1st International Conference on Sensing Technology
November 21-23, 2005 Palmerston North, New Zealand

Stereo Vision Controlled Humanoid Robot Tool-kit

C. H. Messom
IIMS, Massey University, Albany, New Zealand
C.H.Messom@massey.ac.nz

Abstract

This paper introduces a novel stereo vision based simulation system for a Humanoid Robot. This stereo vision system and robot controller is simulated dynamically and its performance evaluated for a standard humanoid robot benchmark problem. The vision system presented has been tested by simulating the dynamics of the robot system as well as the image processing subsystem. The real-time image processing and control algorithms allow the unstable dynamic model of the biped robot to be controlled

Keywords: Stereo vision, Biped Humanoid Robot, Dynamic Simulation

1 Introduction

A vision based humanoid robot system requires a high-speed vision system that does not introduce significant delays in the control loop. This paper presents a stereo vision system for biped control that performs in real-time. The humanoid robot used to test the system is a 12-degree of freedom biped robot [1]. The images from the cameras attached to the top of the robot is processed to identify positions of obstacles as well as any landmarks in the field of view. Obstacles near the robot can be accurately placed relative to the robot, and with the identification of land marks the robot can be accurately localized and a map of the obstacles developed in world coordinates.

Much early work in biped and humanoid robotics has focused on the basic dynamics and control of the biped robot system [2-5]. However more recently researchers have started to address the higher-level functionality such as biped robot vision for navigation and localization. To test and develop this functionality toolkits that support full simulation of the vision and control system have been developed [6]. This study builds upon the vision enabled robot simulation environment using the 12-degree of freedom m2 biped robot [1]. Typical views from the robot in an environment with obstacles is shown in figure 1. The key objective of the stereo vision system is to identify the objects in view, their position and depth in the image given changing viewing angles and lighting conditions. This requires object's characteristic color and size to be continuously updated based on current conditions.

Recently researchers have investigated biped vision strategies based on both simulation [7, 8] and real robot systems [7 – 10]. Braunl reported some of the problems associated with a "reality gap" when transferring results from a simulated system to a real robot system. Ensuring that the systems developed in simulation are not dependent on specifics of the simulation system ensure that this "reality gap" can be closed to the point that simulated solutions are useful for solving the real problem.

Figure 2 shows the image processing pipeline for the system presented in this paper. The core image processing algorithm used is the RLE [11, 12] based image segmentation and object tracking. This subsystem provides a real-time object tracking algorithm [13 - 15]. Once the key features or objects in the environment have been identified the stereo correspondence algorithm identifies matching features in the stereo images. The depth map of the objects in view is calculated based on the disparity between the features position in the two images.

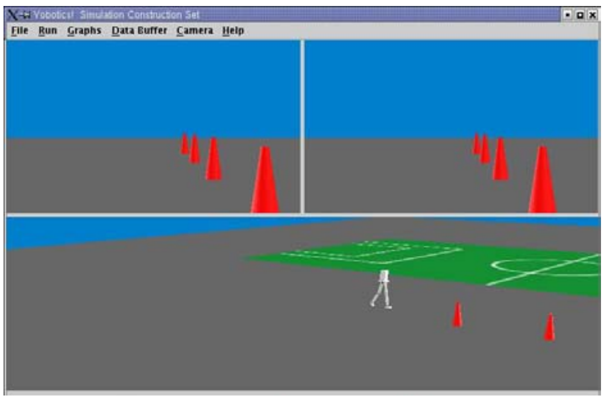

Figure 1: Stereo Vision Simulation System

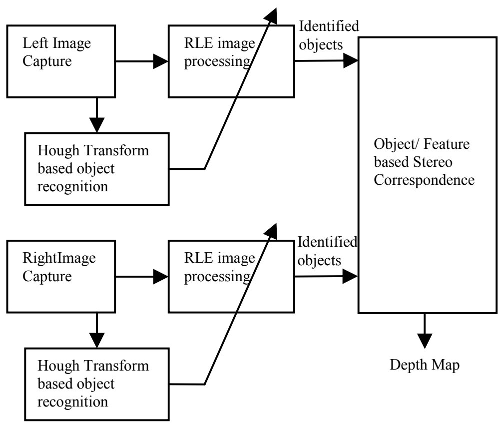

Figure 2: Image Processing pipeline

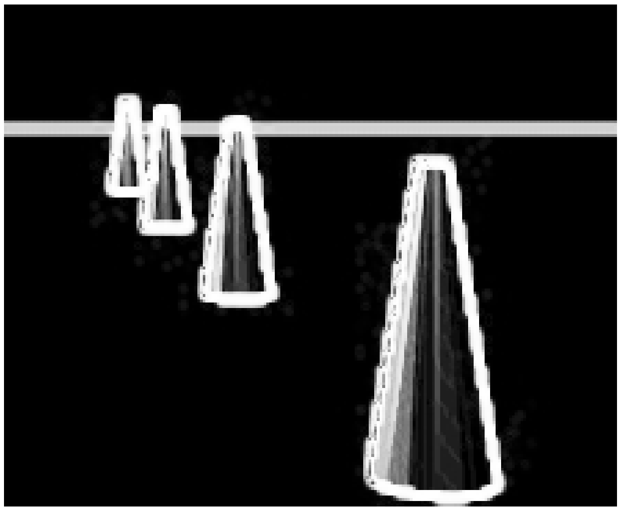

Figure 3: Edge detected image

2 Biped Systems

Biped humanoid robot structures have been investigated for many years [16-18], but it is only recently that the costs of the robot have been reduced to the point that it is possible to consider placing humanoid robots in everyday working environments. Before we can put humanoid robots in work related situations such as health care, aged care and miscellaneous service roles, machine intelligence must develop to the point that it is adequate to solve the problem. This research aims to develop robust humanoid robot controllers that can work reliably in dynamic environments along with people in a safe, fault tolerant manner.

Currently systems with a small number of links and degrees of freedom (say 12 as in a biped robot) can be simulated in real time on a single processor machine. Simulating a larger system (say 12 joints, their motors and sensors as well as the robots vision system) can not be completed in real time on a single processor machine, so a multiprocessor approach must be adopted.

This study is significant as real-time or faster than real-time simulation of robotic systems is required for intelligent controller design. Currently many techniques make use of kinematic models as a first approximation of a dynamic system so that the time complexity of the algorithms can be improved. While this approach is suitable for behaviours that are well understood and do not have complex dynamics, it is not suitable for investigating unknown or hard to model scenarios. For example when investigating vision based control of robots, a kinematic model would be suitable for a slow moving wheel based mobile robot that has as small and constant delay for image processing, but will not be suitable for fast moving, legged robot systems with highly dynamic delays as is the case with many modern image processing algorithms.

Early biped robot research focused on the control algorithms for static and dynamic gait control [19-22] however as these problems have essentially been solved researchers have turned their attention to higher level control algorithms. Several groups have taken the visual servo control [23-24] approach particularly motivated by the Hurosot benchmark problems, these benchmarks are classified as simulation and real robot benchmarks. Many of the active researchers have focused on the real robot system, particularly due to the difficulty of transferring a controller that works on the simulation model to one that will work on a real robot system and also because of the difficulty in transferring between different simulation models and environments. This research aims to overcome this limitation of simulation models by using realistic models of real biped robot systems that model the dynamics of the sensors and actuators used in the system as well as the robot morphology.

Many research groups have built biped robot structures beginning with Waseda's biped[16]. The last few years with the benchmark problems proposed by RoboCup Humanoid and FIRA Hurosot [24] many university based research groups have developed biped robot structures including Hanyang University [25], Singapore Polytechnic [4-6], University of Auckland (and Manitoba University) [26], National University of Singapore [27], University of Tokyo [28] plus many more.

The morphology of the biped robot used in this study is based on the M2 MIT robot from the MIT AI Lab [1] a 7-link, 12-degree of freedom biped robot (see figure 1) that uses series elastic actuators [29].

3 Stereo Vision Simulation System

The stereo vision simulation system is based on Yobotics Inc Simulation Construction set (SCS), a Java based 3 D dynamical simulation and visualisation system. The system provides morphological input of the robot and environment. This work extends the SCS system by adding robot vision components to the robot morphological tools available in the SCS. The robot vision components are configured to support simultaneous stereo imaging of the environment. The simulated stereo cameras are attached to the robot and moves with the robot. The cameras can be positioned and oriented at any point of the robot. Given a suitable controller the cameras can pan, tilt and rotate to any desired viewing direction.

3.1 Vision Processing

The images from the cameras attached to the top of the robot are processed to identify positions of obstacles as well as any land marks in the field of view. Close obstacles can be reasonably accurately placed relative to the robot, but land marks allow the robot to know where it and the obstacles are in world coordinates.

Edge detection is often used to identify objects and regions of interest in an image where there can be significant variation in size and colors of the objects of interest or the colors of the objects of interest are not known. In this paper edge detection is used to identify landmarks in the image, particularly the horizon and any unknown large obstacles. Once these areas have been detected the RLE algorithm is seeded with the required convex colour subspace required to identify the object.

A 5x5 RGB Sobel edge detection filter is applied to the raw RGB image producing the edge detected image (figure 3). This edge detection technique is computationally relatively expensive as compared to RLE, however in the situation where color identifiers are unknown it provides a suitable image that can be further processed to find information about the environment in which the robot is operating.

In the simulated domain presented one of the key features that can be identified from the edge-detected image are the horizon from which the body position of the robot can be inferred (this is useful if the joint angles are not explicitly available to the system). The second types of feature available in the edge-filtered image are land marks such as large obstacles which can be used to aid robot localization. In this study the colors of the obstacles where known and so were identified using RLE rather than the edge detection techniques.

Identifying the horizon means that a long almost horizontal (at least not near vertical) line must be identified. A similar approach will need to be adopted if there are walls or corridors in the image, that is, long lines in the image are identified before further processing.

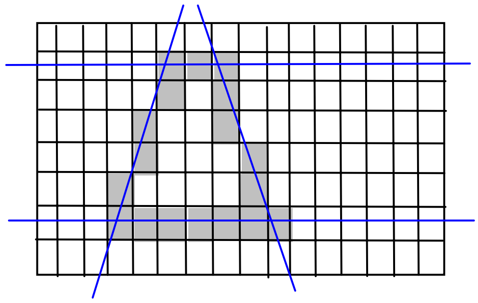

Figure 4: Grouped Edge detected pixels and candidate lines in object

Identifying straight lines in an image requires a first order Hough transform to be applied. Where the lines in the parameter space of the Hough transform intersect represents the equations of the lines in the image, this is also the position where there is a peak of data points in the parameter space. The edge-filtered image is also run length encoded so identifying the connected lines of contiguous objects (fig 4).

The Hough transform of the RLE edges is performed on each object separately [30]. This reduces the computational complexity of the algorithm as interactions between the objects are not added to the parameter space model, reducing the interference effect between the different lines in the image. With run length encoding only the start and end positions of the pixels in each horizontal row are recorded, so random pixels within the run length are selected for the Hough Transform to parameter space.

Each obstacle in the example image (figure 3) consists of two long straight lines and two short straight lines. The Hough transform is biased towards the long lines as they provide more candidate points in the parameter space. The two short lines are also identified since they provide two peaks in parameter space after applying the neighbor grouping algorithm. This is the case for the base line of the obstacles even though this line is barely straight. Figure 4 illustrates the selection of the selection of the candidate lines in the given object.

The lines and points formed by the intersection of the lines in each object define the boundaries of the objects of interest in the image. The pixels within this boundary are used to calculate the color space range values to be used by the RLE algorithm.

4 Stereo Correspondence

Stereo correspondence, finding matching areas, features or objects in a fast, reliable and robust manner is a challenging problem. Real images have a rich texture that allows area based correspondence [31]. The artificial images generated by the simulation currently have a very plain texture making area based approaches unsuitable.

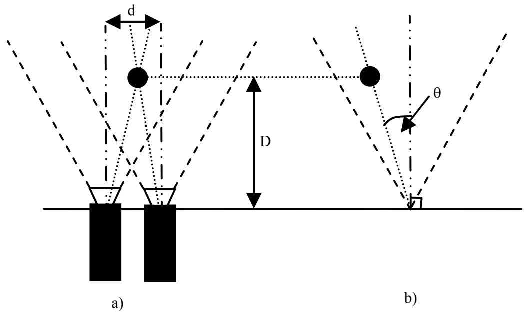

Figure 5: a) Stereo Cameras with object in field of view at depth D and interocular distance d. b) Angle of object (θ) relative to perpendicular to epipolar line.

A feature based approach is preferred in the current system, even though it yields a sparse disparity map as compared to the areas based approaches. Due to the lower dimension of the feature space the algorithm is faster than texture based approaches. Hardware approaches can be adopted to implement real-time texture based stereo correspondence algorithms [32].

In this study the correspondence algorithm uses features that are the identified objects found in each image. Based on the camera configuration (shown in figure 5) the epipolar lines in the image are horizontal so the correspondence algorithm makes use of the constraint that identical objects in each image should appear on the same horizontal line. Due to the errors in the calculated position of the objects (±1 pixel) this constraint is relaxed to include neighboring horizontal lines. Other constraints used include the size and color of the object. Slight variations between the two images are acceptable but a match only occurs when they are within specified tolerances.

5 Disparity and Depth Maps

Once the corresponding objects have been identified in the two images its depth can be calculated based on the disparity between the objects positions in the two images. The angle of the object in the image is calculated based on the model of the camera such as focal length, size of image plane and the position of the object in the image (1).

$$\theta^i = \arctan((\psi/2 - x)/\rho f) \quad (1)$$

where θ^i is the angular position of the object in the image perpendicular to the plane of the camera, ψ is the number of pixels in the horizontal direction of the image plane, ρ is the pixel density of the image plane, f is the focal length of the camera lens and x is the position of the object in the image.

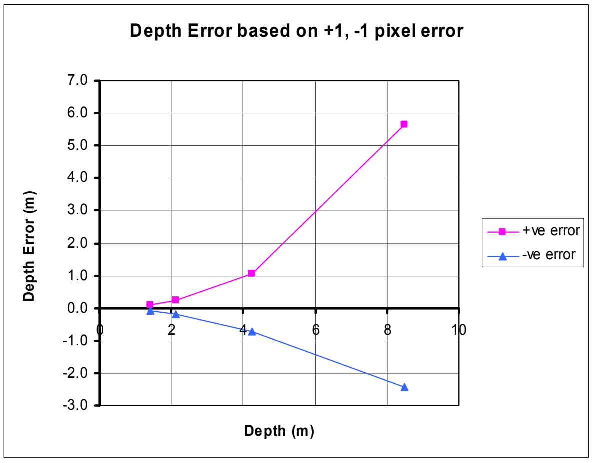

Figure 6: Variation in Depth measured due to ±1 pixel error in object's position

If the cameras are not aligned the angle from the perpendicular to the epipolar line include an angle offset as well as the actual angle in the image (2)

$$\theta = \theta^i + \alpha \quad (2)$$

where θ is the angular position of the object relative to the perpendicular to the epipolar line, θ^i is the angular position of the object in the image perpendicular to the plane of the camera and α is the angle offset of the camera relative to the perpendicular to the epipolar line.

The depth of the object based on the angles in the two images are given by (3) derived by using similar triangles.

$$D = d/(\tan(\theta_R{}^i + \alpha_R) - \tan(\theta_L{}^i + \alpha_L)) \quad (3)$$

Where $\theta_L{}^i$ and $\theta_R{}^i$ are the angular positions of the object in the left and right images perpendicular to the plane of the camera and α_L and α_R are the angle offsets of the left and right cameras relative to the perpendicular to the epipolar line.

The calculation of the depth is numerically stable for close objects but (as illustrated in figure 6 for a typical interoocular distance of 30cm, 35mm lens, horizontal field of view angle of 74^o and horizontal resolution of 320 pixels) for objects a long distance away (say >10m) even a small error in calculated object position in the two images (say ±1 pixel) has a potentially very large error in depth. This means this approach is only useful for accurately positioning objects that are close to the robot.

6 Conclusion

This paper presented a stereo vision based biped control system simulation toolkit that can operate in real-time when executing on two or more processors. This toolkit can form the basis for automatically developing vision based control systems using genetic programming and neural network learning techniques that can be rapidly developed before testing on real robot systems. Future work will extend the system so that multiple collaborating robots can be simulated in real time using multiple processors and machines.

The real time stereo image-processing algorithm is based on run length encoding, first order Hough transform and feature based stereo correspondence. This system can be implemented on real biped robot systems to detect obstacles quickly in the field of view.

This simulated system used in this study provides clean images and so does not reflect reality where there are often variations in the image due to sensor noise. Localization of the robot relative to the obstacles and mapping the environment with sensor noise requires particle filtering and optimal filtering approaches as discussed in [33]. Future research will model sensor noise and will require these additional techniques to provide reliable recognition of obstacle positions and localization of the robot.

7 Acknowledgements

The authors would like to acknowledge the use of Massey University's parallel computer the "Helix" for the computational experiments that supported the results presented in this paper.

8 References

[1] Pratt, J. and Pratt, G., "Exploiting Natural Dynamics in the Control of a 3D Bipedal Walking Simulation", Proceedings of the International Conf on Climbing and Walking Robots, 1999.

[2] Zhou, C. and Meng, Q., "Dynamic balance of a biped robot using fuzzy reinforcement learning agents", Fuzzy Sets and Systems, Vol.134, No.1, pp.169-187, 2003.

[3] Jagannathan, K., Pratt, G., Pratt, J. and Persaghian, A., "Pseudo-trajectory Control Scheme for a 3-D Model of a Biped Robot", Proc of ACRA, pp 223-229, 2001.

[4] Jagannathan, K., Pratt, G., Pratt, J. and Persaghian, A., "Pseudo-trajectory Control Scheme for a 3-D Model of a Biped Robot (Part 2. Body Trajectories)", Procs of CIRAS, pp 239-245, 2001.

[5] Baltes, J., McGrath, S. and Anderson, J., "Feedback Control of Walking for a Small Humanoid Robot", Proceedings of the FIRA World Congress, Vienna, Austria, 2003.

[6] Messom, C.H., "Vision Controlled Humanoid Toolkit", Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Artificial Intelligence vol 3213, pp 218 – 224, Springer Verlag, Berlin Heidelberg 2004.

[7] Boeing, A., Hanham, S. and Braunl, T., "Evolving Autonomous Biped Control from Simulation to Reality", Proc of 2nd International Conference on Autonomous Robots and Agents, pp 440 – 445, 2004.

[8] Chestnutt, J., Kuffner, J., Nishiwaki, K. and Kagami, S., "Planning Biped Navigation Strategies in Complex Environments", Proceedings of Int ernationalConf on Humanoid Robotics, 2003.

[9] Ogino, M., Katoh, Y., Aono, M., Asada, M. and Hosoda, K., "Vision-Based Reinforcement Learning for Humanoid Behaviour Generation with Rhythmic Walking Parameters", Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.1665--1671, 2003.

[10] Lorch, O., Albert, A., Denk, J., Gerecke, M., Cupec, R., Seara, J.F., Gerth, W. and Schmidt, G.,"Experiments in Vision-Guided Biped Walking", Proceedings of IEEE/ RSJ International Conf on Intelligent Robots and Systems, 2003.

[11] Sen Gupta, G., Bailey, D., Messom, C., "A New Colour Space for Efficient and Robust Segmentation", Proceedings of IVCNZ 2004, pp 315-320, 2004.

[12] Bruce, J., Balch, T. and Veloso, M., "Fast and Inexpensive Colour Image Segmentation for Interactive Robots", IROS 2000, San Francisco, 2000.

[13] Messom, C.H., Demidenko, S., Subramaniam, K. and Sen Gupta, G., "Size/Position Identification in Real-Time Image Processing using Run Length Encoding", IEEE Instrumentation and Measurement Technology Conference, pp 1055-1060, ISBN 0-7803-7218-2, 2002.

[14] Sen Gupta, G., Messom, C.H. and Demidenko, S., "Real-Time Identification and Predictive Control of Fast Mobile Robots using Global Vision Sensing", IEEE Transactions on Instrumentation and Measurement, vol 54, No 1, pp 200-214 2005.

[15] Messom, C.H., Sen Gupta, G. and Demidenko, S., "Hough Transform Run Length Encoding for Real-Time Image Processing", Proceedings of IEEE Instrumentation and Measurement Technology Conference 2005, 2198-2202.

[16] Hashimoto, S. and Takanobu, H., "Humanoid Robots in Waseda University - Hadaly-2 and WABIAN", IEEE-RAS International Conference on Humanoid Robots CD-ROM, 2000.

[17] Hired, K., Hirose, M., Hedkawa, Y. and Takenaka, T., "The development of Honda humanoid robot", In Proc of the Int Conf on Robotics and Automation, vol 2, pp 1321-1326, 1998.

[18] Shih, C.L., "The Dynamics and Control of a Biped Walking robot with seven degrees of freedom", Vol 118, Num 4, pp 683-690, 1996.

[19] Raibert, M.H., Brown, H.B. and Chepponis, M., "Experiments in Balance with a 3D One-Legged Hopping Machine", Int Journal of Robotics Research, Vol 3, No 2, pp 75-92, 1984.

[20] Okada, K., Kino, Y., Kanehiro, F., Kuniyoshi, Y., Inaba, M. and Inoue, H., "Rapid Development System for Humanoid Vision-based Behaviours with Real-Virtual Common Interface", Proc. Int. Conf. on Intelligent Robotics and Systems(IROS), 2002.
[21] Cupec, R., Denk, J. and Schmidt, G., "Practical Experience with Vision-Based Biped Walking", Proc 8th Int Symp on Experimental Robotics (ISER'02), 2002.
[22] Hirai, K., Hirose, M., Haikawa, Y. and Takenaka, T., "The development of honda humanoid robot", IEEE Conf on Robotics and Automation, 1998.
[23] Ishida, T., Kuroki, Y., Yamaguchi, J., Fujita, M. and Doi, T.T., "Motion Entertainment by a Small Humanoid Robot Based on OPEN_R", IEEE/RSJ Int Conf on Intelligent Robots and Systems, 2001.
[24] Pratt, J. and Kun, A., "Hurosot Simulation: Rules of the Game", FIRA web site http://www.fira.net
[25] Park., J.H., "Impedance Control for Biped Robot Locomotion", IEEE Transactions on Robotics and Automation, Vol 17, No 6, 2001.
[26] Baltes, J. and McGrath, S., "Tao-Pie-Pie", Proceedings of the RoboCup Symposium, 2003.
[27] Zhang, R. and Vadakkepat, P., "Motion Planning of Biped Robot Climbing Stairs", Proc FIRA Congress, 2003.
[28] Miura, H. and Shimoyama, I., "Dynamic Walk of a Biped", Int Journal of Robotics Research, Vol 3, No 2, pp 60-74, 1984.
[29] Pratt, G.A. and Williamson, M.W., "Series Elastic Actuators", Proc of the IEEE/RSJ Int Con on Intelligent Robots and Systems (IROS-95), vol 1, pp 399-406, 1995.
[30] Walsh, D.C.I and Raftery, A.E., "Accurate and Efficient Curve Detection in Images: The Importance Sampling Hough Transform.", Pattern Recognition, vol 35, pp 1421 – 1431, 2002.
[31] Veksler, O., "Fast Variable Window for Stereo Correspondence using Integral Images", 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003) (1) 2003: 556-564.
[32] Darabiha, A., Rose, J. and Maclean, J.W., "Video-rate stereo depth measurement on programmable hardware", Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. June 2003, Volume: 1, pp 203-210, vol.1.
[33] Yuen, D.C.K. and MacDonald, B.A., "Theoretical Considerations of Multiple Particle Filters for Simultaneous Localisation and Map-Building", Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Artificial Intelligence vol 3213, pp 203 – 209, Springer Verlag, Berlin Heidelberg 2004.