Realtime Object Distance Measurement Using Stereo Vision Image Processing

B. N. Arunakumari, R. Shashidhar, H. S. Naziya Farheen, and M. Roopa

Abstract In recent years, great progress has been made on 2D and 3D image understanding tasks, such as object detection and instance segmentation. The recent trends in technology driverless cars are making a difference in daily life. The basic principle in these driverless cars is object detection and localization using multiple video cameras and LIDAR and it is one of the current trends in research and development, so attempts to achieve the same on small scale using the available resources. In the proposed method, firstly the stereo images are captured in a dual-lens camera, and secondly, converting the RGB image into a grayscale image. The third step is to apply a global threshold to separate the background, to get the same size of the image using morphological operation. Blob detection is used to detect the points and regions in the image. The fourth step is to detect the object distance and size measurement using the pinhole camera formula. Further, in the proposed work, an effort is made to determine the linear space between the camera and the object from the pictures taken from the camera. Typically, stereo images are used for computation. Binocular stereopsis, or stereo vision, is the capability to derive information about how far left the objects are, grounded uniquely on the comparative places of the object in the two eyes. It depends on both sensory and motor capabilities, using the similar principle the human brain employs, taking two images of the same object taken from two different linearly separated distances. The frame rate of the system can go a

B. N. Arunakumari (✉)
Department of Computer Science and Engineering, BMS Institute of Technology and Management, Bengaluru, Karnataka 560064, India
e-mail: arunakumaribn@bmsit.in

R. Shashidhar
Department of Electronics and Communication Engineering, JSS Science and Technology University, Mysuru, Karnataka 570006, India
e-mail: shashidhar.r@sjce.ac.in

H. S. Naziya Farheen
Department of Electronics and Communication Engineering, Navkis College of Engineering Hassan, Hassan, Karnataka 573217, India

M. Roopa
Department of Electronics and Communication Engineering, Dayananda Sagar College of Engineering, Bengaluru 560078, India

M. D. Borah et al. (eds.), *Big Data, Machine Learning, and Applications*, Lecture Notes in Electrical Engineering 1053, https://doi.org/10.1007/978-981-99-3481-2_2

maximum of up to 15 frames per second. 15 frames per second can be considered as acceptable for most autonomous systems, and it will work in realtime. Effective convolutional matching technique between embeddings are used for localization that leads LIDAR to increase centimeter level accuracy by about 97%.

Keywords Linear regression · Linear distance · Stereovision · LIDAR · Stereo image

1 Introduction

Nowadays, understanding state-of-the-art evolution of object detection and instance segmentation has made significant improvement in 2D image. However, beyond getting 2D enclosing boxes or pixel masks, 3D understanding is eagerly in demand in real world applications such as housekeeping robots, autonomous driving and advance driver assistance system (ADAS) and augmented reality (AR). With rapid advent and development of 3D sensors deployed on mobile devices and autonomous vehicle navigation, 3D data capturing and processing is gaining more and more attention. Studying the rotation and translation parameters—3D object detection and localization with respect to the coordinate system, which classifies the object category and estimates 3D minimum bounding boxes of solid objects from 3D digitized sensor data are attempting to find the linear space among the camera and the object from the pictures taken from the camera. There are two ways to compute the object to camera distance viz. (i) the camera focal length and the object given size, (ii) the point of contact where the object or image meets the ground and the height of the camera. Unlike in the (i) approach, the dimension of object or image is unspecified in the (ii) approach. Typically, the proposed work uses stereo camera for the computation of object to camera distance. However, the binocular stereopsis is the capability to derive an impress of deepness by the superimposition of two images of the same object taken slightly from unlike angles [1]. It depends on the amalgamation of pictorial excitations from matching retinal images into a single image (sensory fusion) and the capability to sustain a single image with curative movement of eyes to carry the fovea around to the essential place (motor fusion), a similar principle in which the human brain employs, is taking two images of the identical object taken from two, unlike linearly separated distances. Once get the left and the right images of the object, apply the algorithm to compute the linear distance.

2 Related Work

In the past few years, couple of techniques have been established for calculating object to camera distance. These methods have been categorized into two types: contact and noncontact approaches. In the contact approach, most of the harvests

have been used to calculate distance. The weakness of this approach is the image or object is destructive. For noncontact approach, many solutions have been proposed namely laser and ultrasonic replication. The drawback of these techniques is image reflectivity is failing in noncontact measurement [2–6]. In [7], the authors have developed an efficient method for object or image camera distance measurement using fuzzy stereo matching algorithm. The algorithm uses a window size of 7×7 pixels within a search range that varies from -3 to $+3$ for accurate computation of disparity. This approach is the optimally moral choice with respect to processing speed and accuracy. But, for realtime applications like autonomous vehicles and robot navigation, the rectification process is necessary, but it is unnoticed in the proposed stereo vision algorithm. In [3], the authors proposed a search detection method that uses depth and edge data and its hardware architecture for realtime object detection. This method has improved the detection accuracy with respect to the sliding window method and its hardware realization on a Field Programmable Gate Arrays (FPGA) is further improved via an application-specific integrated circuit (ASIC) execution or application precise optimization. However, the method has not presented the implementation approaches of three major tasks of the algorithm viz. disparity calculation, classification and edge detection and the influence of algorithms with respect to hardware architecture and performance [8, 9]. According to [10], calculation of eight to thirteen discrepancy images per second with 3D reestablishment on the M6000GPU, attaining a mean square error of 0.0267 m2 in calculating distances up to 10 m. Graphics processing unit (GPU) implementation significantly speeds up the calculation and 3D depth information for shapeless outdoor atmospheres can be produced in realtime, agreeing on suitable close-range plotting of the atmosphere. In addition, the authors have made an attempt to estimate the motion of a stereo camera by decreasing re-projection errors between two successive edges using Random sample consensus and there are no prior knowledge/sensors are needed to estimate the motion of the camera. As many parameters are added to optimize the process which leads to the enhancement of the motion accuracy [11, 12]. However, it is observed that compared to other steps of the stereo matching method, the random sample consensus-based motion is comparably futile. In this paper, an attempt has been made to develop a robust and efficient method to solve such problems using available resources just by taking picture. Furthermore, authors have made an attempt to improve the accuracy and speed of depth estimation through depth from focus and depth from defocus in combination with stereoscope [12, 13]. However, the motion of the image will remain the superior method in terms of cost, if enlarging the baseline is inexpensive than enlarging the lens aperture, enhancements of depth from or depth from defocus by algorithmic improvements are restricted in common implementations by the small aperture dimension. In addition, the system does not ensure the prevention of occlusion and matching problems of monocular structure of focusing and defocusing systems. However, there is a possibility to enlarge the range of focuses that can be segmented to foodstuff, animals and other stuffs.

A novel method to calculate low latency dense-depth maps by means of a particular CPU core of a mobile phone was proposed by [14]. This method can effectively facilitate occlusions for applications viz. AR photos, AR realtime navigation and

shopping that uses only current monocular color sensor. However, this method has the usual constraint of monocular depth approximation systems. When the comparative pose among the recent frame and the designated key frame is indefinite then various components can be impacted. Another constraint comes from hardware limitations that can for occurrence evident in the form of progressing motion blur and shutter artifact. Lastly, as usual in the reflexive stereo works, small surfaced areas are predominantly unclear and performances of implication over them often lead to inappropriate results.

In addition, the estimation of depth map is a significant step in 3D data generation for 3D demonstration method. From the existing methods it has been perceived that the utmost of the literature is done on monocular, sequence and stereo image. Depth map estimation is depending on depth from focus. However, there are complications to approximate the depth of the object at other spaces which are not in intensive areas and also there is an encounter to estimate the depth map from defocused images and 3D translation from that [15, 16]. Furthermore, authors have attempted to estimate low complexity depth and confidence maps based on bloc distinction investigation for close to the pixel implementation and all-in emphasis image rendering. In this method to build all-in focus images the depth criterion with error elimination and noise filtering was proposed by simple median filter of size 21×21 [17]. The median filtering to the unified depth maps is used to observe whether the measured pixel is uncorrupted one and if the pixel is uncorrupted then the next iteration is executed based on 3×3 window size. Thus, a low complexity method to improve depth estimation is at stake, if the numbers of iterations are repeatedly executed based on the size of the image.

Further, attempts have been made to find a distance map for an image capably solved on the linear array with reconfigurable pipeline (LARP) optical bus system [18]. Because of the elevated bandwidth of an optical bus, numerous resourcefully implemented data movement processes are effectively resolved on the LARP model. Hence, this algorithm is completely scalable. However, a real parallel computer founded on the LARP optical bus system may not exist in the near imminent; the algorithms on the LARP optical bus system model are not robust. In [19], authors have developed a methodology for distance map estimation for two-dimension to three-dimension transformation. The developed method is tested under various conditions viz. multiple objects, static cameras, a very dynamic foreground, motion in behind the forefront then minor motion as backdrop. Therefore, this method remains valid for two-dimension to three dimensions in three-dimensional display. However, such a method possesses the limitation as motionless forefront image inside a view is likely to be measured in the process of backdrop as a result of inaccuracy in distance approximation resulting in truthful stereo vision. Moreover, a survey on indoor and outdoor mobile robot navigation has been developed based on structured and unstructured environments [20]. If the objective is to send a movable robot from one organized area to a different organized area, we trust there is adequate accumulated capability in the investigation today to build a movable robot that could do that in a distinctive structure. However, if the objective is to achieve function driven mobile navigation

we are still in the days of yore. Moreover, an automatic method has been developed to estimate the motion of a stereo camera from uninterrupted stereo image pairs. The method has based only on stereo camera and no other sensor or prior knowledge is required to extract point features in the pair of images. Minimization in the re-projection error of successive images leads to good correspondence of the features. The main advantage of this method is that it designs an adapted feature descriptor that can make feature-matching procedure particularly fast while conserving the matching dependability. However, the method included a few features in the optimization process to improve the accuracy of motion estimation [21].

3 Proposed Methodology

First phase of the work was image acquisition using stereo camera. Once the image is obtained next phase was to convert the color space to Gray scale to rescue the amount of computation on 3 layers in case of RGB, and the image is resized to reduce the computation by 24 folds. Now the image is ready to process for further phases. Median filter is useful to eliminate the noise in the image and then threshold to translate the image into binary image. Once the binary image is obtained blob detection is carried out to find the coordinates of the object of attention in the image. Once all these steps are carried out on the two images obtained from the camera, the difference between the pixel values is used to compute the disparity hence the distance as shown in Fig. 1. Fig. 2 explains the proposed method flowchart and, the flowchart, showing the step-by-step procedure of the proposed. Stereo Image capture: Image is captured by Redmi note 4 mobile cameras. Its dual-lens camera yields an effective focal length of 26 mm (in 35 mm full-frame equivalent terms). Its 1/2.9-inch sensor has 1.25 μm pixels. Two images (left and right) are caught by moving the camera in y axis keeping x and z axis constant.

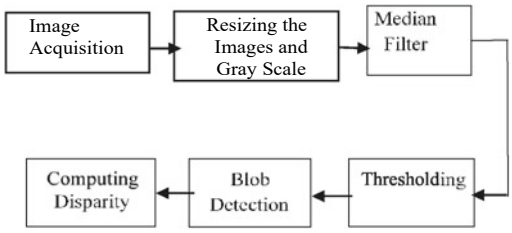

Fig. 1 Object detection and localization

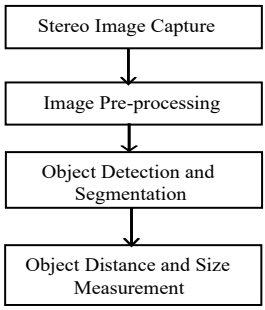

Fig. 2 Flowchart of the proposed method

3.1 Pre-processing of an Image

Pre-processing steps were carried out to meet the requirement of the system. Image is resized to the required dimension. To reduce computation complexity and increase response time RGB image is transformed to grey scale image. Resulting grey scale image is passed through the median filter. It is a non-linear alphanumeric filtering method used to remove noise from an image or signal.

3.2 Object Detection and Segmentation

Background elimination is carried out using thresholding technique. Global thresholding is used to separate the background. Morphological operations are applied. It processes images founded on forms. Morphological procedures adopt constructing features to an input picture, forming an output picture of the equivalent size. The morphological procedure, the cost of every single element in the outcome picture is formulated on an assessment of the conforming element in the incoming picture with its adjacent connects. Blob detection is used to detect points and sections in the image that contrast in properties like illumination or color equated to the adjacent. The main aim is to afford harmonizing information about regions, which is not obtained from edge detectors or corner detectors. It is used to obtain regions of interest for further processing. After object recognition pixel management of both left and right is determined. The difference between pixel coordinates of left and right images gives disparity.

3.3 Object Distance and Size Measurement Using Stereo Camera

Once disparity is found; value of disparity is used in the formula z = (b*f)/a to find the distance. Where z is object distance, d is the disparity, f is the focal length of

camera, b is the steps space among the camera. Once we know the distance object approximate dimension can be found using pin hole camera formula is discussed in Sect. 3.4.

3.4 Mathematical Substantiation for Object Detection and Localization

Figure 3 shows the mathematical proof of the object detection and localization. where d is the disparity i.e., $d = (xl - xr), f$ is the focal length of camera b is the distance between the camera. Triangle APL is similar to triangle $CDLf/z = xl/x$, Triangle BRP is similar to triangle $EFR, f/z = xr/(x - b)$ not ragged.

$$f/z = xl/x = xr/(x - b) \tag{1}$$

$$x - b = zxr/(f) \tag{2}$$

$$zxl/(f) - b = zxr/(f) \tag{3}$$

$$(z(xl - xr))/(f) = b \tag{4}$$

$$Z = bf/((xl - xr)) = bf/(d) \tag{5}$$

$$Z = bf/(d) \tag{6}$$

4 Result and Discussion

Two images (left and right) of LCD monitor are taken from Redmi Note 4 camera separated by 15 cm. The focal length of camera in pixels is 62.40. The disparity is computed to find the distance of the object.

The final result shown in Fig. 4a is the original image and Fig. 4b Processed image. The main objective of the work is to detect the objects using stereo camera and to localize the detected objects in a coordinate system. Finally, we completed the objects with the mathematical evidence. The main applications of the work are to use in self-driving cars, Autonomous robots and Survey drones. Table 1 summarizes different approaches for object detection and localization and their accuracy level. Our proposed method shows better accuracy (97%) compared with the extant methodologies.

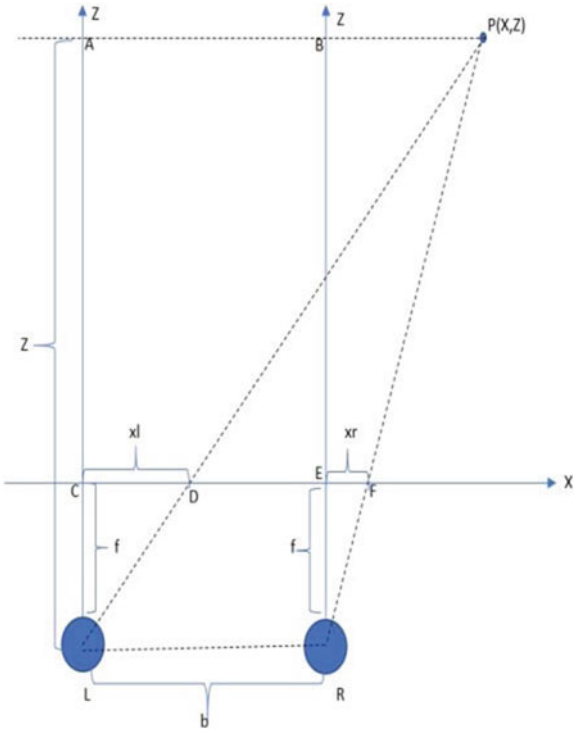

Fig. 3 Detection and localization of object

5 Conclusion

In our proposed methodology an attempt has been made to compute linear distance between the camera and the object using stereo camera and further detected objects are localized in a coordinate system which is economically feasible. The proposed work and its objectives were successfully completed with an accuracy of about 97%. At the end, able to implement the basic concepts of image processing practically and explored specific functions and the mathematics behind it. Major applications of this work are used in self-driving cars, Autonomous robots and survey drones. The recommendation of the future work is to detect multiple objects, detect moving and still objects and classify the object.

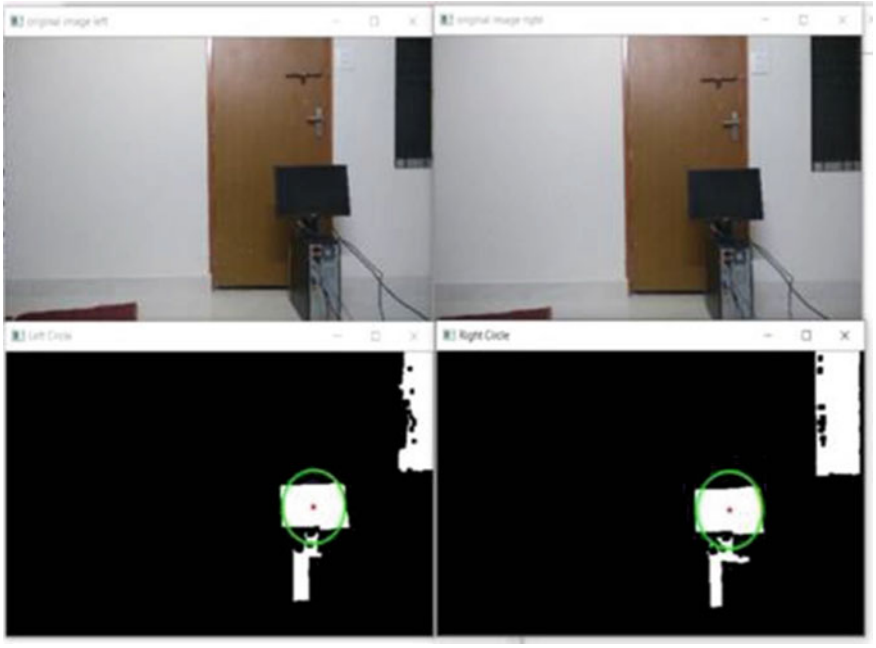

Fig. 4 **a** Original image (top), **b** processed image (bottom)

Table 1 Comparison of different object detection and localization systems

Methods	Accuracy (%)	Remarks
Realtime 3D depth estimation and measurement of un-calibrated stereo and thermal images [14]	90.54	In this approach, authors have made an attempt to perform the rectification of uncalibrated stereo images, realtime depth estimation and realtime distance measurement using stereo images, Intel computes stick webcam and thermal camera used. SAD, triangulation method, epipolar constraints and disparity map for 3D rectification, depth estimation and distance measurement of the object
Distance estimation of colored objects in image [21]	92.1	Determination of accurate HSV values for the lower and upper limits requires a separate experiment. The detection of objects can also be imperfect due to lighting causing different colors in the image with its original color. Imperfection's detection of objects bycolor into account when calculating the distance because the size of the object in the image is different
Realtime object distance measurement using stereo vision image processing (proposed method)	97	Two images (left and right) of the LCD monitor are taken from Redmi Note 4 camera separated by 15 cm. The focal length of camera in pixels is 62.40

References

1. Ma Y, Li Q, Chu L, Zhou Y, Xu C (2021) Real-time detection and spatial localization of insulators for UAV inspection based on binocular stereo vision. Rem Sens 13(2):230. https://doi.org/10.3390/rs13020230

2. Garcia MA, Solanas A (2004) Estimation of distance to planar surfaces and type of material with infrared sensors. In: Proceedings of the 17th international conference on pattern recognition, 2004. ICPR 2004, vol 1, pp 745–748. https://doi.org/10.1109/ICPR.2004.1334298
3. Culshaw B, Pierce G, Jun P (2003) Non-contact measurement of the mechanical properties of materials using an all-optical technique. IEEE Sens J 3(1):62–70. https://doi.org/10.1109/JSEN.2003.810110
4. Klimkov YM (1996) A laser polarimetric sensor for measuring angular displacements of objects. In: Proceedings of European meeting on lasers and electro-optics, pp 190–190. https://doi.org/10.1109/CLEOE.1996.562308
5. Gulden P, Becker D, Vossiek M (2002) Novel optical distance sensor based on MSM technology. Sensors, vol 1. IEEE, pp 86–91. https://doi.org/10.1109/ICSENS.2002.1036994
6. Carullo A, Parvis M (2001) An ultrasonic sensor for distance measurement in automotive applications. IEEE Sens J 1(2):143–. https://doi.org/10.1109/JSEN.2001.936931
7. Chowdhury M, Gao J, Islam R (2016) Distance measurement of objects using stereo vision. In: Proceedings of the 9th hellenic conference on artificial intelligence (SETN '16). Association for Computing Machinery, New York, NY, USA, Article 33, pp 1–4. https://doi.org/10.1145/2903220.2903247
8. Othman NA, Salur MU, Karakose M, Aydin I (2018) An embedded real-time object detection and measurement of its size. Int Conf Artif Intell Data Process (IDAP) 2018:1–4. https://doi.org/10.1109/IDAP.2018.8620812
9. Kyrkou C, Ttofis C, Theocharides T (2013) A hardware architecture for real-time object detection using depth and edge information. ACM Trans Embed Comput Syst 13(3), Article 54 (December 2013), 19 p. https://doi.org/10.1145/2539036.2539050
10. Singh D (2019) Stereo visual odometry with stixel map based obstacle detection for autonomous navigation. In: Proceedings of the advances in robotics 2019 (AIR 2019). Association for Computing Machinery, New York, NY, USA, Article 28, pp 1–5. https://doi.org/10.1145/3352593.3352622
11. Mou W, Wang H, Seet G (2014) Efficient visual odometry estimation using stereo camera. In: 11th IEEE International conference on control & automation (ICCA), pp 1399–1403. https://doi.org/10.1109/ICCA.2014.6871128
12. Wadhwa N, Garg R, Jacobs DE, Feldman BE, Kanazawa N, Carroll R, Movshovitz-Attias Y, Barron JT, Pritch Y, Levoy M (2018) Synthetic depth-of-field with a single-camera mobile phone. ACM Trans Graph 37(4), Article 64 (August 2018), 13 p. https://doi.org/10.1145/3197517.3201329
13. Acharyya A, Hudson D, Chen KW, Feng T, Kan C, Nguyen T (2016) Depth estimation from focus and disparity. IEEE Int Conf Image Process (ICIP) 2016:3444–3448. https://doi.org/10.1109/ICIP.2016.7532999
14. Iqbal JLM, Basha SS (2017) Real time 3D depth estimation and measurement of un-calibrated stereo and thermal images. Int Conf Nascent Technol Eng (ICNTE) 2017:1–6. https://doi.org/10.1109/ICNTE.2017.7947959
15. Valentin J, Kowdle A, Barron JT, Wadhwa N, Dzitsiuk M, Schoenberg M, Verma V, CsaszarA, Turner E, Dryanovski I, Afonso J, Pascoal J, Tsotsos K, Leung M, Schmidt M, Guleryuz O, Khamis S, Tankovitch V, Fanello S, Izadi S, Rhemann C (2018) Depth from motion for smartphone AR. ACM Trans Graph 37(6), Article 193 (November 2018), 19 p. https://doi.org/10.1145/3272127.3275041
16. Kulkarni JB, Sheelarani CM (2015) Generation of depth map based on depth from focus: a survey. Int Conf Comput Commun Control Autom 2015:716–720. https://doi.org/10.1109/ICCUBEA.2015.146
17. Emberger S, Alacoque L, Dupret A, de Bougrenet de la Tocnaye JL (2017) Low complexity depth map extraction and all-in-focus rendering for close-to-the-pixel embedded platforms. In: Proceedings of the 11th international conference on distributed smart cameras (ICDSC 2017). Association for Computing Machinery, New York, NY, USA, pp 29–34. https://doi.org/10.1145/3131885.3131926

18. Pan Y, Li Y, Li J, Li K, Zheng SQ (2002) Efficient parallel algorithms for distance maps of 2D binary images using an optical bus. IEEE Trans Syst Man Cybernet Part A Syst Humans 32(2):228–236. https://doi.org/10.1109/TSMCA.2002.1021110
19. Mulajkar RM, Gohokar VV (2017) Development of methodology for extraction of depth for 2D-to-3D conversion. In: 2017 Second international conference on electrical, computer and communication technologies (ICECCT), pp 1–5. https://doi.org/10.1109/ICECCT.2017.8117848
20. Desouza GN, Kak AC (2002) Vision for mobile robot navigation: a survey. IEEE Trans Pattern Anal Mach Intell 24(2):237–267. https://doi.org/10.1109/34.982903
21. Zhang J, Chen J, Lin Q, Cheng L (2019) Moving object distance estimation method based on target extraction with a stereo camera. In: 2019 IEEE 4th International conference on image, vision and computing (ICIVC), pp 572–577. https://doi.org/10.1109/ICIVC47709.2019.8980940