign objects mingle with the object pixels in a bounding box. Apart from the influence of foreign objects in the research location, lighting is also an important factor in this study.

4. CONCLUSION

Research shows that the main findings produced can be used to produce an accurate human crowd object recognition system that is able to provide information on the value of the object's distance to the camera when the object is stationary with a measurement error of 5.7%. This research has limitations in using a normal processor, this research will have maximum results by using a high-speed processor so there is no need to trim videos manually which results in errors in video frame pairing. This research contributes by developing a calibrated stereo camera method so that it can measure distances in all conditions, both moving and stationary objects. This research can be developed by classifying mass behavior by distinguishing normal or abnormal masses, such as riots, disasters or other abnormal things so that it can assist security forces in detecting crowd behavior.

REFERENCES

Bailo, O.; Rameau, F.; Joo, K.; Park, J.; Bogdan, O.; Kweon, I.S. Efficient adaptive non-maximal suppression algorithms for homogeneous spatial keypoint distribution. Pattern Recognit. Lett. 2018, 106, 53–60

Bian, J.; Lin, W.-Y.; Matsushita, Y.; Yeung, S.-K.; Nguyen, T.-D.; Cheng, M.-M. Gms: Grid-based motion statistics for fast, ultra-robust feature correspondence. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4181–4190

Chai, X.; Gao, F.; Hu, Y. Mirror binocular calibration method based on sole principal point. Opt. Eng. 2019, 58, 094109. [Google Scholar] [CrossRef]

Carrera, G.; Angeli, A.; Davison, A.J. SLAM-based automatic extrinsic calibration of a multi-camera rig. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, 9–13 May 2011; pp. 2652–2659

Gil, G.; Savino, G.; Piantini, S.; Pierini, M. Motorcycles that see: Multifocal stereo vision sensor for advanced safety systems in tilting vehicles. Sensors 2018, 18, 295. [Google Scholar] [CrossRef] [PubMed][Green Version]

Hong Y, Ren G, Liu E. Non-iterative method for camera calibration.[J]. Optics Express. 2015; 23 (18):23992–4003. doi: 10.1364/OE.23.023992 PMID: 26368490

Ji, S.; Qin, Z.; Shan, J.; Lu, M. Panoramic SLAM from a multiple fisheye camera rig. ISPRS J. Photogramm. Remote Sens. 2020, 159, 169–183

Liu Y H, Jia Q X, Sun H X, Su J. Multi-camera calibration based on coplanar projection of points[C]// Computer Design and Applications (ICCDA), 2010 International Conference on. IEEE. 2010:V1-221– V1-224

Ricolfe-Viala C, Sanchez-Salmeron A J. Camera calibration under optimal conditions.[J]. Optics Express. 2011; 19(11):10769–75. doi: 10.1364/OE.19.010769 PMID: 21643333

Rohac, J.; Sipos, M.; Simanek, J. Calibration of low-cost triaxial inertial sensors. IEEE Instrum. Meas. Mag. 2015, 18, 32–38
Su, P.C.; Shen, J.; Xu, W.; Cheung, S.S.; Luo, Y. A Fast and Robust Extrinsic Calibration for RGB-D Camera Networks. Sensors 2018, 18, 235
Semeniuta, O. Analysis of camera calibration with respect to measurement accuracy. Procedia Cirp 2016, 41, 765–770.
Tsai R Y. A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses[J]. IEEE Journal on Robotics and Automation. 1987; 3 (4):323–344
Tu, J.; Zhang, L. Effective data-driven calibration for a galvanometric laser scanning system using binocular stereo vision. Sensors 2018, 18, 197. [Google Scholar] [CrossRef] [PubMed][Green Version]
Ueshiba T, Tomita F. Plane-based Calibration Algorithm for Multi-camera Systems via Factorization of Homography Matrices[C]// IEEE International Conference on Computer Vision, 2003. Proceedings. IEEE. 2003:966–973
Wang, Y.; Liu, L.; Cai, B.; Wang, K.; Chen, X.; Wang, Y.; Tao, B. Stereo calibration with absolute phase target. Opt. Express 2019, 27, 22254–22267
Wang, H.; Mou, W.; Mou, X.; Yuan, S.; Ulun, S.; Yang, S.; Shin, B.-S. An automatic self-calibration approach for wide baseline stereo cameras using sea surface images. Unmanned Systs. 2015, 3, 277–290
Wang, Y.; Wang, X.; Wan, Z.; Zhang, J. A Method for Extrinsic Parameter Calibration of Rotating Binocular Stereo Vision Using a Single Feature Point. Sensors 2018, 18, 3666.
Wang, Y.; Rajamani, R. Direction cosine matrix estimation with an inertial measurement unit. Mech. Syst. Sig. Process. 2018
Yang J, Liu Y, Meng Q, Chu R. Objective Evaluation Criteria for Stereo Camera Shooting Quality Under Different Shooting Parameters and Shooting Distances[J]. IEEE Sensors Journal, 2015, 15 (8):1–1.
Yang J, Yang Q X, Qin P L. Two-dimensional flexible target for calibrating camera[J]. Optics & Precision Engineering. 2011; 19(5):1134–1142
Yang J, Yang Q X, Qin P L. Two-dimensional flexible target for calibrating camera[J]. Optics & Precision Engineering. 2011; 19(5):1134–1142
Zhang G. Unique world coordiantes based global calibration method for multi-vision inspection system [J]. Journal of Beijing University of Aeronautics & Astronautics. 2006; 32(11):1268–1272.
Zhang Z. A Flexible New Technique for Camera Calibration[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence. 2000; 22(11):1330–1334.
Zhou Z, Huang J, Wang J, Zhang K, Kuang Z. Object-Oriented Classification of Sugarcane Using Time-Series Middle-Resolution Remote Sensing Data Based on AdaBoost[J]. Plos One. 2015; 10 (11).