Proceedings of the 1997 IEEE
International Conference on Robotics and Automation
Albuquerque, New Mexico - April 1997

An Eye-Hand System for Automated Paper Recycling†

by

S. Faibish, H. Bacakoglu and A. A. Goldenberg
Robotics and Automation Laboratory, University of Toronto
5 King's College Rd., Toronto, Ontario M5S 3G8, CANADA
e-mail: faibish@me.utoronto.ca

ABSTRACT

This paper presents an robotic system used for detection, sorting and grading of paper objects as a part of an automated paper recycling system. The system uses stereo vision to estimate in real time the spatial position of moving objects on a conveyor to be picked by the robot arm. It was proved in the literature that stereo vision may be successfully used to estimate position of moving objects, with a priory known shape, and generate commands to a robot arm for picking up the object. A special computing architecture is introduced for performing the task in real time. A vector of geometrical and textural features is used for sorting the paper, according to its grade. The grading process uses color vision and additional contact ultrasonic sensors. A new data fusion paradigm, based on supervised learning, is used. The data fusion algorithm combines stereo vision and ultrasonic sensors to detect and grade paper objects.

1. Introduction

The basic principle of the proposed system is the use of vision and ultrasonic sensors for detecting paper objects from a municipality waste stream, and remove them from the scene. In order to enable real-time operation of the vision system, special image processing techniques are used. The basic principle of the image processing algorithm is to divide the image frames of the workspace, sampled by a pair of CCD cameras into image sub-frames. The image sub-frames of the empty scene are stored and used as reference images. The real-time algorithm computes the difference between the sampled images of the sub-frames and the reference images. The paper objects present in the scene are detected using an adaptive threshold process. The computation scheme has several advantages including: (i) only image sub-frames containing objects of interest are analyzed; (ii) smaller images are sampled and processed; (iii) there is no need for matching the two images of an object (from each camera) for the photogrammetry algorithm; and (iv) the efficiency and computation speed of the image processing algorithms are improved. Two issues related to 3D visual sensing from stereo pairs are addressed in this paper. First issue is concerned with the image processing techniques used for enabling real-time position estimates, using photogrammetry methods, as applied to moving objects of unknown structure. Second, the fusion of color vision and ultrasonic data is used for detecting, with high confidence level, special properties of small paper objects. The method is applied to an automated paper recycling system. A special vacuum gripper was designed for gripping paper objects detected by the vision system. Experimental results are presented which prove the efficiency of the proposed techniques.

The main objective of the stereo vision system is to calculate spatial location of objects, in the robot task space coordinates, and enable autonomous vision guided manipulation of objects by the robot arm. The stereo vision system acquires images from two cameras, calculates the position of a pre-defined point in the World Coordinate System (WCS), and use this information to close the loop of a robotic control scheme.

Determination of the object position and location in a factory environment, like material handling or mechanical assembly, is a major task in robot vision. The 3-D position of an object can be found by monocular or stereo vision. Monocular vision has been reported to attract much attention, but it is assumed that object dimensions in the scene are known a priori [6]. In our particular application we have used stereo vision since the object dimensions are not known. Experimental results show that stereo vision compensates for the uncertainties that may occur from one camera. Possible sources of such uncertainties are poor lightening, noise in the image and camera calibration errors.

The vision system acquires images through two fixed CCD cameras with a resolution of 480(V) by 640(H) pixels. The imaging is modeled using a pin hole camera model. A two-step camera calibration method is used to obtain our imaging model [1]. After the calibration is completed for each camera, the location of the point, in world coordinates, can be obtained by extracting its image projections with sub-pixel accuracy (see [7]). The coordinates of the point are transformed to the robot coordinate system by a rotational and a translational transformation, accurately measured. A similar technique can be used for robot positioning and trajectory calibration in automated visual inspection of 3-D parts.

† This work was supported in part by IRIS (Institute of Robotics and Intelligent Systems)

0-7803-3612-7-4/97 $5.00 © 1997 IEEE

2. Brief Description of the System

In order to proof the feasibility of the proposed concept a demonstration system was build. The principle block diagram of the demonstration system is presented in Figure 1.

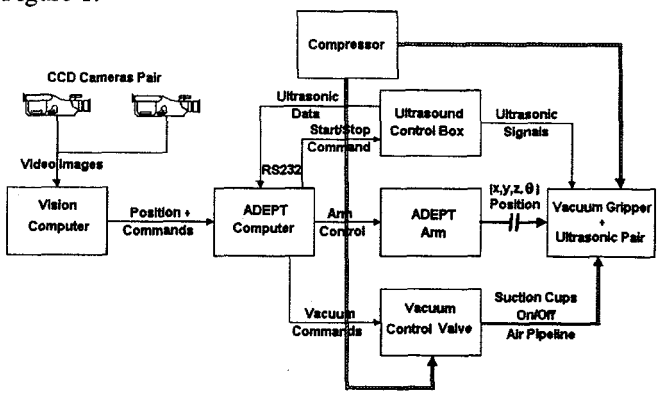

Figure 1: Schematic block diagram of the system

The Vision System: Consists in a pair of color CCD cameras with mono-focal lens and focal length of 4.8 mm, and two uniform illumination sources placed over the scene. The video outputs of both cameras are digitized into 24 bit/pixel RGB data by a Pentium 133 MHz vision computer. The vision software was implemented in Matlab interfaced with the image grabbing board driver, using a Dynamic Data Exchange mechanism. Control commands are sent to a robot arm using a RS232 serial port.

The Robot Arm: The robot arm used, had to be powerful enough to raise the vacuum gripper, the ultrasonic probes and a maximum load of 3 kg. The most adequate robot configuration for manipulation of large loads is a SCARA type robot. An ADEPT SCARA manipulator with four degrees of freedom, able to carry loads up to 5.5 kg, was used. The ADEPT computer was connected by 2 serial ports, to the vision computer and to an ultrasonic sensor.

The Ultrasonic Sensor: Consists in a pair of emitter/receiver probes, containing low frequency resonant crystals, and an electronic control box. The ultrasonic probes are mounted on the vacuum gripper and the control box is mounted on the arm. The sensor measures the time delay of the sound traveling through a tested object.

The Vacuum Gripper: Was designed and built in the laboratory and consists in three components: a vacuum gripper, 2 ultrasonic probes and an aluminum structure (see Figure 2). The vacuum gripper consists in four independent vacuum cells each containing a group of three suction cups. The vacuum generator and the vacuum valve are mounted on the arm. The gripper is able to hold a 50 paper sheets pile and 3 kg weight load.

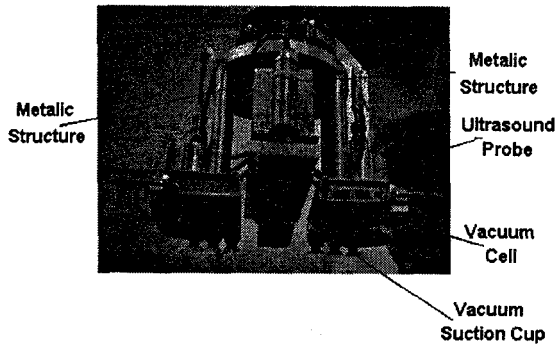

Figure 2: The Vacuum Gripper

3. Overview of Algorithmic Modules

In this section we present the main algorithmic modules of the system. The algorithmic modules are: Vision module, Robot control module and Data fusion and decision module.

3. 1. Vision Module

This section describes the up-to-date progress in the vision algorithms used for paper recycling. The reader is assumed to have a basic knowledge of mid-level vision and pattern recognition.

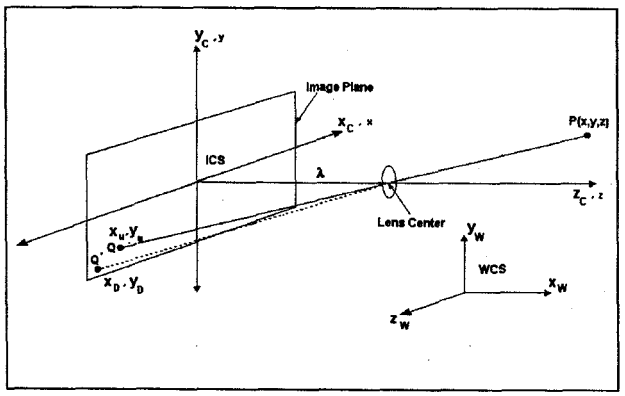

Figure 3: Camera model

Image Acquisition: The vision system acquires images from two color CCD cameras. The system is capable of acquiring 24-bits color RGB images. The imaging system is modeled using a pin hole camera model as in Figure 3. In this model a point (X,Y,Z) in World Coordinates System (WCS) is projected to the image point (x_D, y_D) in the Image Coordinate System (ICS). The lens is characterized by its effective focal length λ.

Calibration: The issue of camera calibration in the context of machine vision is to determine: the *intrinsic parameters* which give information about the optical and geometrical camera characteristics such as focal length, scale factors, lens distortion and the intersection point of

the camera axis with the image plane and the *extrinsic parameters* which give information about the position and orientation of the camera frame relative to a world coordinate system such as rotation and translation.

To obtain these parameters, calibration points with known world and image coordinates are used. Each camera is calibrated independently. The calibration target is a 50 cm Plexiglas cube which has 45 white circles on black background as shown in Figure 4. The black background is specifically chosen to decrease the blurring effect of impulse response of the camera. Matching is carried out between the centers of these circles and their projections on the image plane.

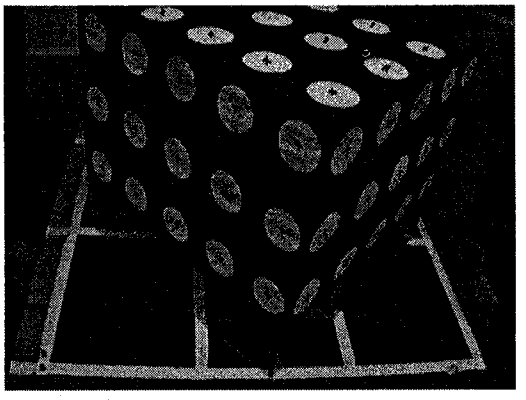

Figure 4: The Calibration Target

Image processing: This is the most important step of processing large data sets to make the results more suitable for classification than the original data. The process does not increase the information content, but it does increase the dynamic range of the data. There are two approaches for image processing: *spatial domain processing* and *frequency domain processing*. We have selected spatial domain method since it is faster and the pixels are manipulated directly.

Object Segmentation: A spatially dependent object segmentation method is used within each sub-frame. The difference image is taken and then filtered with a median filter to remove white noise. Linear gamma correction is used to enhance the image. Since we deal with objects with unknown gray level properties, an adaptive thresholding technique is used to segment the object from background. Because of its speed and robustness, we chose Otsu's Method [5] to convert the gray level image to binary image. This binary image is then used to extract the pixels belonging to the object, Figure 5c. In cases of dark colored objects this method failed and an adaptive threshold parameter depending on object color was used.

Photogrametry: A two camera stereo vision system is used for obtaining the location of objects with respect to the robot. To obtain the location of an image point (x1,y1) in the world coordinates, the same point has also

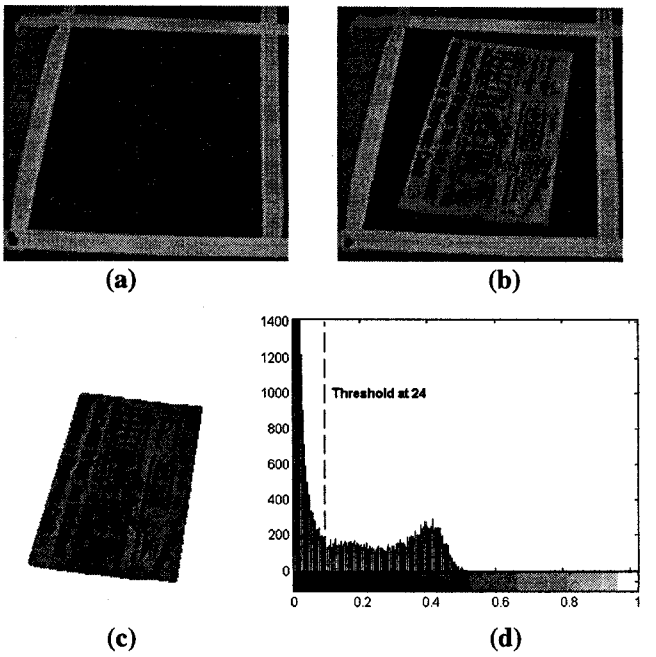

Figure 5: Empty frame (a), frame with object(b), segmented object (c), histogram of the image (d)

to be extracted from the other image with the coordinates (x2,y2). In order to avoid model matching, the centroid of the object in each image is used. The centroid algorithm has also facilitated to work with sub-pixel accuracy. Finally, the four degrees of information obtained from the two cameras is used to calculate the world coordinates (X,Y,Z) via pseudo-inverse least squares solution. At the end of this stage the orientation of the object is computed using Hough transform techniques [4].

Features Extraction: The feature extractor uses two different types of features to describe the structural properties: *shape descriptors* and *textural features*. The shape descriptors used were: circularity, area, perimeter and area-to-perimeter ratio. There are 7 textural features that were derived: *uniformity of the textural energy, entropy of the image, contrast of the gray level variation, inverse difference moment of the image, correlation of the co-occurrence matrix, homogeneity of the texture* and *cluster tendency of the texture*. The computation formulas are presented in Table 1.

Classification: In general, pattern classification techniques are grouped into two: *parametric* and *non-parametric techniques*. In this research a non-parametric pattern classification method was used without assuming that the forms of the underlying densities are known. Several non-parametric classifiers have been tested, including: *Fisher's linear classifier* [3], *Nearest-Neighbor classifier, Condensed Nearest-Neighbor classifier* and *Perceptron classifier*. Among these

classifiers Fisher's method gave the most accurate results, Figure 6.

Uniformity of energy	$\sum_{i,j} P_{ij}^2$
Entropy	$-\sum_{i,j} P_{ij} \log P_{ij}$
Contrast	$\sum_{i,j} \lvert i-j \rvert^k P_{ij}^l$
Inverse difference moment	$\sum_{i,j} \frac{P_{ij}^l}{\lvert i-j \rvert^k}$ i≠j
Correlation	$\sum_{i,j} \frac{(i-\mu)(j-\mu)P_{ij}}{\sigma^2}$, $\mu = \sum_{i,j} iP_{ij}$
Homogeneity	$\sum_{i,j} \frac{P_{ij}}{1+\lvert i-j \rvert}$
Cluster tendency	$\sum_{i,j} (i+j-2\mu)^k P_{ij}$

Table 1: Textural Features

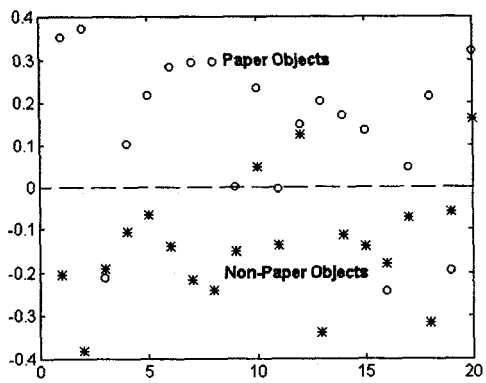

Figure 6: Fisher Linear Classifier, `o'=Paper, `*'= Non-Paper, - - - = Paper border

3. 2. Robot Control Module

In order to implement the data fusion and manipulation of the objects an ADEPT robot arm was used. The arm was controlled by the ADEPT computer. The control tasks includes the next functions: Trajectory planning, Motion control, Vertical force/position control, Ultrasonic sensor control and Vacuum gripper (x,y) position control.

Trajectory planning: Consists in the selection of the start and end position on the trajectory in task coordinates, transform the position into joint coordinates and generate the point to point trajectory in joint coordinates. There are no restrictions on the trajectory, in the task space, between the point to point trajectory in joint coordinates. There are no restrictions on the trajectory, in the task space, between the start and end joints position. The resulting trajectory consists in four legs: gripper up; move gripper over frame center; move gripper to object's location and orientation as received from the vision computer; lower the gripper down until the object or table is reached.

Motion control: Refers to the generation of the closed loop control commands to all the joints. A PID controller was used for each joint high level. The controller uses trapezoidal velocity profiles. The closed loop position controller was implemented using a joint based control scheme with Cartesian path input. The position control loop consists in a PID controller receiving input from the inverse kinematics block. This control scheme is not very robust and precise but was proven sufficient to meet the position accuracy requirements of the system.

Vertical force/position control: The purpose of this function is to keep the contact force between the vacuum gripper and the object within a given range and in the same time to estimate the height of the contact object. The contact force value is the desired torque of joint 3 computed from feedback information in the closed loop. The value of the contact force was experimentally selected to ensure optimal ultrasonic measurements. The additional purpose of this function is to estimate the height of the object in contact with the gripper by comparing the measured contact height with the learned height of the sub-frame.

Ultrasonic sensor control: The contact ultrasonic sensor is used for grading paper objects detected by the vision system with low confidence. The ultrasonic sensor measures the time delay between the emission and detection of the return signal traveling through the tested material. In order to facilitate accurate measurements two requirements must be fulfilled: the contact force of both ultrasonic probes is identical and equal to the set value, and secondly that entire contact area of both ultrasonic probes is in contact with the tested objects. As single ultrasonic measurements may be erroneous, additional filtering is needed in order to eliminate noise. A maximum slew rate (MSR) non-linear filter [2] was used, applied to series of 10 samples (sampled at 2 Hz).

Vacuum gripper control: The vacuum gripper consists in four vacuum cells each containing 3 suction cups and was designed such that each cell will be able to grip independently from the other cells. This design was used in order to enable gripping of planar objects of different sizes independently. The control of the gripper is a task of the ADEPT computer. In order to grip an object the vision system computes the location, orientation and size of the object in ADEPT coordinate system. Depending on the estimated size of the object the ADEPT computer may decide to grip it using four, two or

one cell. The vacuum, in the vacuum cells, is actuated/de-actuated by the ADEPT computer.

3. 3. Data Fusion and Decision Module

This section describes the data fusion scheme and the decision algorithms used for detecting paper objects and discriminate between the different grades. The principle diagram of the data fusion paradigm and decision algorithm used is presented in Figure 7. The decision algorithm is based on a multi-level active sensing process. The active sensing includes action and sensing tasks. The actions are performed by the ADEPT arm and the sensing is performed by Stereo Vision, Ultrasound and Color Vision sensing. The action tasks are: Gripping, Dropping, Touching Object and Out-of-Scene, and the sensing tasks are: Detecting Visual Features, Detecting Color of Objects, Measuring Ultrasonic Delay. A similar data fusion paradigm is used in Automatic Target Recognition (ATR) systems [8].

The data fusion and decision algorithm may be described as a feedback control process including: Perception, Response Action (feedback signal) and Decision. This closed loop system is in fact a Reasoning Intelligent Controller which combine the perceptions and actions in order to make an assertion and to validate the hypothesis. The first stage of the algorithm is the *Sensor Image Processing* of the stereo vision sensor. At the end of this stage the objects of interest are detected and separated from the background. The *Geometrical Parameters* (Physical Variables) of the objects of interest, are computed using the a priori calibrated parameters of the CCD cameras pair. Furthermore, additional *Algorithms* are used to estimate the relevant *Shape and Texture Features*. The features are analyzed and the object is classified using a *Continuous Inference Scheme* based on a priori Supervised Learning. As a result of the classification process, the objects are partitioned into *Representational Classes(A to E)*. In the next stage a decision is made concerning the Actions (manipulations) to be taken in order to clarify the behavior type of the object by analyzing the response of the sensors to the Actions. A specific action is associated to each behavior type, i.e. Ultrasonic Sensing (*Behavior Type* α), Color Sensing (*Behavior Type* β), Remove Object (*Behavior Type* γ). Depending on the type of the Action the Vision Computer send commands to the execution level, i.e. the ADEPT computer. The ADEPT arm executes the Action and decide about the Class and Behavior Type of the object as indicated by the Ultrasonic sensor output (*Direct Feedback Action*) or gets out of the scene and enable the Vision computer to analyze object's color (*Indirect Feedback Action*). As a *Response* to the color analysis a decision is taken, by the Vision Computer, concerning the Validation/Rejection of the initial hypothesis of the Vision sensor.

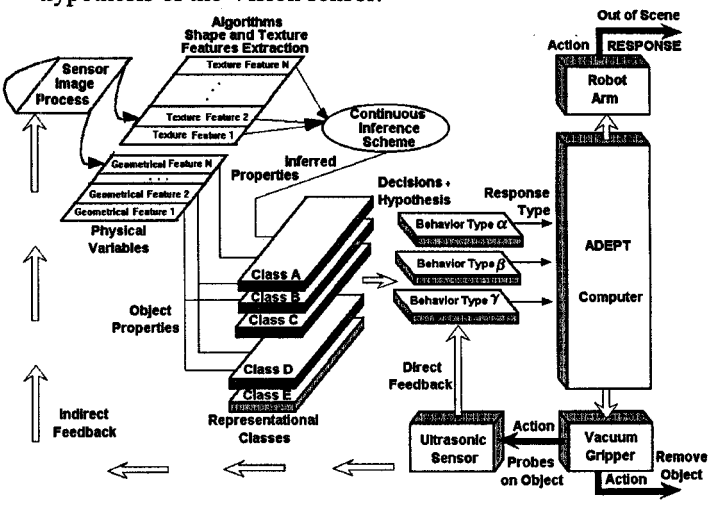

Figure 7 : Data Fusion and Decision Paradigm

4. Experimental Results

The software implementation of the algorithmic modules was a critical task. The software was written in Matlab and Borland C. All the Image Processing and Vision algorithms were implemented under Windows. The ADEPT arm control algorithms were written in Turbo C.

Parameter Tuning: The performance of the implementation in software is dramatically influenced by the parameters of the sensing scheme. The relevant parameters are: *Object detection threshold* used for detection of an object in a sub-frame; *Object content threshold* used for finding the pixels belonging to the interior of an object; *Color threshold* used for detecting the area of the unprinted region of a paper object; *Noise threshold* used for selecting a noise free area of the image of an object; and *Ultrasonic threshold* used for paper/non-paper object discrimination from ultrasonic measurements. Another set of parameters are the weights of the linear Fisher classifier. A linear classifier is not the optimal solution. Non-linear classifiers need to be used for improving the classification. Figure 8 presents a 2D features space and shows the linear and non-linear solutions. The linear solution leaves paper objects classified as non-paper and the other way.

Experimental Results: The experiments were conducted for the next purposes: find the optimal parameter set for best detection and grading performance, evaluate the performance of the system, test the robustness of the algorithms when applied to real objects picked from Municipal Solid Waste stream and prove the feasibility of the proposed detection and grading method.

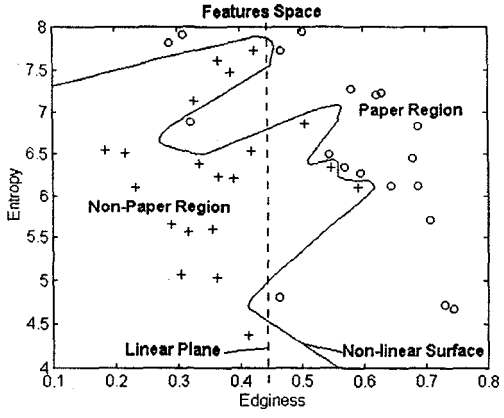

Figure 8: Linear and non-linear classifiers

The results of the experiments were very useful for improving the image processing algorithms and tune the parameters. The problems of image processing were: non-uniform illumination, segmentation of dark objects with low reflectance (cloths) and detection of the bounds of the sub-frames. At the end of the image processing stage a set of 50 images of segmented objects was defined and used as the learning set for the vision system. The false detection rate of paper objects was of 8% and the false alarm rate of non-paper object as paper was 2%.

An example of the image processing results is the object presented in Figure 5. The object detection threshold was 24 while the best fitting threshold value was 30 (the first local minim in the histogram of Figure 5d). Using this value all the pixels belonging to the object were detected with no false pixels. The position estimation errors were (0.012, 0.02, 0.002) m and 2 deg. Those values are enough for the robot to place the ultrasonic sensor in the correct measuring location and orientation. The resulting value of the cost function was 1.01 (<1.2), corresponding to an object with unknown structure. Additional ultrasonic tests were needed for sorting. The result of the ultrasonic sensing was 69 nano-seconds which is in the range of paper objects.

5. Conclusions and Remarks

As a result of the experimental results two types of improvements appear to be critical to the performance of an effective industrial automated paper recycling system: (i) the sensing needed for detection and grading; and (ii) the manipulation of the paper objects.

The vision system is the core sensing device for both detecting and grading of paper objects. The image processing algorithms are robust to small changes in illumination and color balance, but the confidence of the detection is reduced, due to high image noise level. The adaptive threshold mechanism needs to be improved using additional parameters, and not only the object color as is done now. Improved algorithms should be used for automatic tuning of the image processing parameters.

The stereo vision, based on photogrammetry, proved to be precise enough for the needs of such a system, but it is too slow (80 msec/sub-frame) for industrial applications. The computational time was acceptable but 90% of the time was spent on the 3D position computations.

The additional ultrasonic sensor used in the experiments was an off-the-shelf item and was not tuned for this specific use. Even so, the ultrasonic sensing functioned well. The detection of paper objects was correct in most of the cases (90%) and in all the experiments no non-paper object was detected as paper.

The manipulation of paper objects using the vacuum gripper was efficient. The suction power was enough to grip 50 paper sheets at a time. The gripper was able to pick small as well as large objects including books, wood, metal, plastic and glass objects.

It is clear that, in order to cope with large amount of objects, as needed by the industry, contact manipulation and sensing needs to be eliminated. The learning based vision sensing alone, will be able to detect paper and non-paper objects and also separate the different grades of paper using non-linear classification algorithms.

References

[1] H. Bacakoglu, *Two-Step Camera Calibration for Dual Camera Photogrammetry*, M.Sc Thesis, University of Waterloo, 1995.

[2] S. Faibish and I. Moscovitz,"A New Closed-Loop Non-Linear Filter Design",in *Proc. of 1-st European Cont. Conf.*, Grenoble, France, July,1991.

[3] R. A. Fisher, "The use of multiple measurements in taxonomic problems", *Ann. Eugenics*, vol. 7, 1950.

[4] J. Illingworth and J. Kittler,"A survey of the Hough Transform", *J. of Computer Vision, Graphics, and Image Processing*, Vol. 44, pp. 87-116, 1988.

[5] N. Otsu, "A threshold selection method from gray level histograms", *IEEE Trans. Syst. Man Cybern.*, vol. SMC-9, no. 1, pp. 62-66, 1979.

[6] Y. C. Shiu and S. Ahmad,"Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX=XB," *IEEE Trans. on Robotics and Automation*, Vol. 5, No. 1, pp.16-29, February 1989.

[7] M. A. Sid-Ahmed and M. T. Borai,"Dual Camera Calibration for 3-D Machine Vision Metrology, " *IEEE Trans. on Instrumentation and Measurement*, Vol. 39, No. 3, pp. 512-516, June 1990.

[8] J. A. Stover, D. L. Hall and R. E. Gibson, "A Fuzzy-Logic Architecture for Autonomous Multisensor Data Fusion", *IEEE Trans. on Ind. Electronics*, vol. 43, no. 3, pp. 403-410, June 1996.