2nd International Conference on Autonomous Robots and Agents
December 13-15, 2004 Palmerston North, New Zealand

Relative Visual Servoing: A Comparison with Hybrid Intelligent Models

Shahram Jafari Ray Jarvis
Intelligent Robotics Research Centre (IRRC)
Monash University
VIC 3800, Australia
{ Shahram.Jafari, Ray.Jarvis }@eng.monash.edu.au

Abstract

In this paper, we compare four methods to perform eye-to-hand visual servoing using primitive tactile information. They are relative visual servoing, direct fuzzy servoing, fuzzy correction, and interpolation. These methods just need a monocular eye-to-hand camera and primitive tactile sensors. The fuzzy methods are tuned [1] by an Adaptive Neuro-Fuzzy Inference System (ANFIS). The set-up is part of a robotic platform named COERSU. Experimental results from COERSU are provided to validate these methods. An evolved visual perception system based on genetic algorithms was used to have the best possible combination of object segmentation and classification parameters [2]. As the experimental results show, relative visual servoing represents a significant improvement over the other methods in terms of reliability.

Keywords: Visual servoing, eye-to-hand, adaptive neuro-fuzzy inference system (ANFIS), genetic algorithm.

1 Introduction

Currently, there are numerous research projects conducted to study fast and reliable cooperation between a camera and a robot's arm [3-6]. Most of them focus on the estimation of a feature Jacobian matrix [6, 7] mapping the motion of the robot arm to the changes in the image frame. These approaches concentrate on robots with cameras on their hands [8, 9]. Since human-like robots do not have eye in hand, we have avoided this approach. Another reason is that we wanted to observe the whole arm in the context of the obstacles in the scene (Fig. 1) so as to easily avoid them.

Image-based visual servoing using stereo vision (binocular camera) has been reported in the literature as in [10, 11]. The methods we propose require only monocular vision. However, in the presence of stereo vision, part of our method can be used to verify the noisy result of the disparity images [12] to get more accurate range information. In relative visual servoing, the robot arm could be easily manipulated to avoid object collision. This makes planning the trajectory easier in this approach. In our set-up, first the robot tooltip goes directly above the target and then moves downwards to the target to grasp and pick it up. This is a reasonably good trajectory planning strategy in the context of obstacle avoidance in table-top scenarios.

Our motivation for implementing different approaches was to compare the reliability and speed of the relative visual servoing with the other approaches in a typical table-top scenario. A genetic tuner was also applied for visual perception in all the methods considered here [2]. Although using this tuner makes the system more complicated, it is preferable to apply it to obtain a higher degree of precision. The genetic tuner makes the individual processes such as edge detection, segmentation, finding the tooltip in the image, and measuring the centroid of objects more robust.

2 Implemented methods

2.1 Method 1: Relative visual servoing

Biologically-inspired visual servoing is currently receiving more focus in the literature [13]. A human-inspired approach for locating the target using a

Figure 1. COERSU and a table-top scenario.

monocular camera can be achieved as follows: First, a few planned movements of the hand is made. Then the relative information between the target pose and the hand, picked up by the eye, and the idea about the hand movements are used to judge the position of the target. We could emulate this behavior by three major steps: initially, lining up the centroid of the target with the center of the robot tooltip; secondly, moving downwards maintaining this alignment; and finally grasping the object when the tooltip senses the target. This approach has the following disadvantages:

1. This approach is time-consuming because of the necessity to maintain the condition of alignment, which in turn, requires a continuous visual and mechanical feedback.
2. If we have other restrictions on the trajectory of the tooltip, for instance, to avoid obstacles the above approach is not suitable. It is possible that the wrist of the robot and/or the tooltip disturbs the objects surrounding the target (due to the structure or bulkiness of the wrist of the robot such as RTX UMI robot arm).
3. Considering the target, this approach can disturb a vertically positioned target by knocking it down.

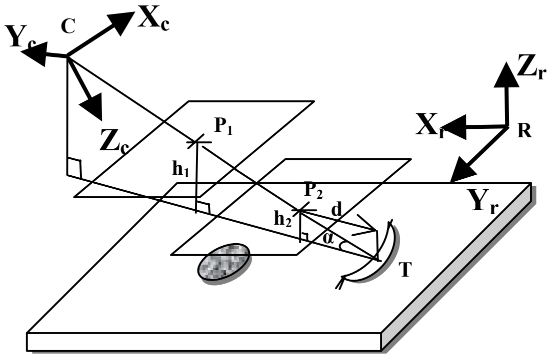


Figure 2. Relative Visual Servoing

R, Robot coordinate frame; C, camera coordinate frame; P_1, P_2, two alignment points; h_1, h_2, their heights from the table; d, horizontal motion adjustment; α, view point angle

Therefore, in our approach we modified this behavior by taking just two points of alignment and then using this information to find the horizontal motion required to place the tooltip directly above the target and then move downwards to pick it up as shown in Fig. 2. For further details refer to [14].

2.2 Method 2: Direct fuzzy servoing

Application of fuzzy logic in the context of visual servoing was reported for different purposes. Fuzzy integration of the cues for visual servoing was researched in [15]. Also, applying fuzzy logic as an estimator function in the context of visual servoing was simulated in [16].

In our attempt based on fuzzy inference, we found a direct mapping from the target position in the image frame to a proper position of the tooltip directly above the target. The salient feature of this approach in comparison with our other methods is its speed because it does not need to align the tooltip in the image. Another prominent feature is its independency from any kind of perspective matrix transformation or calibration. Therefore we directly reach to the desired position above the target then move downwards to pick it up. We considered a fuzzy case-based approach to solve the problem. Here, the cases were training data preferably with enough distribution over the domain. These cases were collected by placing the objects in different locations on the table. This idea has come from the fact that human patterns for picking up objects on a table have some common features. It was stated in [17] that humans were able to pick up objects with very sparse visual information. We believe that this is because of the previous training/practice they had from their childhood.

In order to obtain the proper position of the tooltip in 3D frame, two different fuzzy systems work in parallel. Each of them receives the position of the target in the image (row and column) and provides either x or y coordinate of the tooltip in 3D frame as its output. The tooltip is moved along a z-plane (height) fairly far from the table. In order to tune the fuzzy systems, two ANFIS networks are used.

2.3 Method 3: Fuzzy correction followed by alignment

Compared to the above method where we use very little online visual information and rely heavily on training, in the alignment followed by fuzzy correction technique we rely more on visual data and depend on training for just the last stage of adjustment. In this method, we first align the tooltip with the centroid of the target in the image and then according to two fuzzy systems find the horizontal motion Δx and Δy required to place the tooltip directly above the target, move downwards and grasp the target. The final adjustment for the tooltip is done based on the previously observed cases (i.e. fuzzy case-based). Similar to Method 2, two ANFIS structures are used to tailor the fuzzy systems.

2.4 Method 4: Linear interpolation method

In this approach, first we align the tooltip with the target in the image, and then according to an estimation function, we interpolate the horizontal motion required to place the tooltip directly above the target and then move downwards to pick it up. This estimation function was derived based on the training data. The training data was chosen to be the two extreme points visible in the image in both x and y directions as input and their corresponding adjustments as output. In order to simplify the estimation function, we considered a straight line fitting the training data. We believe that for small working areas the adjustment function can be considered semi-linear. Actually, the results obtained seem to validate our assumption.

3 Hardware description

In order to validate our methods, we set up COERSU to pick and place some objects on a dining table. The major hardware components of COERSU are (Fig. 3): 1. an anti-speckle USB color camera of resolution 640*480 acting as a visual sensor, 2. a gray-scale stereo camera pair producing left and right gray-scale interlaced images and a rough estimation of range is obtained by the disparity image [12]. 3. an RTX UMI robotic manipulator acting as an arm for COERSU, 4. a set of primitive tactile sensors around the tooltip of the robot and their relevant interface circuitry, 5. a pan-tilt unit resembling a neck for COERSU. 6. a frame-grabber which writes left and right frames given by the stereo camera pair directly to the disk.

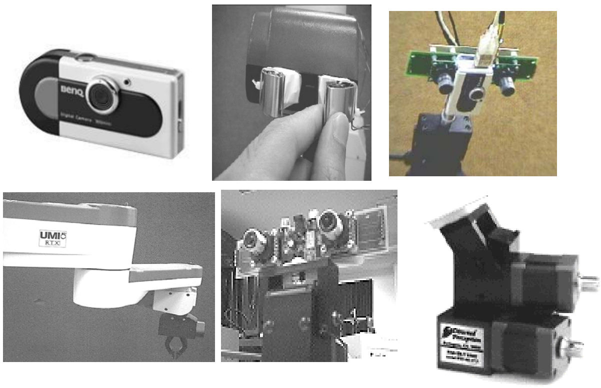


Figure 3. Basic hardware components (row-wise from left): Color camera, Primitive tactile sensors, Head of COERSU, Arm of COERSU, Grey scale stereo vision system, Pan-tilt unit (neck).

Therefore, COERSU can easily move its eye to adapt to movements of the arm and different locations in the scene (Fig. 1). The noisy result of gray scale stereo vision system [12] was outperformed by the accurate measurement of depth using tactile sensors.

3.1 Tactile Sensors

Some of the advantages in our implementation of tactile sensors can be summarized as: 1.Grasping soft objects such as fruit on a dining table without squeezing them. 2. Grasping free-size objects autonomously. 3. Assisting the vision system in obstacle avoidance by verifying the existence of fixed objects (estimating table depth by touching it. 4. Touching and verifying in the last stage before final grasping (Section 3.2).

3.2 The last verification stage

Provided that the target sufficiently stimulates the tactile sensors, (i.e. not hollow inside like a mug or liquid) we can consider the last stage of moving downwards to make sure that the target is located underneath. Apparently, the last verification is not necessary for the relative visual servoing where we find the final proper position fairly accurately.

4 Experimental results and analysis

In order to compare the efficiency of our methods, we carried out eleven experiments for each of our methods (Fig. 4) to pick up a cucumber in different positions. It was placed among other fruits on the table to validate our visual perception method. Note that the experiment numbers are sorted according to the distance of the target from the camera.

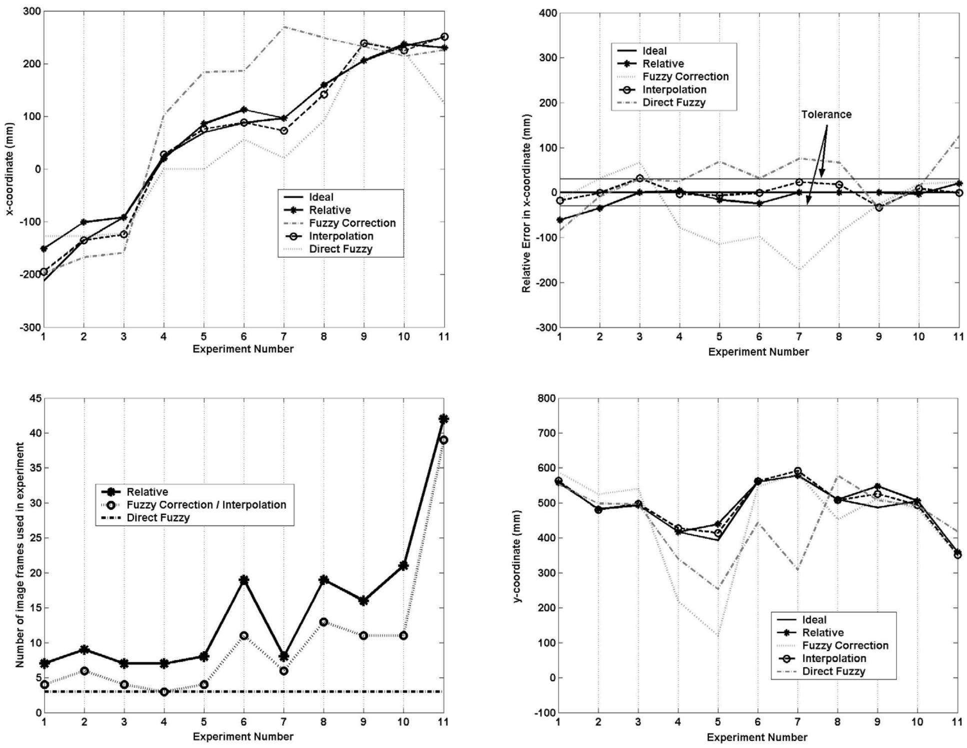
Accuracy: The success of each method employed in grasping the target is decided by how accurately they position the tooltip directly above the target. Therefore, the error of the tooltip position achieved in both x and y coordinate with respect to the target is plotted in Fig. 4b,e. The acceptable tolerance (denoted by two thin horizontal lines in Fig. 4b,e) helps to provide a measure of the success rate of each method. We also provide the Fig. 4a,d so as to give a clear perspective on the absolute displacement of the target in each of the experiments.

Time: The time spent for each experiment depends on both the image processing time and robot arm movements. The number of image frames processed reflects both of these operations. Therefore, the number of image frames processed during each experiment is shown in Fig. 4c to compare the speed of these methods. The time in seconds spent to perform the whole task from the beginning to the end of picking up the target is highly dependent on the processor specifications. Typically, the time spent ranged from 20 seconds to 30 seconds.

Referring to the Fig. 4c, one can see that in seven cases out of the total eleven experiments the number of frames processed increases with an increase in the distance of the target from the eye. This can be attributed to the fact that the error in the real world is magnified by the perspective effect. In other words, a small distance in the image corresponds to a large distance in the robot frame. The consequence is that it is more difficult to align the tooltip with a target which is far from the camera than one that is near.

After comparing the results (Fig. 4c) of our methods, we observed that, in all the cases, relative visual servoing performs slower than other methods due to the fact that two alignments were needed.

Another issue is that for a long target object such as a cucumber or a banana, the grasping technique is not that much dependent to the ideal positioning of the center of the tooltip on the exact center of the principal axis. The reason is that robot is going to pick up the object with a wrist perpendicular to the principal axis of the target. That is why the tolerance in y-coordinate was set more than that in x-coordinate.



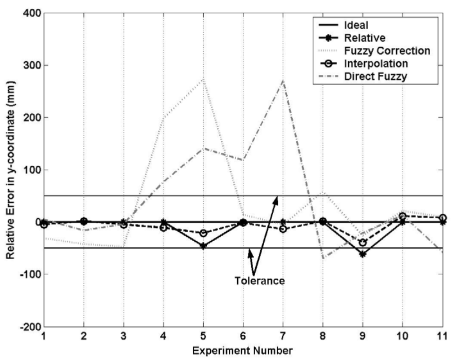


Figure 4. Evaluation of our four methods in eleven different scenes. points are connected to have a clear view (row-wise from top):

a) The desired and the actual x-coordinate of the tooltip in robot's 3D frame. b) Error between the desired x-coordinate and the actual x-coordinate of the tooltip in robot's 3D frame. c) Number of image frames processed. d) The desired and the actual y-coordinate of the tooltip in robot's 3D frame. e) Error between the desired y-coordinate and the actual y-coordinate of the tooltip in robot's 3D frame.

A study of the results of our eleven experiments shows that the fuzzy systems have the maximum error in both x and y coordinates of the robot. This is due to its tight dependency on the training data of ANFIS. However, the inherent common sense of fuzzy control forms the basis of knowledge that tends to provide a quick and reasonable result [18].

Considering the results of the interpolation method, one can see that it is fairly accurate. We believe that this is so, because for small variations of the target range the final adjustment function shows some linearity.

Since direct fuzzy servoing is based on look-then-move approach, it is the fast method. However, in terms of reliability, it does not perform well in comparison to the relative visual servoing.

5 Conclusion

In this paper we introduced four human-inspired methods for visual-servoing and hand-eye coordination. Although these methods relied on a monocular camera and tactile sensors, they can be applied to other robotic configurations.

The fuzzy expert systems tuned by ANFIS behave well when the target is located somewhere very close to its training input data. However, the behavior of the system changes when the target is located far from all the training data. The errors in the fuzzy systems are because of insufficient training samples, over-fitting and existence of conflicting samples due to inaccurate measurements. The consequence is that to get the best possible output from the fuzzy systems we have to provide the system with significantly large number of accurate non-conflicting training data, which does not seem to be feasible at least for our case.

Direct fuzzy servoing specifically works well when we need to place the tooltip fairly close to the target in one go of the robot arm. And it is independent from any kind of matrix transformation or calibration. In terms of accuracy, since it needs an accurate set of training data it does not perform as well as the relative visual servoing.

Although the calibrated linear interpolation technique shows an accurate result, it is not robust to the pan-tilt movements of the neck i.e. a small changes in the position of the camera means finding a new interpolation function based on the new training data.

Relative visual servoing outperforms the three other methods in terms of accuracy and reliability. However, the relative visual-servoing is slower than the other three methods because of the need for two alignments. One could trade-off accuracy for time by taking the camera position as one of the points in the horizontal correction. The time constraint is negligible in some applications and in some others, reliability and accuracy of grasping is more important. To its further credit, relative visual servoing does not need any off-line training, as do the other methods.

6 References

[1] S. Jafari, F. Shabaninia, and P. A. Nava, "Neural network algorithms for tuning of fuzzy certainty factor expert systems," presented at IEEE/WAC World Automation Congress, 2002.

[2] S. Jafari and R. A. Jarvis, "A Genetic Off-line Tuner for Robotic Humanoid Visual Perception," presented at IEEE Int. Congress on Evolutionary Computation, 2003.

[3] J. A. Piepmeier, G. V. McMurray, and H. Lipkin, "Uncalibrated Dynamic Visual Servoing," *Robotics and Automation, IEEE Transactions on*, vol. 20, pp. 143-147, 2004.

[4] F.-Y. Wang, "A Simple and Analytical Procedure for Calibrating Extrinsic Camera Parameters," *Robotics and Automation, IEEE Transactions on*, vol. 20, pp. 121-124, 2004.

[5] E. Malis, F. Chaumette, and S. Boudet, "2½D visual servoing," *Robotics and Automation, IEEE Transactions on*, vol. 15, pp. 238-250, 1999.

[6] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *Robotics and Automation, IEEE Transactions on*, vol. 12, pp. 651-670, 1996.

[7] K. Hosoda and M. Asada, "Versatile visual servoing without knowledge of true Jacobian," presented at IEEE/RSJ/GI Int. Conf. on Intelligent Robots and Systems, 1994.

[8] E. Malis, "Visual Servoing Invariant to Changes in Camera-Intrinsic Parameters," *Robotics and Automation, IEEE Transactions on*, vol. 20, pp. 72-81, 2004.

[9] K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura, "Manipulator control with image-based visual servo," presented at IEEE International Conference on Robotics and Automation, 1991.

[10] K. Hosoda, K. Sakamoto, and M. Asada, "Trajectory generation for obstacle avoidance of uncalibrated stereo visual servoing without 3D reconstruction," presented at IEEE/RSJ Inter. Conf. on Intelligent Robots and Systems, 1995.

[11] E. Cervera, F. Berry, and P. Martinet, "Is 3D useful in stereo visual control?," presented at IEEE Int. Conf. on Robotics and Automation, 2002.

[12] K. Konolige and D. Beymer, "SRI Small Vision System User's Manual, Software version 2.2f," 2002.

[13] D. Kragic and H. I. Christensen, "Biologically motivated visual servoing and grasping for real world tasks," presented at IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS 2003). 2003.

[14] S. Jafari, R. A. Jarvis, and T. Sivahumaran, "Relative Visual Servoing," presented at IEEE Conference on Robotics, Automation and Mechatronics (RAM), 2004.

[15] D. C. Kragic, H.I., "Cue integration for visual servoing," *Robotics and Automation, IEEE Transactions on*, vol. 17, pp. 18-27, 2001.

[16] D.-M. Chuang, S.-C. Wu, and M.-K. Hor, "Adaptive fuzzy visual servoing in robot control," presented at IEEE International Conference on Robotics and Automation (ICRA 1997), 1997.

[17] A. Hauck, M. Sorg, G. Farber, and T. Schenk, "What can be learned from human reach-to-grasp movements for the design of robotic hand-eye systems?," presented at IEEE Int. Conf. on Robotics and Automation, 1999.

[18] J. Durkin, *Expert Systems Design and Development*: Macmillan, 1994.