A Short Review of Neural Network Techniques in Visual Servoing of Robotic Manipulators

Dhanesh Ramachandram & Mandava Rajeswari
School of Computer Science
Universiti Sains Malaysia
11800 Minden,Penang,Malaysia

Abstract

Robotics is one of the most challenging applications of soft computing techniques. It is characterized by direct interaction with a real world, sensory feedback and a complex control system. This paper reviews the application of soft computing approaches, particularly neural networks, in the domain of visual servoing of robotic manipulators. Various robotic tasks within the scope of visual servoing are identified, and the issues involving the application of soft computing approaches in solving these problems are discussed. The paper provides some practical suggestions in applying neural networks for these tasks.

1 Introduction

The field of robotics in general involves several types of signal transformation. Firstly, there is the forward and inverse kinematic mappings for position control of the robot. Here, there is a need to translate Cartesian coordinate of the workspace in terms of joint or motor variables. This task can be easily carried out with the use of a built-in position controller which robot manufacturers provide. In such cases, it is sufficient for the robot controller to have access to an inverse kinematic mapping, i.e., that provides joint coordinates as a function of the position and orientation of the end effector. To realize a task that necessitates sensor feedback, an appropriate sensorimotor mapping relating sensor patterns to motor commands is needed. All these aforementioned mappings are often highly non-linear and it is difficult (though not impossible) to derive them analytically. Furthermore, because of environmental changes or robot wear and tear, the mappings may vary in time and one would have to adapt the control structure to these variations. Classical control methods usually rely on a reference model, whose discrepancy with a real system may lead to considerable errors.

It is highly desirable to have a method of learning these mappings automatically. Neural networks are good candidates for approximating non-linear transformation functions because they possess the following desirable features. Firstly, neural networks have the capability to learn from experience. They do not require explicit programming to acquire the approximate model. Secondly, neural networks may approximate arbitrary non-linear mappings subject to the availability of unlimited number of processing units. Thirdly, because of their massive parallel architecture, the data processing is fast. In the last decade research into neural networks has advanced to an extent that they are no longer viewed as mere black boxes [1], [2], [3].

Most successful applications in robotics utilize neural networks to implement some kind of signal transformation that may not be computed satisfactorily by other means; owing to the lack of knowledge about the underlying process, or because the conventional approach would turn out to be complicated and computationally expensive. In the field of robotics, neural networks have been applied in the following problems: to solve the inverse kinematic problem of robots, to map the non-linear relationships in robot dynamics as an inverse dynamics controller, in path or trajectory planning, to map sensory information for robot control and in task planning and intelligent control. The next section provides a brief overview of visual feedback for robot manipulators.

2 Visual Servoing

The development of modern industrial robots dates back to the 1950's and 60's when George Devol and Joe Engleberger created the "Unimate" robots. However, it was only in the 1970's when the research community increasingly began to accept the robot as a machine able to interact with its external environment with some degree of autonomy. This sparked the interest to use visual sensors to direct the motion of the robot. By strict definition, the term 'visual servoing' refers to the use of vision at its lowest level to provide closed loop position control of a robotic system. The approach uses a visual sensor to measure the error between the current location of the end effector and its goal location. The use of closed loop control increases the overall accuracy of the system by providing real-time feedback of error for the position control of the robots. Corke [4] provides an early review of visual servoing. More recent reviews of this research area may be found in [5], [6] and [7].

There are numerous methods of classification for visual servoing, i.e. in terms of control architecture, image features, or hardware configuration. A commonly adopted classification of visual servoing architectures is given by Weiss et.al [8]. According to their classification, a distinction between position-based and image-based control is made. In position-based visual servoing, image features are used in conjunction with a geometric model of the target and a known camera model, to estimate the pose of the target with respect to the camera. The feedback signal to control the motion of the robot is computed by reducing the positioning error in the estimated pose space. In image-based visual servoing, the location of features on the image plane is directly used for computing the feedback signal in the robot positioning system. In general, various approaches using the latter method focus on computing or estimating the image Jacobian - a function that relates the rate of change of a robot's pose to the rate of change of observed image features [5], [9],[10].The primary advantage of the image-based approach over position-based approaches is that it is less sensitive to camera calibration errors [11]. Image-based approaches also hold an advantage over position-based control in terms of computational load, making real-time control feasible.

Besides eye-hand coordination applications, visuomotor mappings underlie other robotics tasks such as visual robot positioning. The goal of visual positioning is to move a camera so that the image captured matches a given reference image. This has many applications, such as inspection, grasping, and docking.

3 Neural Networks for Visual Positioning

In visual control of robot manipulators, most applications of neural networks are aimed at approximating the non-linear image Jacobian. The neural network approach is attractive because it does not require any a priori knowledge of the controlled system, and is able to adapt to configuration changes in the robotic system during operation or by mere re-training. More importantly, neural networks are capable of the direct learning of the image Jacobian, as well as the possibility of avoiding the costly matching of image features in the current and reference images.

One of the earliest approaches to learning sensorimotor control was proposed by Miller [12] in which, a CMAC (Cerebellar Model Articulation Controller) network model learns the mapping between the current and desired images to the joint angle displacement of the manipulator. Since the last decade, there has been a steady stream of reported research adopting the learning approach to visual servoing of robots. The most predominant neural network architecture used for approximating sensory-motor transformation is the Multi Layer Perceptron model (MLP). This may be attributed to its proven function approximation capability [13] and fast recall due to inherent parallelism. Kubota and Hashimoto [14] used a MLP to learn the non-linear mapping between image deviations of four projected points of a viewed object with respect to a desired image, and the corresponding joint angles of a robot with an eye-in-hand configuration. Wei and Hirzinger [15] demonstrated a MLP neural network that performs visual servoing of a manipulator using the multi-sensor fusion of a laser range sensor and an eye-in-hand camera. While most work relied on off-line learning, on-line learning systems have also been implemented. Van der Smagt et al.[16] demonstrated visual servoing of a manipulator using a MLP capable of on-line learning using two computer workstations running in parallel. Other MLP based approaches include [17], [18], [19] and [20].

Besides the MLP neural network, other architectures have also been investigated. Wunsch et al.[21] asserted that network topology must be chosen in accordance with the representation of the 3D orientation of the object. Pose estimation is performed using a modified Kohonen self-organizing map, which learns by using synthetic views, generated by 3D CAD like models. Blackburn and Nguyen [22] conducted a comparative study on four different neural network architectures for vision directed reaching tasks. They concluded that the optimal algorithm to be used depended upon the availability of memory, the necessity of on-line adaptation and training speed. Other types of network models that have been used for visual servoing include Linear Local Mappings network [23] and hybrid neural networks [24].

Multi-sensor fusion and integration have been investigated in order to enable the visually-guided robot to perform complex tasks such as grasping, insertion, micro assembly and tele-operation more effectively than using a single sensor such as the camera. Among the advantages of using multiple sensors is that they provide an extended spatial, temporal or spectral coverage of the associated phenomenon. In addition, multiple sensors create overlap in observations, and thus redundancy. The key issue in sensor fusion is the accurate conversion of the physical measurements of a sensor to an internal model to which the actual fusion method would be implemented. For this, soft computing approaches based on fuzzy systems [25] and neural networks have proven to be successful [26], [15].

4 An Example Application

We briefly describe our implementation [27], [28], [29] of a neural network based visual positioning system. A 5DOF robot manipulator is used with the camera mounted on the end effector or placed on a tripod observing the target object. Our emphasis is on formulating efficient global image descriptors for a class of objects which exhibits little, if any, distinctive features on its surface to extract local information. We utilized a structured lighting approach to project a square array of grid lines from a modular laser unit which is fixed onto the end effector. Our system uses a standard MLP neural network and the conjugate gradient training algorithm for learning. Training is performed off-line in a supervised manner before a recursive positioning method is utilized during its task execution. The task to be performed by the robot is to move to a pre-defined reference pose from any arbitrarily chosen pose within its workspace by virtue of the visual information gathered by the sensor. The neural network learns the mapping between various global image descriptors of the scene and the relative Cartesian position of the end effector from its reference pose. In our experiments, we have achieved positioning accuracy up to 0.1mm for the Cartesian axes and up to 0.5 degrees for rotational axes. Figure 1 illustrates the recursive positioning errors during a typical positioning trial.

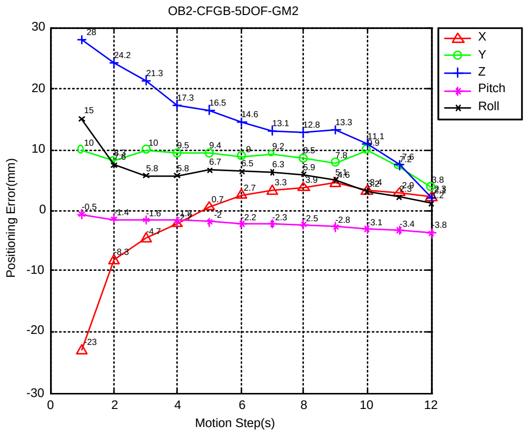

Figure 1: Recursive Positioning Result - 5DOF

5 Implementation Issues

Based on our experience, there are several factors that need to be considered before implementing a neural network to learn visuo-sensorimotor mapping for robot. These factors include:

- Using a suitable image representation

- Formulating suitable input-output pairs
- Using a reasonable number of training samples
- Finding out optimal neural network structure

Ideally, the observed image should be described by a fixed-length image feature vector for any views of the object. This is especially critical in the use of neural networks to realize the sensory-motor mapping as missing features may not be tolerated during training and execution of the system. In this regard, global image features such as geometric moments and Fourier descriptors have been successful [20], [17], [28] since these descriptors always provide a fixed set of image features for any image.

The image features often used in visual servoing may be correlated with each other, such that a motion along one axis of the robot may cause various image feature values to change in a complex manner. It is desirable to have a set of features that independently describes separate axes of the robot, but in practice, this is generally not realisable.

Many robots may be controlled using Cartesian movement commands or direct joint commands. Consequently, a conversion of the extracted image information into one of these forms of feedback is necessary. The complexity of this transformation has given rise to many approaches to vision-based control. For example, Miller [12] used an eye-in-hand configuration to track an object on a conveyor. In order to command the robot such that it achieves a predetermined reference pose with respect to the object, the neural network learned the mapping between the current and the reference image and the robot joint angle displacements used to move the end effector. Since the same relative hand to object position in different robot configurations would require different joint angle displacements to achieve the reference position, the robot's current joint angle configuration should be involved in the input space of the mapping. This increases the computational complexity of the mapping and violates the principle with which a correct mapping may be learned. Kubota and Hashimoto [14]made a simplification to Miller's method by considering the relative positioning with respect to a static object without having to involve the current hand configuration in the input space of the neural network. The disadvantage is that the learned relationship is pose-dependent i.e. it only applies for positioning with respect to the target object in a particular location. A method to solve this problem is to allow the neural network to learn the Cartesian motion of the end effector instead of the joint angles, provided the kinematic of the robot is known *a priori*. This makes the system suitable for both stationary and moving targets without increasing the dimensionality of the input space [15].

Another common concern in using neural networks for visual positioning would be the issue of having a reasonable training time. In order to design a system that is practical, a reasonable training time must be ensured. Real robots move slower than simulated ones, mainly to ensure fast damping and jitter-free, focussed images. Unacceptably long training time is not practical especially for dynamic environments. Therefore, a finite set of training samples is usually used in real-world applications. However, a finite set of training samples may be insufficient for the neural network to approximate the mapping to a desired accuracy, therefore, there exists a trade-off between sample size and training duration. To circumvent this issue, it is suggested that an initial supervised learning phase is implemented to allow the neural network to quickly learn an approximate mapping and during the execution, an on-line learning module enables the neural network to improve its generalization as and when it encounters new data. This may minimize the training time needed, while allowing the system to be adaptive. Nevertheless, this approach calls for an implementation of a hybrid neural network architecture that supports both off-line and on-line learning.

Although the MLP has been the *de-facto* neural architecture used in visual servoing, a noted limitation of the architecture is that, it is not known what size of network works best for a given task. This problem is unlikely to be resolved since each task demands different capabilities from the network. The optimal structure of the neural network may be determined using trial and error or *rules of thumb* [30]. Advanced methods for optimal network structure determination include the bayesian framework for model selection [31], weight decay [32] and weight elimination [33]. To overcome over-fitting, training the network using the *Early-Stopping Criterion* is recommended. The MLP should be replaced by a more dynamic neural network architecture which is capable of on-line learning and incremental learning. This ensures a dynamic model of the sensory-motor mapping is learned and the system is capable of adapting to changes in the environment. Hybrid neural networks which consist of both supervised and unsupervised learning such as the Growing Neural Gas or Fuzzy-ARTMAP neural networks are possible choices.

6 Conclusion

This paper has provided a brief review of the learning approach to visual servoing. The use of neural networks in visual servoing requires no elaborate calibration or analytical modeling of the sensor-robot system prior to using the system to perform its designated task. Some implementation issues and practical suggestions to overcome them have also been discussed. In conclusion, the neural network approach is relatively easier to implement and is more robust than its model-based counterparts. A good set of image features, coupled with a flexible, but effective neural architecture may hold the key to a successful implementation

References

[1] J. Benitez, J. Castro, and I. Requena, “Are artificial neural networks black boxes?,” *IEEE Transactions on Neural Networks*, vol. 8, no. 5, pp. 1156–1164, 1997.

[2] L. Bilbro and D. Van den Bout, “Learning theory and experiments with competitive networks,” in *Advances in Neural Information Processing Systems*, pp. 846–852, Morgan Kaufmann Publishers Inc, 1991.

[3] G. Towell and J. Shavlik, “The extraction of refined rules from knowledge based neural networks,” *Machine Learning*, vol. 31, pp. 71–101, 1993.

[4] P. Corke, “Visual control of robot manipulators - a review,” in *Visual Servoing* (K. Hashimoto, ed.), pp. 1–31, Singapore: World Scientific, 1993.

[5] S. Hutchinson, G. Hager, and P. Corke, “A tutorial on visual servo control,” *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 651–670, 1996.

[6] B. Espiau and R. Horaud, “Visual servoing with calibrated cameras - a review,” Tech. Rep. 26247, VIGOR, Esprit-IV reactive LTR project, 1998.

[7] D. Kragic and H. Christensen, “Survey on visual servoing for manipulation,” Tech. Rep. ISRN KTH/NA/P-02/01-SE, Center for Autonomous Systems, Numerical Analysis and Computer Science, Univ.of Stockholm, 2001.

[8] L. Weiss, A. Sanderson, and C. Neuman, “Dynamic sensor based control of robots with visual feedback,” *IEEE Journal of Robotics and Automation*, vol. RA-3, no. 5, pp. 404–417, 1987.

[9] R. Horaud, F. Dornaika, and B. Espiau, “Visually guided object grasping,” *IEEE Transactions of Robotics and Automation*, vol. 14, no. 4, pp. 525–532, 1998.

[10] E. Cervera, F. Berry, and P. Martinet, “Image-based stereo visual servoing : 2d vs. 3d features,” in *Proceedings of the 15th Triennial World Congress of the IFAC*, 2002.

[11] P. Corke, *Visual Control of Robots : High Performance Visual Servoing.* New York: John Wiley, 1996.

[12] T. Miller, “Sensor-based control of robotic manipulators using a general learning algorithm,” *IEEE Journal of Robotics and Automation*, vol. RA-3, no. 2, pp. 157–165, 1987.

[13] G. Cybenko, “Approximation by superposition of a sigmoidal function,” *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303–314, 1989.

[14] T. Kubota and H. Hashimoto, “Visual control of robotic manipulator based on neural networks,” in *Neural Networks for Robotic Control - Theory and Applications*, pp. 218–244, Ellis Horwood, 1996.

[15] G. Wei and G. Hirzinger, “Multisensory visual servoing by a neural network,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 29, no. 2, pp. 1–6, 1999.

[16] P. van der Smagt, F. Groën, and B. Kröse, "Robot hand-eye coordination using neural networks," Tech. Rep. TR CS-93-10, Dept. of Computer Science, University of Amsterdam, 1995.

[17] H. Yakali, B. Krose, and L. Dorst, "Vision-based 6-dof robot end effector positioning using neural networks," tech. rep., RWCP Novel Functions SNN Laboratory, Univ. of Amsterdam, 1996.

[18] P. Luebbers and A. Pandya, "Vision-based path folowing by using a neural network guidance system," *Journal of Robotic Systems*, vol. 11, no. 1, pp. 57–66, 1994.

[19] J. Cooperstock and E. Milios, "An efficiently trainable neural network based vision guided robot arm," in *Proceedings of the IEEE International Conference on Robotics and Automation*, (Atlanta), IEEE, May 1993.

[20] G. Wells, C. Venaille, and C. Torras, "Promising research : Vision-based robot positioning using neural networks," *Image and Vision Computing*, vol. 14, pp. 715–732, 1996.

[21] P. Wunsch, S. Winkler, and G. Hirzinger, "Real-time estimation of 3-d objects from camera using neural networks," in *Proceedings of the International Conference on Robotics and Automation*, pp. 3232–3237, 1997.

[22] H. Blackburn and H. Nguyen, "Learning in robotic vision directed reaching : A comparison of methods," in *Proceedings of the Image Understanding Workshop*, vol. 1, (Monterey, CA), pp. 1143–1150, 1991.

[23] A. Cimponeriu and J. Gresser, "Precision requirements for closed-loop kinematic robot control using linear local mappings," *Neural Networks*, vol. 11, no. 1, pp. 173–182, 1998.

[24] J. Buessler and J. Urban, "Visually guided movements with modular neural maps in robotics," *Neural Networks, Special Issue on Neural Control and Robotics : Biology and Technology*, vol. 11, pp. 1395–1415, 1998.

[25] M. Ribo and A. Pinz, "A fuzzy framework for multi-sensor fusion in mobile robot navigation," in *Proceedings of EuroFusion98 International Conference on Data Fusion*, (Great Malvern, UK), pp. 15–21, 1998.

[26] J. van Dam, B. Krose, and F. Groen, "Neural network application in sensor fusion for an autonomous mobile robot," in *Reasoning with Uncertainty in Robotics (RUR'95) Int. Workshop Proceedings* (L. Dorst, M. van Lambalgen, and F. Voorbraak, eds.), pp. 263– 277, Springer-Verlag, 1995.

[27] D. Ramachandram and M. Rajeswari, "Neural network based visual servoing using global image desciptions," in *1st. Regional Malaysia-France Workshop on Image Processing in Vision Systems and Multimedia Communications*, (Sarawak, Malaysia), UTM, 2003.

[28] D. Ramachandram and M. Rajeswari, "Visual positioning of a robot manipulator using the wavelet transform of image projections," in *Proceedings of the Int. Conf. AI In Engineering And Technology*, (Sabah, Malaysia), pp. 268–271, UMS, 2002.

[29] D. Ramachandram, M. Rajeswari, and S. Leo, "Structured-lighting approach to enhance pose characterisation using global descriptors for a model free robot positioning problem," in *Proceedings of the SPIE Intelligent Robots and Computer Vision XVIII*, vol. 3837, (Boston, USA), pp. 72–79, 1999.

[30] T. Masters, *Practical Neural Network Recipes in C++*. Academic Press, 1993.

[31] D. Mackay, "A practical bayesian framework for backprop networks," *Neural Computation*, vol. 4, no. 3, pp. 448–472, 1991.

[32] G. Hinton, "Connectionist learning procedures," *Artificial Intelligence*, vol. 40, pp. 185–234, 1989.

[33] A. Weigend, D. Rumelhart, and B. Huberman, "Backpropagation, weight elimination and time-series prediction," in *Proceedings of the 1990 Connectionist Models Summer School*, pp. 65–80, 1990.