TEOH MOOI KHEE MEng 2023 UMS

MONOCULAR VISION-BASED POSITION DETECTION OF QR MARKER USING NUMERICAL COMPUTATION

TEOH MOOI KHEE

FACULTY OF ENGINEERING
UNIVERSITI MALAYSIA SABAH
2023

MONOCULAR VISION-BASED POSITION DETECTION OF QR MARKER USING NUMERICAL COMPUTATION

TEOH MOOI KHEE

THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ENGINEERING

**FACULTY OF ENGINEERING
UNIVERSITI MALAYSIA SABAH**
2023

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS TESIS

JUDUL : **MOPNOCULAR VISION-BASED POSITION DETECTION OF QR MARKER USING NUMERICAL COMPUTATION**

IJAZAH : **SARJANA KEJURUTERAAN**

BIDANG : **KEJURUTERAAN ELEKTRIK DAN ELEKTRONIK**

Saya **TEOH MOOI KHEE**, Sesi **2018 – 2023**, mengaku membenarkan tesi Sarjana ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:–

1. Tesis ini adalah hak milik Universiti Malaysia Sabah
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat Salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat Salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/):

☐	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA 1972)
☐	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/ badan di mana penyelidikan dijalankan)
☐	TIDAK TERHAD	

Disahkan Oleh,

TEOH MOOI KHEE
MK1811018T

(Tandatangan Pustakawan)

Tarikh: 03 Februari 2023

(Dr. Yoong Hou Pin)
Penyelia

DECLARATION

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at Universiti Malaysia Sabah or other institutions

25 October 2022

Teoh Mooi Khee
MK1811018T

CERTIFICATION

NAME	:	**TEOH MOOI KHEE**
MATRIC NO.	:	**MK1811018T**
TITLE	:	**MONOCULAR VISION-BASED POSITION DETECTION OF QR MARKER USING NUMERICAL COMPUTATION**
DEGREE	:	**MASTER OF ENGINEERING**
FIELD	:	**ELECTRICAL AND ELECTRONICS ENGINEERING**
VIVA DATE	:	**25 OCTOBER 2022**

CERTIFIED BY;

Signature

SUPERVISOR

Dr. Yoong Hou Pin

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to Dr.Yoong Hou Pin, my academic supervisor, for all his kind support, guidance and advice during my postgraduate research, Vision-based Position Detection of QR Marker Using Numerical Computation. Without his criticism and encouragement, this thesis would not have been conceivable.

In addition, I remain indebted to the lecturers and friends who gave me the necessary impetus I needed to finish this research study. Moreover, I wish to thank all my lab mates who have worked and grown together with me throughout this research journey.

Last but not least, I'd like to express my heartfelt appreciation to my entire family for their unwavering support. Your understanding and inspiration given were what sustained me when undertaking my studies and writing this report.

Teoh Mooi Khee
25 October 2022

ABSTRACT

This research presents the development of a monocular vision-based positioning system using single Quick Response (QR) code as a landmark by numerical computation. The system is applicable to an Eye In Hand (EIH) type grasping automation system in which the camera sensor is positioned at the end of the robot arm to take a snapshot of the artificial landmark. Using the system developed, three-dimensional (3D) position information such as the depth and orientation can be extracted from a two-dimensional (2D) image that consists of a known-size QR code being used as the marker. The four vertices of the QR code are used as the positioning points to determine the 3D information from the underdetermined system with two known parameters, camera's focal length and QR code's dimension. The pinhole imaging theory and similar triangles rule are the fundamental concepts used to analyse the relationship between the 2D image coordinates and the 3D world coordinates of the QR code's vertices. The 3D coordinates of the QR code marker are then determined using the numerical computation model derived with the appropriate guessing parameters and updating rules, and the orientation is extracted using the rotation matrix. To determine the maximum rotation angle attainable at each point in the coordinate plane, the model is simulated using MATLAB platform with 12 different combinations of rotation around three cardinal axes. The simulation's precision is set to be no more than two-degree angle error and distance difference of less than five mm. The simulation results show that an average 77.28% of the coordinate plane can achieved converged results within 30 iterations. The simulation's output is then validated experimentally in an indoor environment by implementing the positioning model into a hardware setup built using a Raspberry Pi, a camera module and a fixed-dimension QR code marker. In comparison to the simulation results, the experiment results have been verified with satisfactory outcomes of average 73.15% resemblance was attained among the rotation combinations tested. Based on the model simulation and experimental data, the QR code's 3D information in terms of distance and orientation extracted from single image captured at one optical point has been validated successfully.

ABSTRAK

PENENTUDUDUKAN KOD QR BERASASKAN VISI MONOKULAR DENGAN MENGGUNAKAN KAEDAH BERANGKA.

Kajian ini membentangkan penggunaan kaedah berangka dalam perkembangan sistem penentudukan berasaskan visi monokular dan hanya mengguna satu Kod Respons Pantas (kod QR) sebagai mercu tanda buatan. Sistem ini sesuai untuk automasi cengkaman jenis Eye In Hand (EIH) di mana kamera yang berfungsi untuk menangkap gambar dipasang pada hujung lengan robot. Selain itu, sistem tersebut dapat menghitung kedudukan dimensi tiga (3D) seperti kedalaman dan sudut kod QR dengan menggunakan gambar dimensi dua (2D) kod QR yang diketahui saiz. Empat bucu kod QR tersebut digunakan sebagai titik kedudukan untuk mengira informasi 3D daripada sistem tidak tentu yang hanya mengandungi nilai panjang focus kamera dan size kod QR. Model kamera pinhole atau kamera lubang jarum dan segitiga serupa merupakan konsep asas dalam menganalisis pemetaan koordinat 3D kod QR dalam ruang ke koordinat gambar 2D. Di samping itu, model kaedah berangka dengan nilai anggaran dan rumus pelelaran yang bersesuaian digunakan untuk menyelesaikan kordinat 3D kod QR manakala sudut rotasi diperolehi melalui matriks rotasi. Seterusnya, model tersebut disimulasi dengan 12 jenis kombinasi putaran mengelilingi paksi-x, -y dan -z melalui MATLAB untuk dapatkan sudut putaran maksima yang boleh dicapai pada setiap lokasi. Ralat sudut putaran tidak melebihi dua darjah dan ralat jarak kurang daripada 5mm merupakan piawaian yang diterima untuk simulasi. Keputusan simulasi sehingga 30 lelaran menunjukkan bahawa lebih daripada 77.28% purata kawasan analisis dapat mencapai piawaian yang diterima. Eksperimen kemudian dijalankan di dalam persekitaran terkawal terhadap perkakasan model tersebut yang dipasang dari satu Raspberry Pi, satu kamera dan satu kod QR yang telah ditetapkan saiz. Perbandingan antara keputusan simulasi daripada MATLAB dan hasil eksperimen yang diuji menunjukkan bahawa pencapaian eksperimen adalah memuaskan sedangkan purate 73.15% hasil eksperimen telah mencapai persamaan. Berdasarkan keputusan yang diperoleh daripada simulasi dan eksperimen, model yang dibina untuk menghitung 3D informasi seperti kedudukan and orientasi kod QR menggunakan satu gambar yang diambil pada satu lokasi telah disahkan.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

(X_{Li}, Y_{Li}, Z_{Li})	-	Local coordinate of point i
(X_{wi}, Y_{wi}, Z_{wi})	-	Real-world coordinate of point i
(x_i', y_i', z_i')	-	Image coordinate of point i
(x_i, y_i, z_i)	-	3D coordinate of point i
$Point\ i$	-	QR code's vertices i
f	-	Camera's focal length
L_i	-	Actual length i
$L_{i,\text{caculated}}$	-	Calculated length i
$diff_i$	-	Difference between calculated and actual i
$Z_{w,min}$	-	Minimum working distance
$Z_{w,max}$	-	Maximum working distance
α	-	Angle of rotation around x-axis
β	-	Angle of rotation around y-axis
γ	-	Angle of rotation around z-axis
d, i	-	Constant
n	-	Number of iteration loop
n_{max}	-	Maximum number of iteration loop
tf	-	Tuning factor of numerical computation
$tolerance$	-	Error tolerance for calculation
$c_{(i)}$	-	Cosine
$s_{(i)}$	-	Sine
opp	-	opposite of triangle
adj	-	adjacent of triangle
hyp	-	hypotenuse of triangle
$R(\alpha, \beta, \gamma)$	-	Rotation Matrix
$R_x(\alpha)$	-	rotation around the x-axis
$R_y(\beta)$	-	rotation around the y-axis
$R_z(\gamma)$	-	rotation around the z-axis
x_{range}	-	range of translation in x-axis
y_{range}	-	range of translation in y-axis
z_{range}	-	range of translation in z-axis

LIST OF ABBREVIATIONS

1D	-	One-Dimensional
2D	-	Two-Dimensional
3D	-	Three-Dimensional
6DOF	-	6-Degree-of-Freedom
AI	-	Artificial Intelligence
cv2	-	OpenCV
EIH	-	Eye In Hand
EPnP	-	Effective Perspective-n-Point
ETH	-	Eye To Hand
IoT	-	Internet of Things
FFC	-	Flexible Flat Cable
FOV	-	Field of View
PIL	-	Python Image Library
QR code	-	Quick Response code
RAM	-	Random-Access Memory
V1	-	Version 1
V2	-	Version 2

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Research Background

For decades, automation technologies have played an indispensable role in uncountable fields. Day by day, it evolves and infiltrates the industry and our daily lives. As the Covid-19 pandemic strikes, systems with minimal human interaction, such as online grocery shopping and autonomous warehouses using pick-cart robot, are becoming the new norm of life. Automation will take over repetitive and physically demanding tasks such as loading and unloading packages at a fixed, standardized location. Even though the process is simple and fixed, the automated system still needs support from sensing technologies to complete the tasks efficiently.

Sensing techniques can be categorized into contact and contactless types. Both types can be combined with machine hardware to provide feedback information needed by a partially or fully autonomous system. Among the non-contact types, vision-based sensor technologies are preferred and recommended. Numerous researchers have mentioned that vision-based systems are one of the most promising technologies for various applications, such as machining (He *et al.*, 2021), localization (Xin *et al.*, 2019), healthcare (Ahad *et al.*, 2019), and civil engineering (Dong & Catbas, 2021).

According to the review, vision-based sensing technologies are widely used for localization but have its limitations. Positioning and object detection tasks require two or more optical sensors, known as stereo vision or multi-camera systems, to capture positional information accurately. Some researchers can perform localization and positioning with just one camera, but the system requires at least two photos captured from different optical points (He *et al.*, 2017). Multiple images are involved in obtaining the location and orientation of the target object. On the other hand, obtaining depth information from a single two-dimensional (2D) image is challenging. To date, only one research has been found to perform 3D positioning from an image consisting of two Quick Response Code (QR code) landmarks (Pan *et al.*, 2020). Finding depth information using a single 2D image with a single QR code landmark is still an open issue. This is because solving the 3D positional information, including six unknowns, and using four reference points is considered an underdetermined condition. It involves complex and extensive calculations to identify the unknowns.

On the other hand, 2D barcode technologies such as QR codes are becoming popular in positioning systems. QR codes are used for various functions, such as information storage (Babu & Markose, 2018), object recognition (Yu *et al.*, 2019), and artificial landmarks (Beck & Kim, 2017). Using QR codes as artificial landmarks in a vision positioning system can simplify the target recognition and searching process. This is because QR code is easy to recognize and decipherable even when thirty percent of the code is damaged (DensoWave, no yr.). Furthermore, the QR code landmark can be placed anywhere, either on the target object itself and move with it (Yu *et al.*, 2019), or on a fixed working location, such as the ceiling, floor or wall (Cavanini *et al.*, 2017). So, it is an excellent option to use QR codes as artificial landmarks for indoor warehouses, as they are now widely used in product packaging and have the potential for three-dimensional (3D) positioning.

1.2 Problem Statement

Extraction of positional information, such as the orientation and distance of the robot gripper to grocery goods, then feedback to the robot is a vital process to automate a pick-robot without any human control. The vision positioning system is one of the common techniques that can extract 3D information in a non-contact way.

Research has shown that positioning and orientation detection of objects using vision-based automation can be done effectively without human interruption. However, most of the research is based on stereo vision-based systems, which involve a complicated calibration process of multi-camera measurement. All the cameras must be synchronized. It will be a challenging task for general users to set up the system without any prior knowledge on the camera calibration process. Besides, the stereo vision system has restricted measurement range and distance when decreasing the distance between cameras as the overlapped field of view becomes smaller. A monocular vision system can avoid the limited field of view issue, as its simpler structure that requires fewer calibration processes than a cost-prohibitive stereo vision system. However, changing the stereo to a mono-vision system makes information extraction, such as depth location, challenging. The 3D positional information is unable to be extracted completely from 2D information as information loss when dimension reduction mapping happens, or the real-world 3D information is mapped into a 2D surface. The automation needs to capture photos from distinct points, and it requires complex computation for the multi-image matching process to get the required information. Image matching error happens for the image pre-processing process involving more than two images and affects the system's performance and accuracy.

On the other hand, many researchers have been found to apply artificial landmarks to their positioning and navigation systems to aid the computation process. However, the landmarks used in their positioning system are customized and not available to the public. It takes time to design and set up such a customized landmark system. The landmark has to be designed using recognizable shapes or images and attached to a known location. Thus, the function of the customized marker will be limited to acting as a positioning marker, which is not economical. Application of non-customized artificial landmarks into a vision positioning system is therefore suggested. The 2D barcode technologies, specifically QR codes, can be used as landmarks for object identification and to determine the good's position and orientation. Nowadays, the adoption of QR codes has become common. Utilizing ready-printed QR codes on product packaging can expand the QR code's functionality while saving the time and cost of setting up the landmarks. Besides, the QR code is easy to recognize so it can simplify the object recognition and searching process.

In recent years, the number of researchers using QR codes as landmarks has increased. QR codes are used as artificial landmarks for positioning purposes and to store information for identification when necessary. However, a few studies have used QR codes in conjunction with a monocular vision system for pose estimation purposes (Pan *et al.*, 2020; Atali *et al.*, 2018; Babu & Markose, 2018; Beck & Kim, 2017). Generally, the system captures the images of the QR code markers from at least two different optical points. Then, the depth information is extracted based on the images captured after the correspondence process is carried out (He *et al.*, 2017). From the review, *Pan et al.* is the only team that proposes an approach to perform monocular vision-based 3D positioning based on one single image consisting of two QR code landmarks (Pan *et al.*, 2020). A total of six positioning points were used, and the 2D information extracted is enough to calculate the 3D positional information. However, their method requires adding one more QR code landmark parallel to the existing QR code on the product packaging. This will be tedious work and defeat the aims of utilizing the existing QR code on the packaging.

In short, research using one single QR code with a monocular vision system to perform 3D positioning is necessary. The existing QR code on product packaging can be used as an artificial landmark for the automation's vision positioning system. However, finding the depth information using a single 2D image consisting of a single QR code as the landmark remains unsolved. All four vertices of the QR code can be used as reference points, yet the information extracted is less than the unknown parameters, and the system is considered underdetermined. The process of obtaining the 3D positional information from an underdetermined condition is yet to solve.

1.3 Research Questions

This research aimed to answer the research questions as follows:

a) How to perform 3D positioning from an underdetermined system by using information extracted from a single QR code image captured at one fixed optical point?
b) What is the range of rotation angle solvable using the proposed positioning method, at various translation parameters with reference to the camera?
c) What is the performance and potentiality of the proposed system in practical application to extract the 6-degree-of-freedom (6-DOF) information of a QR marker?

1.4 Research Aim and Objectives

This thesis aims to design and develop a monocular vision-based position detection system for single QR code marker. The underdetermined system will be solve using an iterative-based numerical computation. The research objectives formulated to accomplish this research aim are as follows:

a) To develop a numerical 3D positioning computation to solve the underdetermined system with the information acquired from a single QR code image captured at one fixed optical point.
b) To simulate and determine the range of rotation angle solvable using the proposed positioning method at various translation parameters with reference to the camera via computer aided engineering (CAE) software.
c) To experimentally determine the system's accuracy and range achievability to validate the performance and potentiality of the proposed system in practical application.

1.5 Research Scopes

In this research, development of the 3D positioning model is executed with certain conditions or requirements:

a) Monocular vision system is implemented; only one camera is used.
b) Image is captured from one fixed optical point; only one image is used to extract the 3D positional information for each analysis.
c) Only one QR marker is used in this research and every image consists of only one QR code.
d) Decoding the data and information stored in the QR code is not covered.
e) The camera calibration process is diminished; no distortion parameters need to be determined or adjusted.
f) The QR code marker with a known size is fixed on a flat surface target object where the background is 'simple'.
g) The model is performed and verified in an indoor environment; practical application is not included in this research stage.

h) The QR code position and orientation setup and measurement in the experiment is done manually.
i) Experiments are carried out to measure the 6-DOF parameters of the QR code landmark using the positioning model.
j) Only the QR code's intrinsic parameters is extracted.

Requirements (a), (b), (c), (d) and (e) are justified by the research's focus, deriving a 3D positioning model that extracts the positional and orientation relationship between a single fixed camera and a QR code marker with variable pitch angle and distance, by using a 2D image. A camera sensor captures the QR code landmark at one fixed optical point.

Condition (f) and (g) were added to aid the QR code recognition and detection process. In this research, the experiment setup error is minimised by conducting the experiment in a controlled indoor environment, where the surroundings and background are simple without any object other than the experiment prototype captured in the image. Besides, this research does not include any practical application, such as implementing the positioning algorithm into a gripper system.

In addition, a quick and simple prototype is used to validate the feasibility of the positioning model. Condition (h) is set for the low-fidelity prototype to perform the measurement manually. The data acquisition process is done manually, and minor human error is tolerable.

For conditions (i) and (j), there are 6-DOF involved in this research. The QR code is tilted with different rotation combinations around the three cardinal axes with angle $\alpha, \beta,$ and γ respectively, while the camera is translated in x- y- and z-axis direction away from the camera. On the other hand, the positioning model extracts the QR code's intrinsic parameters, and this research does not cover the derivation of extrinsic parameters.