Conference contribution:
Mohamed, A., Yang, C. & Cangelosi, A. *Stereo Vision based Object Tracking Control for a Movable Robot Head.* the 4th IFAC International Conference on Intelligent Control and Automation Sciences (ICONS), Reims, France.
Stereo Vision based Object Tracking Control for a Movable Robot Head

Abdulla Mohamed* Chenguang Yang**,1 Angelo Cangelosi***

* Centre for Robotics and Neural Systems, Plymouth University, UK  
** Zienkiewicz Centre for Computational Engineering, Swansea University, UK  
*** Centre for Robotics and Neural Systems, Plymouth University, UK

Abstract: In this paper, we have developed a visual tracking control system on a purposely built movable robot head mounted on a Baxter robot. The tracking control algorithm used a stereo vision system provided by a Bumblebee camera, facilitated by the MATLAB computer vision toolbox. The objective of this work is to enable a robot to move its head as in a similar manner as our humans beings to focus vision on a moving object. Identification of object and estimation of its coordinate are performed based on image processing techniques. A fuzzy logic technique is applied to control the moving head in order to bring human like motion into the robot. The method developed in this work can be extended to any other movable robot head platform. Extensive experimental studies have been performed to test the effectiveness and efficiency of the proposed method.

Keywords: Fuzzy logic, Stereo vision, Image processing

1. INTRODUCTION

Our human vision system is subject to a limited range of view, but the range can be extended by moving our head and eyes to cover larger surrounding areas. The head of human has 3 degree of freedom (DOF) which are yaw, pitch and roll. These movements allow us to see the surrounding area without moving our body (Song (1996)). Moreover, human have the ability to estimate the distance. Thanks to stereo based vision system, human eyes are able to estimate the depth of visualized objects. Aiming at bringing this ability to robots for capability enhancement, we develop a stereo vision based object tracking system on our purposely designed movable robot head. There is an obvious advantage of moving cameras/eyes that are integrated on a robot, i.e. where it allows the robot to discover the surrounding environment without adding additional movements to the robot body. The movement of the camera make the tracking of an object easier and keeps the object nearer to the center of the camera axis where its reduce the computation time be reducing the area of calculation S.Rougeaux et al. (1993). This function embraces the robot for being more flexible in different environments to adapt new tasks (Rodriguez and Yu (2011)) .

A conventional vision uses one camera to analyze the surrounding environment or the task. By using one camera we can get a flat image (2D information) about the environment such as identify different shapes or colors etc. The conventional vision cannot supply a 3D dimension (Adrian and Gary (2008)). On the other hand, stereo vision has the advantage over the conventional vision by providing a 3 dimensional space. Stereo Vision systems capable of creating a 3D cloud of point to restructure the environment in order to analyze the surrounding area with a measurement by using two cameras (Ramesh C. Jain and Schunck (1953)). The stereo vision can be used as an alternative to measuring sensors such as the laser sensor that is used in robots to measure distance. Where laser scanner are more expensive than using camera.

Computer vision has been wildly utilized in manipulating arms. The vision system is used as a feedback sensor that supplies the information about the object’s position. Both stereo vision and conventional vision use the vision servoing (s.Hutchinson et al. (2002)). This kind of process is called visual servoing (VS) control. There is a great advantage of using visual servoing control, where machines have their on eyes to see. The conventional teaching and training on robots can be eliminated by this method. The manipulator arms can plan their movements around the environment. Visual servoing control offer more accuracy with a less setup time (Low (2006)).

There are two different approaches to control the manipulator arms based on visual servoing, position based visual servo PBVS and image based visual servo IBVS (Corke (2013)). In PBVS the camera fixes on the top of the end-effect of the manipulator arm. Based on the calibration of the camera and the manipulator arm, the position of the target can be found according to the camera coordinate then the position transformed to the world coordinate. In IBVS the camera is fixed in a known coordinate according to the world coordinate in which the coordinate of the target transformed through the camera to the manipulator arm.

There are several research were done on both approaches, many of them were focusing on tracking objects. One of...
experiment was performed on stereo vision to detect the human motor reactivate (Šuligoj et al. (2014)). In this experiment, the camera was fixed in a specific position while tracking the moving object. Another study was conducted on two manipulator arms that complete each other where one of them is holding the camera as a guider in order for the other arm to perform the assigned task. A study also was done on a fixed platform with two rotated cameras to produce a zero disparity by tracking the object and keep it in the center of the cameras (S. Rougeaux et al. (1993)). This experiment’s result proves the difficulty of providing zero disparity, as a result of there system design.

However, humans have the ability to make the decision quickly with less information available; whereas the machine work on precise information hence it is difficult to make a decision as quick as human (Zadeh (1984)). Decision made by human are based on experience and skill gained during life (common-sense) thus, they do not need a precise fact during decision making while probability might be an important factor in the making decision (Guanrong and Trung (2000)). For example, humans do not need to measure the height of a person to say he is tall or short. The fact of a human being tall comes from our experience and classification of heights. Due to the elegance areas of fuzzy logic to find a solution in a complex system which is difficult to describe by the traditional logic (Bojadziev and Bojadziev (1996)).

Fuzzy logic has been widely used in many experiments. In (Farooq et al. (2013)), fuzzy logic has been used in autonomous car experiment to navigate in unknown locations. The experiment shows the stability and the fast response to any action. Experiments have been conducted on manipulator controllers, where the fuzzy logic is used to control the joints of the robot (Smith et al. (2015), McLauchlan et al. (1997), Lea et al. (1993) and Malik et al. (1997)).

In (Pal et al. (2012)), fuzzy logic (FL) has been used in two stages; at the binning FL was used to detect people during image processing, and in the second stage (FLPF) fuzzy logic particle filter used to track people. The tracking system was used depth information to decided which person first detected. A research was done on active vision in autonomous navigation using FLC in (Kundur and Raviv (2000)). In the experiment they used FL control to process the incoming data. FLC allowed the control to be in real-time feedback process.

In this paper, a stereo vision with a two degree of freedom platform was developed to keep the robot tracking an object and send the feedback of the object position. The controller of the camera was designed based on fuzzy logic system. The fuzzy logic was implemented in this experiment since the aim of this work is to simulate the movement of human’s head in robots, in the behavior of moving.

2. BACKGROUND AND PRELIMINARIES

In this paper, we employ a purposely designed movable robot head, which was designed to be fixed on the top of the head of the Baxter robot as illustrated in (Fig.1). This platform is equipped with two servo-motors (Dynamixel servo-motor AX-12) to move the camera in yaw and pitch. These type of motors give 50 RPM. The micro-controller used was Arduino UNO board. The communication between Baxter and the main computer was based on a UDP communication via wireless LAN.

![Fig. 1. Experiment setup consist of two DOFs and bumblebee camera (Stereo Head at A)](image)

2.1 Stereo Vision

The ability of image processing using one camera to identify an object and to find its position \((x,y)\) and orientation \(\theta\) in 2 dimensions is a very useful process. Despite the fact that this process is providing a 2D coordinates of an object, it is still important to measure the depth of that object. This may require adding extra sensors that can measure the depth of the object. Stereo vision system can be alternative of using extra sensors such as laser sensors. The stereo vision uses two cameras. It applies algorithm to find the position of the object in 3 dimensions. The stereo vision algorithm goes through four processes to find the coordinate of the object (Adrian and Gary (2008)).

Calibration process is the first step, where the calibration is a process used to find and correct the parameters of the camera. Because usually camera comes with internal defect due to the manufacturing such as lens distortion, lens misalignment and the alignment of two cameras in stereo camera. To understand the calibration process, first we need to understand the model of the camera (Fig. 2). We will start with a single camera algorithm then we will move to the stereo camera.

Fig.2 shows the simple camera model, where the camera coordinate \((0,0,0)\) C is the center of the camera and the image plane has its own coordinate \((U_0, V_0)\). \((U_p, V_p)\) is the point intersects with the Z axis of the coordinate of the camera and the image plane. \(f\) is the distance from the center of camera to the image plane which called the focal length.

\(p\) is a point in front of the camera plane with attached coordinate \((x,y,z)\). In the image plane, this point \(p\) is present in 2D coordinate \((U_p , V_p)\). Applying a right triangle trigonometry analysis, the following equations (1) and (2) are established.

\[
\frac{u_p}{X} = \frac{f}{Z} \quad (1)
\]

Copyright © 2016 IFAC
Fig. 2. A representation of the simple camera model used in calculation. Modified from by Samarth (2013).

\[ \frac{v_p}{Y} = \frac{f}{Z} \]  

Let us re-write (1) and (2) to represent the equations in the image coordinate. Note that, the image coordinate start from the top left of the image plane for Adrian and Gary (2008). The results are in (3) and (4) below.

\[ \frac{u_p - u_0}{X} = \frac{f}{Z} \]  

and

\[ \frac{v_p - v_0}{Y} = \frac{f}{Z} \]  

The above equations shows that only X and Y coordinates can be estimated as a scalar while Z cannot be calculated. (3) and (4) can be written in matrix form (5).

\[ p = \begin{bmatrix} u_p \\ v_p \\ w \end{bmatrix}, M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, P = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \]  

M is the camera intrinsic matrix which represent the internal parameters of the camera. While, p is the coordinate of the point in image plane and P is the measurement from the center of the camera to the object. Since single cameras provide only 2D images (X,Y), the Z cannot be calculated. Therefore, we assume the X and Y as scalar (Samarth (2013)).

By introducing (5), the calibration can be carried on, the pattern used is dimensional-ly defined like chessboard where the size of the its boxes are known. T is the translation matrix 3 × 1 and R is the rotational matrix 3 × 3 of the (6) is chessboard. M can be founded by designing algorithm that detects the internal corners in the chessboard and using (6) with different images taken from different angles by the camera.

\[ p = M \times [R \times T] \times P \]  

As shown above the calibration of the camera was using single camera, while in the stereo vision requires two cameras. The process of stereo camera calibration work exactly the same as if an extra camera is added (Fig.3). Using the same triangulation calculation we can find the coordinate of an object in the Z dimension (7). T is the distance between the two cameras. \( u_r \) and \( u_l \) are the distances along u axis of the right and the left images respectively. The differences between these two point are known as Disparity (d).

\[ \frac{u_l - u_r}{T} = \frac{f}{Z} \]  

and solve (7) for Z to yield (8)

\[ Z = f \times \frac{T}{u_l - u_r} \]  

The same algorithm of a single camera calibration will be apply in calibrating the stereo camera. In the calibration of stereo camera, the distance T and the mis-alignment between the two cameras are calculated by using (7) and (8).

Fig. 3. Stereo camera model

### 2.2 Fuzzy logic

Conventional digital logic works on decision making either 1 or 0 (true or false). This logic leads to mis-understanding the input value when compared to a fixed fact (Jun et al. (1994)). Fig.4 shows the traditional logic of deciding the speed of a car, where the speed defined as following: very slow (VS), slow (S), median (M), fast (F) and very fast (VF). As can be seen in the figure, at 20 km/h the very slow end and slow stars. This type of logic has a problem where the condition changes from one state to another state in a sudden change JERRY (2002). But the fact is that the car moving at 15 km/h is not the same as if the speed was 19 km/h and this is what the theory of fuzzy logic is all concerns about. This type of logic is similar to human logic of control. Most of the times, humans do not describe the fact with a precise number; for instance, the temperature will be in described in more precise way. The same algorithm of a single camera calibration will be apply in calibrating the stereo camera. In the calibration of stereo camera, the distance T and the mis-alignment between the two cameras are calculated by using (7) and (8).

Fig. 5 shows the fuzzy logic presentation of the speeds. The point at which different speeds crossed each other is called fuzzification (e.g., the VS interact with S under a certain degree). A set in fuzzy logic means a group of members...
have different values. For example the set of very slow speed in Fig.5 is the speed values from 0 to 30 km/h. We use fuzzy rules to make the decision based on the intersect of these sets and their intersection percentage. For example, if we take X from Fig.5 at 20 km/h, the set VS is contribute with set S by 0.5. The sets of fuzzy logic and their contribution can be represented by different shape and different rules.

Fig. 4. Traditional logic (crisp logic)

The fuzzy logic process is divided into three stages. The design of fuzzy logic starts with the fuzzyfication. At this stage, the memberships of different input set are identified. Fuzzifying the input means assigning a range to each input. The second stage, is to build the rules by using “If ... Then” statement and then aggregate the input using and rules to produce the output with percentage of the intersection between the two sets. The third stage, defuzzification is the process used to convert fuzzy output to crisp output by finding the area under the fuzzy sets Leonid (1997).

Where the centroid or mean maximum area is calculate to find the output as a crisp value. There area many algorithms used to do this process Saletic et al. (2002). Fig.6 shows the area of that need to be calculated based on the overlap percentage from the rules.

Fig. 5. Fuzzy logic presentation

The stereo vision is used to find the coordinates of the object. A calibration was done using MATLAB build in function. The algorithm used in the calibration was based on the procedure explained in above section and the work by Zhang (2000). A 2D chessboard was used in the calibration due to its simple structure the chessboard boxes sizes was 35 mm by 35 mm. Twenty five images were taken from each camera in order to use them in the calibration for the purpose of finding the intrinsic camera parameters M (the images from right and left were taken at the same time). The chessboard was presented in different orientations and translations while pictures were taken to enhance the accuracy. MATLAB produces a file contain the parameters of the camera.

When the parameters of the camera are identified, the stereo process takes the place. Because, a video processing was used so the algorithm shown below was done repeatedly for every frame. The algorithm started by images rectification using the calibration results. This process uses algorithm that scan both images to find a match feature. The scan process is done only on the same row level in both images. This leads to rectifying the image Fig.7a. The process is done by finding the same value of the pixel in both right and left images. Note that, the matching process was performed in the same raw level to reduce the processing time. The result of rectification process leads to measuring the disparity mapping (Fig.7b) which provide the distances between the right and left pixels that have the same matching properties. In(Fig.7b) the dark blue is the farthest white the red represent the closest disparity distance. A triangulation (reprojection) was done to create a cloud of 3D points(8). The process of reconstruction was described in section II. A reconstruction to the image was done to produce Fig.7c. Fig.7c presents the point of cloud from the camera side where the dimension shows are X and Y axes in mm.

3. CONTROL DESIGN

With a minor modification to the platform that was used in the previous work. The camera used for this experiment is Point Gray Research camera Bumblebee2. The software that was used in the control and image processing was MATLAB. The experiment is divided into three stages, image processing, object tracking and picking up the object.

3.1 Stereo Vision

The stereo vision is used to find the coordinates of the object. A calibration was done using MATLAB build in function. The algorithm used in the calibration was based on the procedure explained in above section and the work by Zhang (2000). A 2D chessboard was used in the calibration due to its simple structure the chessboard boxes sizes was 35 mm by 35 mm. Twenty five images were taken from each camera in order to use them in the calibration for the purpose of finding the intrinsic camera parameters M (the images from right and left were taken at the same time). The chessboard was presented in different orientations and translations while pictures were taken to enhance the accuracy. MATLAB produces a file contain the parameters of the camera.

When the parameters of the camera are identified, the stereo process takes the place. Because, a video processing was used so the algorithm shown below was done repeatedly for every frame. The algorithm started by images rectification using the calibration results. This process uses algorithm that scan both images to find a match feature. The scan process is done only on the same row level in both images. This leads to rectifying the image Fig.7a. The process is done by finding the same value of the pixel in both right and left images. Note that, the matching process was performed in the same raw level to reduce the processing time. The result of rectification process leads to measuring the disparity mapping (Fig.7b) which provide the distances between the right and left pixels that have the same matching properties. In(Fig.7b) the dark blue is the farthest while the red represent the closest disparity distance. A triangulation (reprojection) was done to create a cloud of 3D points(8). The process of reconstruction was described in section II. A reconstruction to the image was done to produce Fig.7c. Fig.7c presents the point of cloud from the camera side where the dimension shows are X and Y axes in mm.

3.2 Tracking the Object

The process of tracking an object is divided into two parts. First, identify the object via color threshold and then fuzzy logic system is used to keep the object within the camera center.

Tracking the object in image processing The image threshold pass through series of processes in order to identify an object. First of all, the object has to be placed in front of the camera. The user select the object by drawing a polygon around the wanted object. Then, the mean color value of the selected area is to be calculated. After that, the image will be transformed to the LAB color space to
of the camera in X and Y axis. In order to track the object, the camera coordinate was divided into 4 sections (Fig. 8). These sections represent the positive x-axis right (R), the negative x-axis left (L), the positive y-axis up (U) and the negative y-axis down (D). The above definitions represent two sets of fuzzy logic input, the x axis and y axis. These sets have members which defined based on the distance from the object to the center of the camera. The members of x and y axis sets defined as negative median (NM), negative near (NN), positive median (PM), positive near (PN) and Center (C). Tab. 1 shows the input fuzzy sets.

<table>
<thead>
<tr>
<th>X axis</th>
<th>NM</th>
<th>NN</th>
<th>C</th>
<th>PN</th>
<th>PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y axis</td>
<td>RM</td>
<td>RN</td>
<td>C</td>
<td>LN</td>
<td>LM</td>
</tr>
</tbody>
</table>

Table 1.

Fig. 8. Camera Coordinate divided into 4 sections: Right, Left, Up and Down

The output sets were selected as follows: move up (U), move down (D), move right (R) and move left (L). These sets have members to control the angular speed of the movement of yaw and pitch servomotors. The members are defined as fast speed (F), slow speed (S) and center speed (C) which almost equal to zero.

The memberships of the both sets were designed using a triangulation to reduce the computational time. Twenty five rules for the FLC were designed with “And” contribution between the sets. For example, rule one states that

**Fuzzy logic for tracking an object** The tracking control of the object is designed base on Fuzzy Logic controller FLC. FLC was used to keep the object close to the center

Fig. 9. Membership function of FLC where the input is the position of the object (q) and the output is the desired angular velocity of the moving camera (θ).
If the object is in the right median (RM) and is in the up near (UN) Then move the camera at fast speed to the right (FR) and move at slow speed to the up (SP). Fig.9 show an example of the rules design.

Fig. 10. Fuzzy Controller

Therefore, the controller system is designed to have an input of \( q \) which represents the 2D position of the object \( q = [q_1, q_2]^T \) where \( q_1 \) and \( q_2 \) are \( x \) and \( y \) coordinates of the object with respect to the camera center (Fig.10). \( q \) enters the fuzzy logic controller to give an output \( \dot{\theta} = [\dot{\theta}_1, \dot{\theta}_2]^T \), which is the desired angular velocity of the two servomotors. We can calculate the desired angular position of the motors as below:

\[
p^{k+1} = p^k + \dot{\theta}^k \times T
\]  

(9)

where \( p^{k+1} \) is the desire position, \( p^k \) is the current position of the servomotor (this reading is taken from the servomotor before using Fig.9), \( T \) is a sampling constant, and the desired speed \( \dot{\theta}^k \) of the servomotors are provided by Fig.9. The calculation above tell us the desired angular positions of both servomotors for yaw and pitch motions.

3.3 Object coordinate finding

By considering the relation of the object the stereo camera origin, a transformation between the object to the Baxter world coordinate was calculated below (10).

\[
P_{\text{obj}}^W = T_{m1}^W \times T_{m2}^W \times T_{\text{cam}}^m \times P_{\text{cam}}^m
\]  

(10)

\( P_{\text{obj}} \) is the vector from the world coordinate to the object, \( T_{m1}^W \) is the transformation between robot world coordinate and the motor one coordinates, \( T_{m2}^W \) is the transformation matrix between motor one and the motor two coordinates, \( T_{\text{cam}}^m \) is the transformation matrix between motor two and the camera coordinates and, \( P_{\text{cam}}^m \) is the vector between the camera and the object. This This vector is the coordinate of the object from the stereo camera.

The transformation between motor one and motor two are function of their angular positions \( \theta_{m1} \) and \( \theta_{m1} \), where these values depend on the object position in the world.

The operating system was divided into five stages (Fig.12). The system starts by scanning the surrounding area to find the target. Once the target is identified, the transformation from the object to the fixed world coordinate is calculated. The important result of the transformation was the translation coordinate \([X \ Y \ Z \ 1]^T\) of the object with respect to the world coordinate. The orientation of the object was ignored in this experiment. The fixed world coordinate is the robot center.

The camera will keep following the target object until the object stay stills. The system checks the position of the object and decided which arm is closer to the object. When the decision is taken the robot will move to pick up the object.

4. EXPERIMENT STUDIES

In order to evaluate the system stability and the ability of tracking the object smoothly with stereo vision and fuzzy logic controller, two sets experiments were performed. The first set of experiments were done by placing the object on the left side by 0.5 m from the center of Baxter robot and the camera set in its original position. When the system turn motors move to keep the object within the limited range. Two type of experiments were done: one without as shown in Fig.13 FLC and the other with FLC as shown in Fig.14.
Fig. 13 and Fig. 14 show the results of the first set experiments. The FLC shows a quick response by moving the camera to keep the object in the center of the field of view and in frame 15 the camera was placed in the required position as illustrated in Fig. 13. The result shows there is a fluctuating in pitch due to the changing of lighting and depth error. On the other hand, the system was unable to stabilize the camera to place the object in the center of the camera without using FLC as shown in Fig. 14, and this is obliviously effected by the image processing speed.

5. CONCLUSIONS

In this paper, we have developed a stereo vision based object tracking control system on a 2 degree of freedom movable robot head to simulate human head motion. The tracking control algorithm was designed using fuzzy logic theory, which lead to more reliably tracking system with less computational time. Stereo vision enables the function to measure the distance by using cameras only. These two features will reinforce the capability of the autonomous robot in an unknown environment. The experimental results shows how the moving head can be useful in discovering the surrounding area and tracking the moving object. Further work will be done on the platform and Baxter robot to increase the ability of fully automation and decision making.
ACKNOWLEDGMENT

The authors would like to acknowledge the financial support from Engineering and Physical Sciences Research Council (EPSRC) under grants EP/L026856/1 and EP/J004561/1 (BABEL). The authors thank Dr. Phil Culverhouse for his technical support, Snruti Amarjyoti and Peidong Liang for their experimental support, Junshen Chen for his technical assistance, and Jaafar Mohamed for his assistance.

REFERENCES


