# Calibration, Recalibration and Self- Calibration of a Hybrib Stereo Camera System

Muhammad Izzul Azri Bin Zainol Azman, Aliza Binti Sarlan, David Fofi, Omar Tahri

# Calibration, Recalibration and Self- Calibration of a Hybrib Stereo Camera System

**Conference Paper** · August 2022

DOI: 10.1109/ICCUBEA54992.2022.10011027



**4 authors**, including:

Muhammad Izzul Azri Zainol Azman
Universiti Teknologi PETRONAS
**2** PUBLICATIONS **8** CITATIONS
SEE PROFILE

David Fofi
University of Burgundy
**168** PUBLICATIONS **1,995** CITATIONS
SEE PROFILE

Omar Tahri
University of Burgundy
**67** PUBLICATIONS **1,134** CITATIONS
SEE PROFILE

# Calibration, Recalibration and Self- Calibration of a Hybrid Stereo Camera System

Muhammad Izzul Azri Bin Zainol Azman, Aliza Binti Sarlan
Computer & Information Sciences Department
Universiti Teknologi PETRONAS
Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia
muhammadizzulazri@yahoo.com, aliza_sarlan@utp.edu.my

David FOFI and Omar TAHRI
ImViA VIBOT
Université de Bourgogne – Centre Universitaire Condorcet
720 Avenue de l'Europe, 71200 Le Creusot, FRANCE
david.fofi@u-bourgogne.fr, omar.tahri@u-bourgogne.fr

***Abstract*— Nowadays, the risen number of losses of hearing and loss of eyesight has become alarming to all doctors and researchers to speed up the finding on how to tackle the problems that occur. The difficulty of transporting the drug through the inner ear and indirect visualization of the surgical targets has become the reason for the risen number of lost hearing and lost eyesight. The risk of the inaccuracy of delivering the drug to the targeted place as the cochlea which is the part of the inner ear organ has led us to conduct this research study. While for retinal surgery, the need for direct visualization of the surgical targets and tactile feedback has also become the factor that this study is focusing on overcoming the need for that. This study is focusing on overcoming the criteria needed for the drug transporting in the inner ear and the retinal surgery by having the accuracy for the drug transportation and with the stereo vision that we are proposing can tackle the visualization problem of the surgical target and the tactile feedback. This study is validated by having a real-time scenario of the estimation rigid transformation between the controller of the drug delivery and the ability to update the location of the drug throughout the delivery process.**

**Keywords— Computer Vision, Hybrid Stereo Vision, 3D Reconstruction**

## I. INTRODUCTION

Nowadays the need for high precision and accuracy for surgery operation or drug transporting in a small part of the human body is needed. A lot of methods have been proposed to suppress this problem. As one of the applications that needed this kind of system is drug delivery inside the human inner ear or as known as the cochlea. The cochlea is the inner part of the ear that plays an important role in hearing. In the inner ear, there are a lot of important organs for hearing which is the cochlea of it and for balance, it is the vestibular system which is embedded deep within the skull close to the brainstem in the petrous bone which is one of the hardest bones in our human body. The cochlea is one of the sensitive and very complex structures of organ in our inner ear which requires very high precision and accuracy for making any volume changes[1]. It has always been a challenge for the physician in the treatment of inner ear disorders as the need to deliver the drug to the inner ear. The challenges can be described as in the inner ear of human consists of the bony labyrinth which is a system of passage that compromise the two main parts which is the cochlea for hearing and the vestibular system for balance. The blood labyrinth barrier (BLB) and round window membrane (RWM) are the reason these parts constitute the barriers of the drug delivery [1]. Despite many other diseases that have caused of higher risk of disease, the need for drug delivery through the human inner ear is to help with otologic practice which used for sensorineural hearing loss (SSNHL) and autoimmune inner ear disease (AIED) treatments [2].

The next application that can be beneficial to this study is retinal surgery which involves the procedure in ophthalmology that involves peeling the retinal membrane to improve human vision. The procedure is vitreoretinal microsurgery which is labelled as the most technically challenging of the minimally invasive surgical technique. This surgery is so challenging due to it requires very high precision because of working in micron-scale targets presented by the retina and the need of maneuvering in a tightly constrained and very fragile workspace which can lead to permanent blindness if there is a mistake or inaccuracy through the whole microsurgery process [3]. With the limitations of the human eye to operate on a single-digit micron-scale target is one of the limitations that can lead to higher risk if it is operated without the vision system assisted.

This study started with the possibility of transporting the drug to the inner ear (cochlea) of a patient by using magnetic particles actuated with the help of a camera for recovering the position of the magnetic particles and the limitations of the human eye in processing single-digit micron targets. Besides that, the need for this vision system is not only limited to drug transportation, as mentioned the retinal surgery is also of the applications that require high-precision vision systems. The application that has been discussed is the application that required non-invasive drug transportation and minimally invasive for retinal surgery.

This study is focused on the calibration, recalibration and self-calibration of the perspective camera and the stereo camera. The application of this system is now focused on two main types of drug transporting in the inner ear (cochlea) and retinal surgery which both require high precision and a microscopic view for higher precision and lower the risk of fatal injury to the patients. The drug transporting particles application is based on the stereo microscope camera that overviews the cochlea for observing the behaviour of the particles respected to the actuator, while the perspective camera will be observing the actuator that is going to control the particles.

While for retinal surgery, the concept of this application setup can be implemented as it is almost the same method just the different changes in the magnetic actuator with the robot surgical tools and the view of the stereo microscope system. As the proposed application setup, we can validate the control law and model the actuators and the particles motion model for the inner ear drug transportation while for the retinal surgery, we can prove the concept of our system able to provide the required requirements. Retina surgery is highly requiring a microscopic view which can view objects less than 250 micrometers as the working area are very small. The concept of our visual system can be implemented as the only thing that needs to be changed is the actuator with the surgery tools.

## II. Hybrid Stereo Camera System Vision

In this section, we are going to discuss the problem statement and the research objectives.

### A. Problem Statements

Despite the increased number of methods to perform drug delivery in human inner ear parts and retinal surgery, the problem of inaccuracy and the increased risk of the procedure has risen throughout the years. As in the past decade, there is a method of drug delivery that has been developed. An article [2] has been published with a comprehensive review on inner ear drug delivery in 2008. Throughout the same year, there is another author reviewed the status of the inner ear delivery vias the intratympanic and intracochlear routes. Since that, the field of inner ear drug delivery has gained a lot of researchers to help develop a method that can increase the accuracy and reduce the risk of the procedure.

While for retinal surgery, the popularity and attention have risen since the ability of robots with cameras able to perform or assist the surgical procedures and produce less risk and high precisions. Plus, the ability of the stereo camera system which can help for three-dimensional reconstruction helps to increase the surgical precision [4].

The enclosed challenges that we are going to face for these two main applications were the difference in the scale between the perspective camera and the stereo microscope camera. As mentioned before this, the setup for this system consists of three cameras divided into one for a perspective view of the overall actuator or the robotic surgical tools. The other two are the stereo microscope system which is to observe the position of the microrobots in their workplace or the position of the microsurgical tools when conducting surgical procedures.

The difference in the scales is meant by the perspective camera will observe the bigger image view of the whole view and the stereo microscope camera will observe only the microrobots area or the surgical area which is at the scales less than 250 micrometers.

The next challenges were the non-overlapping fields of the sensors between the three cameras. This meant by the field of view of each camera is not overlapped between the perspective camera and the stereo microscope cameras. As the overlapping has easily been overcome, the challenges that might arrive are the intrinsic calibration and extrinsic calibration of the microscopes that require a high number precision because of the low field of view and depth of view. This meant by the movement of the calibration pattern is limited as it cannot exceed the number of degrees for the rotation as it will exceed the camera frame. It also occurs with the height of the calibration pattern as, during the calibration, a few sets of the pattern high are trying to be achieved but the stereo microscope camera loses its focus due to a low number of depths of view. As a solution for that, a custom pattern has been designed to overcome the problem, the details of the custom pattern are discussed in the result and discussion section.

As mentioned before this the different scale between the perspective camera and the stereo microscopes camera has become a challenge to this hybrid vision system. This challenge led to difficulties in estimating the rigid transformation and the relative pose between the perspective camera and the stereo microscope cameras. A few techniques have been proposed to overcome the problems which is one of the researchers estimate the pose relative between the rigidly linked cameras by moving them and using the motions that have been formed from the moving to estimate the relative of the pose [5]. This technique has been a great solution for overcoming the problem in general but unfortunately, it does not apply to our hybrid vision system because our system is a fixed system where the three cameras will not be moved. In the following chapter, we will enclose more on this matter and the proposed method to solve the arrived challenges and problems.

### B. Objectives

This hybrid stereo vision system that has been proposed is to achieve a stereoscopic vision system capable of providing 3D flow dense and accurate enough for the 2D or 3D registration stage. The objective of this study is to:

- Calibrate perspective camera (short focal length) and stereo camera which is consist of two microscope cameras (long focal length).
- Recalibrate the three cameras for achieving the homogeneous transformation according to their drift which is the estimation pose between all cameras and the magnetic actuator.
- Self- Calibration of the three cameras is to optimize the intrinsic parameters and extrinsic parameters for ease of use in all conditions. It is to adapt the new intrinsic and extrinsic according to the new drift.

## III. Hybrid Stereo Camera System Methodologies

The methodologies section is the most crucial part that discussing on the equations and techniques that have been used to solve the problem and goal that we are going to achieve. It is starting with achieving the optimum calibration pattern, then achieving the re-calibration of the hybrid stereo vision system, formulating the estimation reprojection equation, optimization of the intrinsic and extrinsic parameters and last but not least, the 3d object reprojection estimation.

### A. Optimum Calibration Pattern Used

Several types of calibration have been used throughout the study for achieving the low reprojection error of 2D point and 3D point from the perspective camera and the stereomicroscope systems. The calibration is performed in a standard way of a calibration process which requires an object with features that present to be detected like a square chessboard, symmetrical circle grid or asymmetrical circle grid. Here we are experimenting with the calibration with two types of calibration patterns that is by using a square chessboard pattern and an asymmetrical circle grid pattern.

The calibration pattern has been designed by me to prove the concept of the estimation of the relative pose between the perspective camera and the stereo microscopes cameras by using the two rigidly linked calibration patterns. The pattern is set to have a different scale with the known coordinates in the plane size of 210 millimetres by 297 millimetres. The need for different scaled between two of the calibration patterns is because the field of view and depth of view of the stereo microscope system and the perspective camera is different.

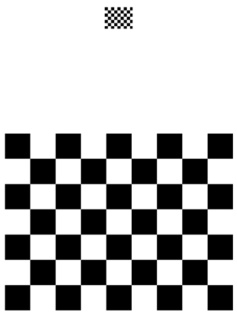

Fig. 1. Square Chessboard with Two Rigidly Linked Pattern

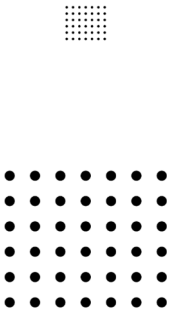

Fig. 2. Symmetrical Circle Grid with Two Rigidly Linked Pattern

Fig. 1 and Fig. 2 is a custom design calibration which is has been used throughout the whole research study for the calibrations of the perspective camera and the stereo microscope camera. The process of designing the pattern has been conducted on open-source software photo editing Inkscape. As mentioned before the two of the patterns have been using the plane size of 210 millimetres by 297 millimetres which will fit the size of general A4 paper for the feasibility of printing. By knowing the size of the plane, each of the patterns has been set to a certain location on the plane which later will help in calculating the homogeneous transformation of the two different sizes and scales of the pattern.

Homogeneous transformation matrices are defined as a matrix that define the position and the orientation of an object [6]. The equation (1) shows the derived homogenous transformation from the two different object position and orientation in a two-dimensional plane which is a combination of rotation matrix and translation vector as follows:

$$HomogenousTransformation = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \quad (1)$$

As we referred to the Fig. 1 and Fig. 2, we can assume that there is no rotation between the small pattern and big pattern of each calibration pattern. The rotation will be as follows:

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

While for the translation vector, there is only the translation for the x-axis and y-axis as the pattern was set on two-dimensional planar only, thus the translation will be as follow:

$$\mathrm{t} = \begin{bmatrix} t_x \\ t_x \\ 0 \end{bmatrix} \quad (3)$$

If we referred to Fig. 1 or Fig. 2, there are two different sizes of pattern are being placed in one plane. The small pattern is labeled as m while for the big pattern are labeled as M. There will be two homogenous transformations being derived between the translation of m to M and M to m. The first homogenous transformation is mTM, which is a transformation from the small pattern to the big pattern as follows:

$$mTM = \begin{bmatrix} 1 & 0 & 0 & -t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

The second homogenous transformation is MTm, which is a transformation from the big pattern to the small pattern as follows:

$$MTm = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & -t_y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

### *B. Re- Calibration of Hybrid Stereo Camera System*

This hybrid stereo vision system that has been proposed is to achieve a stereoscopic vision system capable of providing 3D flow dense and accurate enough for the 2D or 3D registration stage. The objective of this study is to:

- To estimate the relative pose between the perspective camera and the microscope camera 1
- To estimate the relative pose between the perspective camera and the microscope camera 2
- Self- Calibration of the To estimate the homogeneous transformation between the small pattern and the big pattern which was explained in the upper section on how the values of the translation x and the translation y is obtained.

This estimation of the relative pose between the perspective camera and the two microscope cameras are depends on the homogeneous transformation of small pattern and big pattern which is as mentioned before this the possibility of estimating the relative pose between the three camera is by having the two rigidly linked patterns.

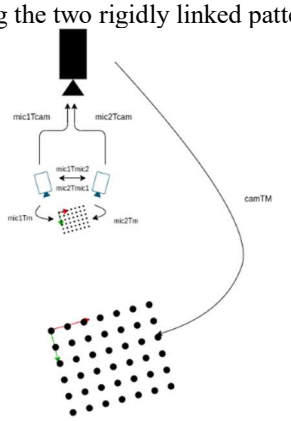


Fig. 3. The Hybrid Stereo Camera System Architecture

Before we started to derive the equations that we develop for estimating the relative pose, I will describe first the notation that we use throughout the research as referred to the Fig. 3. As we can see the black painted camera is the perspective camera which we give the notation for the

intrinsic parameters are $K_{cam}$ while the two other cameras which painted in blue line is the microscope 1 and the microscope 2 with the notation for the intrinsic parameters are $\boldsymbol{K}_{mic1}$ and $\boldsymbol{K}_{mic2}$. The next notation that I am going to describe is the transformation of the camera respect to the pattern which will start first with perspective camera are $cam_{\mathbf{T}M}$. *M* is used to label the big pattern while *m* is used to label small pattern. While for the microscope 1 transformation to the pattern is $mic1_{\mathbf{T}m}$. The microscope 2 transformation to the pattern is $mic2_{\mathbf{T}m}$. The transformation that I declare on the notation for the intrinsic parameters is as follows:

$$\boldsymbol{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

While for the transformation to the pattern is a homogeneous transformation which have the full rotation matrix and translation vector which has been acquired from the extrinsic parameters. The decompose of the transformation is as follows which *i* is the distinguish of different between the each of images:

$$cam_{\mathbf{T}Mi} = \begin{bmatrix} r_{i11} & r_{i12} & r_{i13} & t_{ix} \\ r_{i21} & r_{i22} & r_{i23} & t_{iy} \\ r_{i31} & r_{i32} & r_{i33} & t_{iz} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

$$mic1_{\mathbf{T}mi} = \begin{bmatrix} r_{i11} & r_{i12} & r_{i13} & t_{ix} \\ r_{i21} & r_{i22} & r_{i23} & t_{iy} \\ r_{i31} & r_{i32} & r_{i33} & t_{iz} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}$$

$$mic2_{\mathbf{T}mi} = \begin{bmatrix} r_{i11} & r_{i12} & r_{i13} & t_{ix} \\ r_{i21} & r_{i22} & r_{i23} & t_{iy} \\ r_{i31} & r_{i32} & r_{i33} & t_{iz} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

The equation to acquire the homogeneous transformation in equation (7), (8) and (9) for each of the camera is as follow which the r is the full rotation matrix that has been acquired from the function of OpenCV which is Rodrigues and the t is the translation vector:

$$cam_{\mathbf{T}M} = \begin{bmatrix} cam_r & cam_t \\ 0 & 1 \end{bmatrix} \tag{10}$$

$$mic1_{\mathbf{T}m} = \begin{bmatrix} mic1_r & mic1_t \\ 0 & 1 \end{bmatrix} \tag{11}$$

$$mic2_{\mathbf{T}m} = \begin{bmatrix} mic2_r & mic2_t \\ 0 & 1 \end{bmatrix} \tag{12}$$

With the transformation that we have from each of the camera to their respected pattern and the transformation between rigidly linked pattern, we can now develop the formula for the estimating the relative pose of perspective camera and microscope 1, and perspective camera and microscope 2. For the first transformation of the perspective camera and the microscope 1, it is estimated through the knowing of the relative pose of the perspective camera to the big pattern, microscope 1 to the small pattern and the small pattern to the big pattern. The transformation is formed by the transformation of the microscope 1 to the small pattern, then from the small pattern we know the transformation to the big pattern and finally from the big pattern we can know the transformation to the perspective camera. The flow of the points being estimated as follows:

$$mic1_{\mathbf{T}cam} = mic1_{\mathbf{T}m} \cdot m_{\mathbf{T}M} \cdot cam_{\mathbf{T}M}{}^{-1} \tag{13}$$

While for the second transformation of the perspective camera and the microscope 2, it is estimated through the knowing of the relative pose of the perspective camera to the big pattern, microscope 2 to the small pattern and the small pattern to the big pattern. The transformation is formed by the transformation of the microscope 2 to the small pattern, then from the small pattern we know the transformation to the big pattern and finally from the big pattern we can know the transformation to the perspective camera. The flow of the points being estimated as follows:

$$mic2_{\mathbf{T}cam} = mic2_{\mathbf{T}m} \cdot m_{\mathbf{T}M} \cdot cam_{\mathbf{T}M}{}^{-1} \tag{14}$$

### *C. Reprojection Equation for Hybrid Stereo Camera System*

This subsection is created to explain the process of how we are projecting the points from the homogeneous transformation of the perspective camera to microscope 1 and the perspective camera to microscope 2. The next following subsection is the result of the reprojection error from both homogeneous transformations. For the reprojection equation of the perspective camera to microscope 1 and perspective camera to microscope 2, we have developed two reprojection equations that will transform the detected corners or feature points from the perspective camera to microscope 1 and microscope 2.

For the first reprojection equation, we will use utilize the homogenous transformation of $mic1_{\mathbf{T}cam}$, $cam_{\mathbf{T}M}$, $M_{\mathbf{T}m}$ and by using the intrinsic parameters of microscope 1 with object point of microscope 1which will give us the full reprojection equation was:

$$\boldsymbol{x}_{mic1} = [\boldsymbol{K}_{mic1} 0_{3*1}], mic1_{\mathbf{T}cam} \cdot cam_{\mathbf{T}M} \cdot M_{\mathbf{T}m}, \boldsymbol{Obj}_{mic1} \tag{15}$$

While for the second reprojection equation, we will use utilize the homogenous transformation of $mic2_{\mathbf{T}cam}$, $cam_{\mathbf{T}M}$, $M_{\mathbf{T}m}$ and by using the intrinsic parameters of microscope 2 with object point of microscope 2 which will give us the full reprojection equation was:

$$\boldsymbol{x}_{mic2} = [\boldsymbol{K}_{mic2} 0_{3*1}], mic2_{\mathbf{T}cam} \cdot cam_{\mathbf{T}M} \cdot M_{\mathbf{T}m}, \boldsymbol{Obj}_{mic2} \tag{16}$$

### *D. Optimization of The Intrinsic Parameters and The Extrinsic Parameters*

The result obtained from the previous experiment is by using the intrinsic parameters and extrinsic parameters that we obtained from the last calibration and the homogeneous transformations that we obtained from the mean value of the rotations and translation vectors. As we can see, the results are already good as the average reprojection error is not more than four pixels errors. The way of reprojection error is calculated as mentioned by calculating the difference between the projection point and the original extracted image points. However, the needs of optimization are necessary to reduce the error close to zero as in the real-time situation, there are a lot of factors that can affect the accuracy of the reprojection which will lead to inaccuracy and not able to perform well.

In this section, we are focusing on developing the cost function that will consider optimizing the intrinsic parameters and extrinsic parameters. The cost function will be five as the first one is to optimize the intrinsic parameters of the perspective which will improve the extraction of the $cam_{\mathbf{T}M}$. The main reason why we want to include this first cost function is that the calibration for the perspective camera has a lot of limitations in reaching a certain orientation and pose of the calibration patterns. The need of reaching a certain

orientation and pose is for having a better intrinsic parameter. The next cost function will be optimizing the homogenous transformation of the $cam_{\mathbf{T}M}$ which respect to the microscope 1 and microscope 2. As mentioned in the first cost function, that cost function is solely to improve the intrinsic parameters which we are conducting the calibration without the microscope 1 and microscope for extracting better intrinsic parameters and extrinsic parameters which we can use the intrinsic parameters later for extracting the extrinsic parameters for the calibration that included the microscope 1 and microscope 2.

The next cost function is to optimize the microscope 1 intrinsic parameters and extrinsic parameters for a better reprojection of the points. The cost function will include to optimize the $\boldsymbol{K}_{mic1}$, $mic1_{\mathbf{T}cam}$, and $mic2_{\mathbf{T}cam}$. While for the last cost function is to optimize the microscope 2 intrinsic parameters and extrinsic parameters for a better reprojection of the points. The cost function will be more similar to the last cost function which is used to optimize microscope 1 intrinsic parameters and extrinsic parameters. This four of the cost functions will be added to become one big cost that we are trying to reduce its downs. Plus, with the cost function that has a minimum number of parameters to optimize, this has added to us a more challenger which is known as the bundle adjustments [7].

The goal of this optimization or more accurately bundle adjustment is to refine the three-dimensional feature coordinate and the best parameters for the intrinsic parameters and extrinsic parameters that will reduce the reprojection error by using our custom cost functions. The name "bundle" is for the light rays that leave each of the 3D features and converging on each of the camera centers with the name refers to the "adjusted" for optimally with the respect to both feature and camera positions [7]. The bundle adjustment is usually formulated as the nonlinear least squares problem. The cost function is specially designed to be assumed to be quadratic in the feature reprojection errors [8] and robustness is provided by the explicit outlier screening.

As we proposed before this regarding the cost function parameters we are planning to integrate all of our homogeneous transformations and the intrinsic parameters of each camera that has been in use in the hybrid vision system. We will have four sub cost functions which will fit all of the homogenous transformation and the intrinsic parameters of each camera, and we will have one main cost function that will be the root mean square error between the original extracted feature points from each of the cameras and its reprojection with the homogeneous transformation as follows:

- For the single perspective camera:

$$\boldsymbol{e}_{camSingle} = \sum_{i=1}^{n} \sum_{j=1}^{xe_{camSingle}} \left\| \boldsymbol{x}_{camSingle_{ij}} - f_1\left(\boldsymbol{K}_{camSingle}, cam_{\mathbf{T}M_i}, \boldsymbol{X}_{wcamSingle_j}\right) \right\|^2 \quad (17)$$

- For the perspective camera:

$$\boldsymbol{e}_{cam} = \sum_{i=1}^{n} \sum_{j=1}^{xe_{cam}} \left\| \boldsymbol{x}_{cam_{ij}} - f_2\left(\boldsymbol{K}_{cam}, cam_{\mathbf{T}M_i}, \boldsymbol{X}_{wcam_j}\right) \right\|^2 \quad (18)$$

- For the microscope 1:

$$\boldsymbol{e}_{mic1} = \sum_{i=1}^{n} \sum_{j=1}^{xe_{mic1}} \left\| \boldsymbol{x}_{mic1_{ij}} - f_3\left(\boldsymbol{K}_{mic1}, mic1_{\mathbf{T}m_i}, \boldsymbol{X}_{wmic_j}\right) \right\|^2 \quad (19)$$

- For the microscope 2:

$$\boldsymbol{e}_{mic2} = \sum_{i=1}^{n} \sum_{j=1}^{xe_{mic2}} \left\| \boldsymbol{x}_{mic2_{ij}} - f_4\left(\boldsymbol{K}_{mic2}, mic2_{\mathbf{T}m_i}, \boldsymbol{X}_{wmic_j}\right) \right\|^2 \quad (20)$$

As explained before, the four of the sub cost function is developed to reduce the reprojection error in the two-dimensional image plane of the perspective camera, microscope 1 and microscope 2 by finding the best parameters of the intrinsic parameters and extrinsic parameters with the equation (17) for the single perspective camera, equation (18) for the perspective camera, equation (19) for the microscope 1 and equation (20) for the microscope 2. This optimization is not completed without the main global cost function which is defined as the sum of the four mentioned equations before this as follows:

$$\boldsymbol{e}_{Global} = \boldsymbol{e}_{camSingle} + \boldsymbol{e}_{cam} + \boldsymbol{e}_{mic1} + \boldsymbol{e}_{mic2} \quad (21)$$

Finally, after all of the cost function has been well defined then we can use the algorithm of Levenberg-Marquardt to solve the non-linear least square problem. The Levenberg-Marquardt algorithm is used to find the minimum number of the cost function that can achieve for the optimization purpose. The Levenberg-Marquardt algorithm will require the Jacobians which are the partial derivatives of the extracted image coordinates w.r.t (x, y) the intrinsic and the extrinsic parameters of the camera and the coordinates of the 3D points or as known as object points. The algorithm has been published by Kenneth Levenberg in 1994 and later rediscovered by Donald Marquardt in 1963 [9].

### *E. 3D Object Reprojection Estimation*

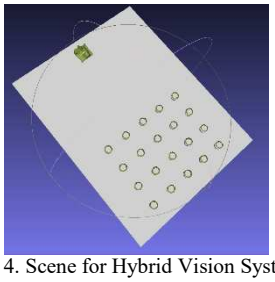

Fig 4. Scene for Hybrid Vision System

This section is focused on reprojecting the detected 3D object surface into the perspective camera frame and the two microscopes. To be able to do that we have come up with developing a new frame for the 3D object and the 3D scene which can be seen in Fig. 4 as the frame can be used for matching with the perspective camera frame and the two microscopes frame. We have a scene for our hybrid vision system scenario, from that we will obtain the 3D point coordinates of the circle grid pattern and the 3D point coordinates of the object as in Fig. 5.

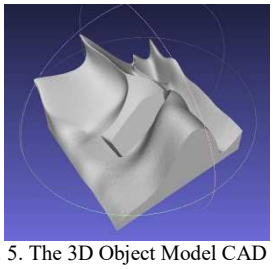

Fig. 5. The 3D Object Model CAD

From the circle gird pattern in the scene, we will pick three circle that can be form an 90˚ right angle triangle. As we can see in Fig. 6 is the process of acquiring the complete 20 points of the first circle which we will take only on the top layer of the points. The down layer of the point is not visible, and we want it to be a zero z-axis values.

Fig. 6. Finding the Coordinates of The Object

The three circles that we have picked is labeled as $P1_i$, $P2_i$, and $P3_i$ where i is the number of the points. From that we will find the mean of that as follows:

$$P1_{mean} = \frac{1}{m}\sum_{i=1}^{m} P1_i, P2_{mean} = \frac{1}{m}\sum_{i=1}^{m} P2_i, \mathrm{P3}_{mean} = \frac{1}{m}\sum_{i=1}^{m} \mathrm{P3}_i \quad (22)$$

After the mean value is acquired, we will then build the frame base of x, y, and z as follows:

$$\vec{x} = P2_{mean} - P1_{mean} \quad (23)$$

Where the $x$,

$$x = \frac{x}{norm(x)} \quad (24)$$

For y frame,

$$\vec{y} = P3_{mean} - P1_{mean} \quad (25)$$

Where the $y$,

$$y = \frac{y}{norm(y)} \quad (26)$$

Then final z is the cross product of x and y,

$$z = \|x \times y\| \quad (27)$$

Then we can have a homogenous transformation of T,

$$T = \begin{bmatrix} x & y & z & P1_{mean} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (28)$$

Where the transformation for the circle grid to world frame is:

$$T_{world^M} = (T)^{-1} \quad (29)$$

With that equation 29, we can proceed to check the estimation of the 3D projection the microscope camera with the first we calculate the camera object circle grid express in camera frame as follow,

$$T_{M^{CAM3D}} = T_{world^M} \times camObj3d \quad (30)$$

Then, we can compute the microscope projection frame,

$$T_{CAM^{MicOBJ_i}} = CAM_{TM_i} \times mic1_{\mathbf{T}cam} \times T_{M^{CAM3D}} \quad (31)$$

As for that, we can see the result of the estimation projection as Fig. 7.

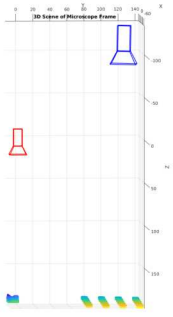

Fig. 7. 3D Estimation of Reprojection

## IV. Hybrid Stereo Camera System Results

In this section, we are going to discuss the detailed results achieve from the calibration of the perspective camera, microscope 1 and microscope 2, the reprojection result of the hybrid stereo camera system, the after optimization result of reprojection, and the validation of the calibration and the 3d estimation reprojection.

### *A. Calibration Result of Hybrid Stereo Camera System*

TABLE. I. Calibration Result of Perspective Camera

| Type of Calibration Pattern | Reprojection Error, px | Standard Deviation, px |
|---|---|---|
| Square | 2.7236 | 0.5004 |
| Circle | 1.3966 | 0.2798 |

TABLE II. Calibration Result of Microscope 1 and Microscope 2 Camera

| Type of Calibration Pattern | Mean Reprojection Error, px | |
|---|---|---|
| | Microscope 1 | Microscope 2 |
| Square | 0.2825 | 0.3382 |
| Circle | 0.3211 | 0.1791 |

### *B. Reprojection Result for Hybrid Stereo Camera System*

The Fig. 8 shows the example images that we use to extract the corner image points and by that, we will reproject it at microscope 1 and microscope 2 projection image by using the $\boldsymbol{x_{mic1}}$ and $\boldsymbol{x_{mic2}}$ for the reprojection calculations.

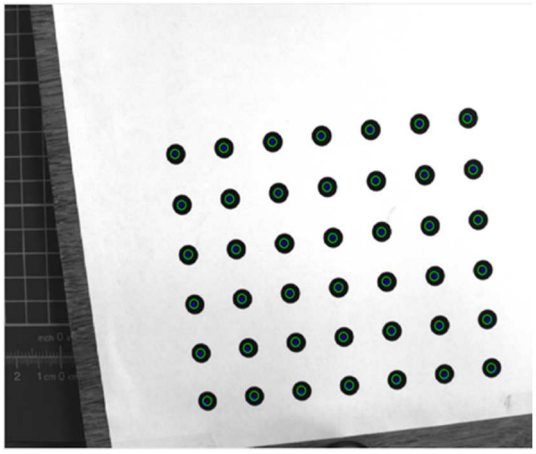

Fig. 8. Example of Image from Perspective Camera

#### *1)* $\boldsymbol{x_{mic1}}$

In The experiment of demonstrating the estimation reprojection from the feature extracted in the perspective camera with the symmetrical circle grid pattern and microscope 1, we have increased the number of images up to 19 images. In Fig. 9 we can see the mean error of the reprojection are **2.1600** pixel which says that our estimation is good enough and with the increment of number of images we are able to lower the mean reprojection error. The mean error is calculated by normalized the difference between the reprojection image point with the extracted image point in the frames.

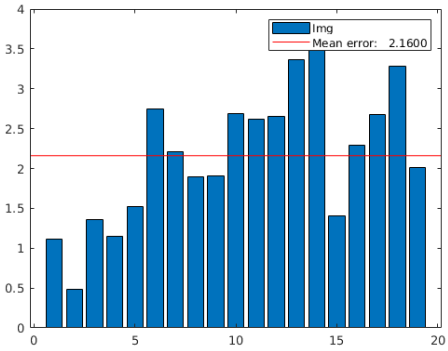

Fig. 9. $x_{mic1}$ Reprojection Error

The reprojection are being calculated with the $x_{mic1}$ that includes the intrinsic parameters of microscope 1, the homogeneous transformation and the object point that defines the world points in the microscope 1 frames. The visualization of the reprojection point from the microscope 1 with the x_mic1 can be seen in Fig. 10 as the marked blue 'X' is the reprojection point and the green circle is the visualization of the original extracted image points.

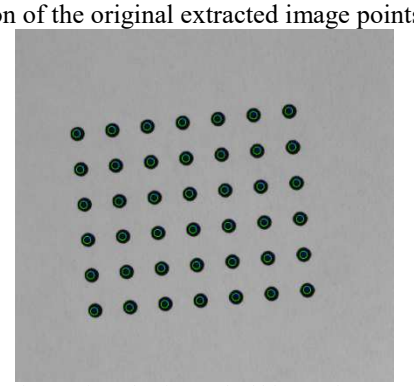
Fig. 10. Example reprojected point marked blue 'X' from extracted feature of perspective camera image, Fig 8

*2)* $x_{mic2}$

While for the experiment of demonstrating the estimation reprojection from the feature extracted in the perspective camera are also carried with microscope 2. In Fig. 11 we can see the mean error of the reprojection are **2.2479** pixel which says that our estimation is good enough. The mean error is calculated by normalized the difference between the reprojection image point with the extracted image point in the frames.

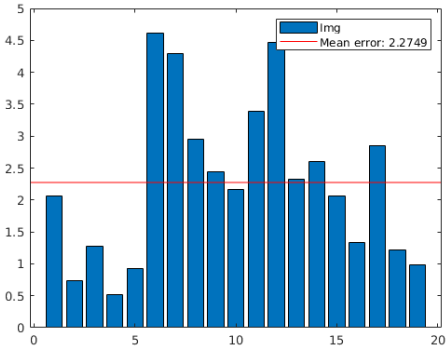

Fig. 11. $x_{mic2}$ Reprojection Error

The reprojection are being calculated with the $x_{mic2}$ that includes the intrinsic parameters of microscope 2, the homogeneous transformation and the object point that defines the world points in the microscope 2 frames. The visualization of the reprojection point from the microscope 2 with the $x_{mic2}$ can be seen in Fig. 12 as the marked blue 'X' is the reprojection point and the green circle is the visualization of the original extracted image points.

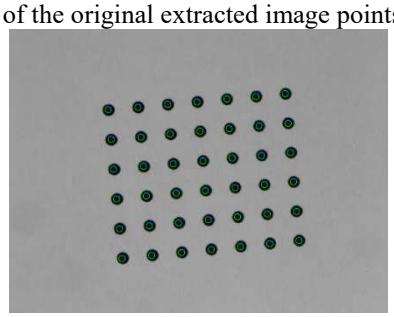
Fig. 12. Example reprojected point marked blue 'X' from extracted feature of perspective camera image, Fig 8

## C. *Enhancement Result of Reprojection Accuracy Through Optimization*

*1)* Result Comparison Before and After Optimization

| | Before | | After | |
|---|---|---|---|---|
| | Reprojection Error, px | Standard Deviation, px | Reprojection Error, px | Standard Deviation, px |
| Cam | 1.0682 | 0.5004 | 0.8059 | 0.1888 |
| $x_{mic1}$ | 2.1600 | 0.8505 | 0.2758 | 0.0576 |
| $x_{mic2}$ | 2.2749 | 1.2592 | 0.2344 | 0.0606 |

## D. *Validation of Calibration*

In this section, we are focusing on implementing the hybrid vision system in real-time. We have been using the Robot Operating System (ROS) as the platform for receiving the real-time footage from our three cameras and apply the homogeneous transformation equation that we have been optimizing together with the intrinsic parameter of the three cameras for the projection of estimating point received from the perspective cameras [10].

For the real-time experiment, we have been using the ROS for communicating with the three of the cameras and run the computation of the transformation for the reprojection to microscope 1 and microscope 2. We can see in Fig. 13, the live featured point extraction that we want to reproject in microscope 1 and microscope 2.

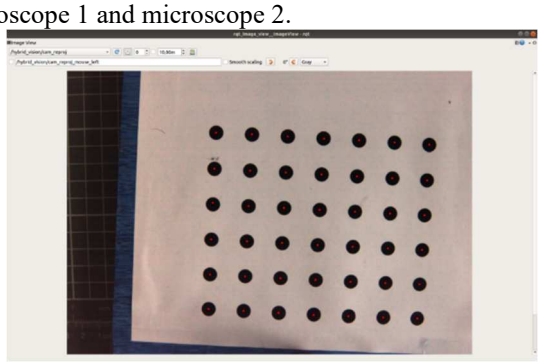
Fig. 13. Real Time Point Extraction from Perspective Camera

The reprojection of the extraction featured point from the perspective camera has been executed by subscribing to the perspective camera topic message. After the subscription is successful, the execution of finding the blob or finding the feature in the images is being executed. Fig. 13 is the new node that we publish after it has successfully extracted the featured points.

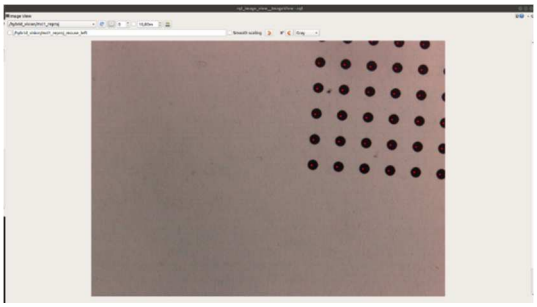

Fig. 14. Real Time Reprojection from Microscope 1

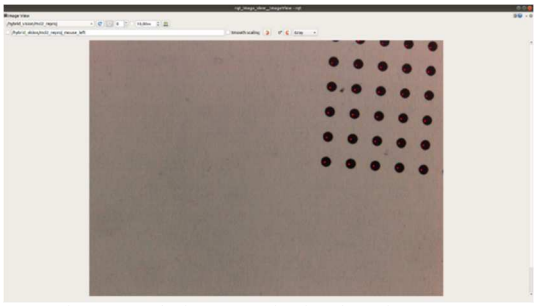

Fig. 15. Real Time Reprojection for Microscope 2

After the process of extraction is completed, the next process was to reproject the extracted feature points in microscope 1 and microscope 2. The process of reprojecting the points was by subscribing the microscope 1and microscope 2 image topics for getting the images and process it to project the extracted point from the perspective camera. The estimation of the reprojection point can be seen on Fig. 14 and Fig. 15, which is the marked red dot. We can see that the projected points have an average reprojection error of fewer than 1 pixel which is already good enough and ready to be implemented in the surgery.

### *E. 3D Estimation Reprojection*

The experiment has successfully given an output of the accurate 3D object estimation as you can see in Fig. 16 is the snapshot of it.

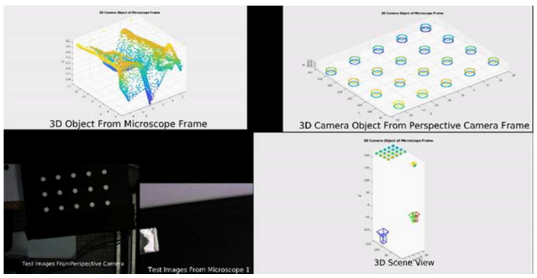


Fig. 16. Estimation 3D Reprojection of 3D Object

## V. Conclusion and Future Works

To conclude this project, the objective of a stereoscopic vision system capable of providing 3D flow dense and accurate enough for the 2D or 3D registration stage has been successfully achieving its objective.

Hence, this research study will help to contribute to fighting the loss of hearing and eyesight in the future. This study has shown the possibility of creating a vision system that allows having a clear view of the targeted surgical place and a clear view of the drug flow throughout the delivery.


## Acknowledgment

This work is supported by University of Burgundy and Universiti Teknologi PETRONAS (UTP). I would like to express my gratitude to my supervisors and colleagues from France and Malaysia for giving me a valuable resource and helping through the project.


## References


[1] J. Casale, P. F. Kandle, I. Murray and N. Murr., "Physiology, Cochlear Function," Treasure Island (FL): StatPearls Publishing, January 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK531483/. [Accessed 20 January 2021].

[2] E. E. L. Swan, M. J. Mescher, W. F. Sewell, S. L. Tao and J. T. Borenstein, "Inner Ear Drug Delivery for Auditory Applications," *Advanced drug delivery reviews,* vol. 60, no. 15, p. 1583–1599, 2008.

[3] M. Roizenblatt, T. L. Edwards and P. L. Gehlbach, "Robot-assisted vitreoretinal surgery: current perspectives," *Robotic Surgery: Research and Reviews,* vol. 5, pp. 1-11, 2018.

[4] A. Molaei, E. Abedloo, M. D. d. Smet, S. Saf, H. Ahmadieh, M. Khorshidifar, M. A. Khosravi and N. Daftarian, "Toward the Art of Robotic-assisted Vitreoretinal Surgery," *Journal of ophthalmic & vision research,* vol. 12, no. 2, pp. 212-218, 2017.

[5] Lébraly, P. a. Royer, E. a. Ait-Aider, O. a. Deymier, C. a. Dhome and Michel, "Fast calibration of embedded non-overlapping cameras," *2011 IEEE International Conference on Robotics and Automation,* pp. 221-227, 2011.

[6] P. Fonseca, "Application of Homogeneous Transformation Matrices to the simulation of VLP systems," 2017.

[7] B. Triggs, P. F. McLauchlan, R. I. Hartley and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," *International Workshop on Vision Algorithms,* p. 298–372, 2000.

[8] S. Recker, M. Hess-Flores and K. I. Joy, "Statistical Angular Error-Based Triangulation for Efficient and Accurate Multi-View Scene Reconstruction," in *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, Clearwater Beach, FL, USA, 2013.

[9] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the society for Industrial and Applied Mathematics,* vol. 11, no. 2, pp. 431-441, 1963.

[10] A. Dattalo, "ROS Introduction," Open Robotics, 2018. [Online]. Available: http://wiki.ros.org/ROS/Introduction.