DISCLAIMER

CONF-960405--5

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

AN ASYNCHRONOUS MODULATION/DEMODULATION TECHNIQUE FOR ROBUST IDENTIFICATION OF A TARGET FOR 3-D POSE DETERMINATION

Regina K. Ferrell
W. Bruce Jatko
David N. Sitter, Jr.

Oak Ridge National Laboratory*
Instrumentation and Controls Division
P.O. Box 2008, Oak Ridge, Tennessee 37831-6011

March 1996

To be presented at
AeroSense '96 Symposium
Signal Processing, Sensor Fusion, and Target Recognition V
Orlando, Florida
April 8-12, 1996

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-96OR22464. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

*Managed by Lockheed Martin Energy Research Corp. for the U.S. Department of Energy under contract DE-AC05-96OR22464.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

An asynchronous modulation/demodulation technique for robust identification of a target for 3-D pose determination

Regina K. Ferrell
W. Bruce Jatko
David N. Sitter, Jr.

Oak Ridge National Laboratory*
P.O. Box 2008, Bldg. 3546, MS 6011
Oak Ridge, TN 37831-6011

ABSTRACT

Engineers at Oak Ridge National Laboratory have been investigating the feasibility of computer-controlled docking in resupply missions, sponsored by the U.S. Army. The goal of this program is to autonomously dock an articulating robotic boom with a special receiving port. A video camera mounted on the boom provides video images of the docking port to an image processing computer that calculates the position and orientation (pose) of the target relative to the camera. The control system can then move the boom into docking position. This paper describes a method of uniquely identifying and segmenting the receiving port from its background in a sequence of video images.

An array of light-emitting diodes was installed to mark the vertices of the port. The markers have a fixed geometric pattern and are modulated at a fixed frequency. An asynchronous demodulation technique to segment flashing markers from an image of the port was developed and tested under laboratory conditions. The technique acquires a sequence of images and digitally processes them in the time domain to suppress all image features except the flashing markers. Pixels that vary at frequencies within the filter bandwidth are passed unattenuated, while variations outside the passband are suppressed. The image coordinates of the segmented markers are computed and then used to calculate the pose of the receiving port. The technique has been robust and reliable in a laboratory demonstration of autodocking.

Keywords: autodocking, object recognition, target identification, digital signal processing

1. INTRODUCTION

In some applications, it is desirable or even necessary to exploit special-purpose machines to supplement a human's abilities. A familiar example of such an application is teleoperated robots, where an operator controls the motion of a remote robot. To grasp an object with the robotic gripper, the operator advances the remote arm while viewing the approaching target on an external camera. The operator has an inherent handicap because the camera is a 2-D device and the robot moves in a 3-D environment. Typically, multiple cameras are used to give an additional angle of view, but this method is cumbersome, forcing the operator to observe two or more video displays during the approach.

Oak Ridge National Laboratory (ORNL) has investigated computer-assisted autonomous docking for a 6-D.F. robotic arm. The goal of the ORNL effort is to develop a system that will augment the operator's abilities by determining the pose of a docking port. The pose consists of the location of the reference axes in 3-D space and the orientation about those axes. By precisely determining the pose of the docking port, the robot can be instructed to move to the docking position without operator intervention. We refer to this "hands-off" operation as autodocking.

*Managed by Lockheed Martin Energy Research Corp. for the U.S. Department of Energy under contract DE-AC05-96OR22464.

The components of our autonomous docking system, shown schematically in Fig. 1, are (1) a robotic arm, (2) a motion control system to move the robot to the desired position, (3) a docking port, and (4) a measurement system.

The measurement system uses a 2-D video camera mounted at the end of the robotic arm so that the camera is aimed in the direction of forward motion. The vision system reconstructs a 3-D representation of the docking port from the camera output. To calculate the pose from the 2-D camera, the vision system performs the three functions shown in Fig. 2: (1) image capture, (2) target identification, and (3) pose determination.

Target Illumination. The target illumination function in Fig. 2 is accomplished with an array of light-emitting diodes (LEDs) attached to the docking port. The LEDs are image features that mark the major vertices of the target (port). The vision system uses the LEDs to uniquely identify the docking port from the surroundings.

Image Capture. A sequence of 20 video images of the docking area is captured with a 512×512 progressive scan video camera and a 6-mm fixed-focal-length lens. The camera signal is digitized to 8 bits by a digital image processing board.

Target Identification. The purpose of the target identification function is to isolate the LED fiducial markers on the docking target and determine the spatial coordinates of each marker. The LED markers are segmented from the background by a digital demodulator that is described in the following section.

Pose Determination. The pose determination function calculates the position and orientation of the target from the image coordinates of the LED markers. Each detected marker is evaluated according to size, shape, and relative location to confirm that it is an LED and not an artifact. The host computer aligns a digital wire-frame model of the target with the true vertices, represented by the LEDs in the camera image.¹

Fig. 2. Functions of the autodocking vision system. The 2-D camera sends video images to the autodocking vision system. The image capture function digitizes a sequence of 20 2-D image frames for processing by the target identification function. The pose determination function calculates the 3-D pose of the port from the image coordinates of the LED target markers.

Fig. 1. An autonomous docking system.

In the initial implementation of the target identification algorithm, the vertices were marked with statically illuminated, infrared LEDs. An optical bandpass filter, placed over the camera lens, combined with a global intensity threshold was used to segment the LEDs from the background. This method proved to be highly sensitive to nearby fluorescent and incandescent fixtures that also emitted significant infrared energy.

An improved method for target identification was developed to reduce the possibility of falsely detecting ambient light sources in the image. The target LEDs were modified to be strobed at a fixed frequency. A digital demodulation algorithm was developed to filter the flashing LEDs from the background. The demodulator is necessary because the LED modulation is asynchronous to the camera sampling rate in both frequency and phase.

2. TARGET IDENTIFICATION THEORY

Consider the case with a stationary camera viewing the target. The signal series obtained from a single pixel, $p(x,y)$, focused on a flashing LED will alternately be high and low, representing changes in the LED illumination. This pixel, shown by the circle in the lower left corner of each frame in Fig. 3(a), will vary at the LED frequency, as will other pixels focused on an LED. Figure 3(b) shows the time-domain variation in intensity for the single pixel in Fig. 3(a) over seven frames.

Let $p(n)$ represent the intensity of any pixel location x, y in the image, where n is the frame number. If $p(n)$ varies at the LED frequency ω_0 , then its Fourier transform $P(\omega)$ will have a strong response at that frequency. We desire to segment the target markers from the camera image, retaining the pixels representing LEDs and discarding all others. One technique is to shift each pixel signal in the frequency domain and then filter the signal, as shown in Fig. 4.

Fig. 3. Variation in pixel intensity over sequential video frames. In (a), the small circle represents a single pixel, $p(x,y)$, in each of the n camera frames. The camera is pointed such that a flashing light is illuminating $p(x,y)$ but not the surrounding pixels. The time-variation of $p(x,y)$ is shown in (b).

Fig. 4. Block diagram of the digital demodulator. Illustration of the method to isolate the flashing LED markers in the camera video.

If $p(n)$ has a Fourier transform $P(\omega)$, then a frequency shift of ω_0 is a translation along the frequency axis that is represented by the Fourier pair:

$$P(\omega + \omega_0) \Leftrightarrow p(n)e^{-j\omega_0 n} \quad (1)$$

Let $h(n)$ be the time-domain representation of a low-pass windowing function and let $H(\omega)$ be its Fourier transform. The frequency-shifted signal $P(\omega + \omega_0)$ can be filtered by multiplying by $H(\omega)$

$$P(\omega + \omega_0)H(\omega) \Leftrightarrow [p(n)e^{-j\omega_0 n}] \otimes [h(n)] \quad (2)$$

where $\otimes$ denotes the convolution operator. Note that Eq. (2) is not a spatial convolution frequently seen in image processing but a temporal convolution, more often associated with 1-D signal processing.

The exponential term in the right-hand side of Eq. (2) can be expanded into cosine and sine components to give

$$P(\omega + \omega_o)H(\omega) \Leftrightarrow \{p(n)[\cos(\omega_o n) - j\sin(\omega_o n)]\} \otimes h(n) \quad (3)$$

The magnitude of Eq. (3) will be large if it arises from pixels viewing an LED flashing at the reference frequency ω_o . Conversely, the magnitude will be small if it is from pixels that are unchanging or varying at another frequency. In this application, we are only interested in the magnitudes of the signals and can ignore their phase components. Let $g(n)$ be the magnitude of Eq. (3), the frequency-shifted and filtered signal, where

$$g(n) = \sqrt{[p(n)\cos(\omega_o n) \otimes h(n)]^2 + [p(n)\sin(\omega_o n) \otimes h(n)]^2} \quad (4)$$

3. IMPLEMENTATION OF THE TARGET IDENTIFICATION ALGORITHM

Because the sampling rate of the camera is 30 Hz, then, by the Nyquist criteria, the highest frequency that can be accurately resolved is 15 Hz. The target LEDs are strobed at 12 Hz, corresponding to a digital frequency of $\omega = 2\pi fT$, where f is the flash rate and T is the sampling period. The 12-Hz modulation frequency was chosen to minimize modulation artifacts with the 30-Hz video sampling rate.²

The input $p(n)$ is convolved with low-pass filter $h(n)$ in Eq. (2). The low-pass filter was implemented with a Hamming window because of its sharp frequency attenuation beyond the frequencies of interest. The Hamming window was implemented in the following form:

$$h(n) = 0.54 - 0.46 \cos\left(2\pi \frac{n}{N-1}\right), \quad 0 < n < N-1 \quad (5)$$

where N is the total number of samples (images) in the set.³

Note that when Eq. (5) is substituted into Eq. (4), then $\cos(\omega_o n) \otimes h(n)$ is constant for any frame n . Those terms can be predetermined for each frame and stored in a lookup table we designate $K_c(n)$. Similarly, the terms for $\sin(\omega_o n) \otimes h(n)$ can be predetermined and stored in lookup table $K_s(n)$. Because we are only interested in the signal magnitude, we can ignore the phase components arising from the convolutions in the right-hand side of Eq. (4). When the phase components are dropped, the convolutions in Eq. (4) can be expressed as the sum-of-products,

$$p(n)\cos(\omega_o n) \otimes h(n) = \sum_{n=0}^{N-1} p(n)K_c(n) \quad (6)$$

and

$$p(n)\sin(\omega_o n) \otimes h(n) = \sum_{n=0}^{N-1} p(n)K_s(n) \quad (7)$$

We now substitute Eqs. (6) and (7) into Eq. (4) and simplify to the computationally simpler form,

$$g(n) = \sqrt{\left[\sum_{n=0}^{N-1} p(n)K_c(n) \right]^2 + \left[\sum_{n=0}^{N-1} p(n)K_s(n) \right]^2} \quad (8)$$

Equation (8) is implemented by the block diagram in Fig. 5 with the following sequence: (1) multiply each image frame by the predetermined sine and cosine constants for that frame, (2) sum the results into separate accumulators, (3) square the contents of each accumulator, (4) add the contents of the two accumulators, and (5) take the square root of the accumulator sum. The algorithm was implemented with hardware lookup tables for both the trigonometric and square root operations described above.

Fig. 5. Signal flow diagram for the digital demodulator.

One frame from a sequence of 20 consecutive frames is shown in Fig. 6(a). Processing the entire sequence using Eq. (8) results in the image of Fig. 6(b). The intensity values for the nonflashing background pixels were less than 2 units out of 255. The frequency response of the algorithm has a bandwidth of 4 Hz, centered at 12 Hz, as shown in Fig. 7.

4. ACCURACY

In this section, we describe the results of a measurement accuracy test by comparing the measured coordinates of the LED markers with their known positions.

The vision system determines the coordinates of the LEDs relative to the camera coordinate frame which originates inside the lens and is physically inaccessible. Therefore, we measure the accuracy indirectly, moving the camera by a known amount on a calibrated stage and comparing the change in the calculated LED position. We compute the measurement error ϵ in percent of the actual value by

$$\epsilon = \frac{\rho(\text{measured}) - \rho(\text{actual})}{\rho(\text{actual})} \cdot 100 \quad (9)$$

where p is the position coordinate of the camera. The camera was positioned perpendicularly to the target and moved in 5-mm increments toward the target. A total of 30 measurements was recorded beginning at a distance of 400 mm and ending at 105 mm. The mean positional error, computed according to Eq. (9), was -0.3%; the median error was -0.35%, and the maximum error magnitude was 0.9%.

Fig. 6. Photograph of the input signal to the demodulator and the resultant output. At left is one frame $f(19)$ of a six-point target. The right is the resultant image after filtering.

Fig. 7. Frequency response of the digital demodulator.

5. SUMMARY

A method of identifying modulated LED markers in a sequence of video images was developed and tested. The technique uses a digital demodulator to filter the markers from their background. The demodulator was developed for an autodocking vision system designed to locate a docking target and compute its 3-D position and orientation coordinates.

6. ACKNOWLEDGMENTS

This research was sponsored by the U.S. Army's Project Manager, Crusader program, under Interagency Agreement 1892-A078-A1 between the Department of Energy and the Armament Research, Development and Engineering Center at Picatinny Arsenal.

7. REFERENCES

1. J. S. Goddard, W. B. Jatko, R. K. Ferrell, and S. S. Gleason, "Robust Pose Determination for Autonomous Docking," ANS 6th Topical Meeting on Robotics and Remote Systems, Monterey, Calif., February 1995.
2. D. N. Sitter, Jr., J. S. Goddard, and R. K. Ferrell, "Method for the Measurement of the Modulation Transfer Function of Sampled Imaging Systems from Bar-Target Patterns," *Applied Optics*, **34**(4) (February 1995).
3. A. V. Oppenheim and R. W. Shafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1975.