Upper Limb Position Sensing: A Machine Vision Approach

Dianna Han[1, 2]
dianna@cs.ucsb.edu
diannah@aemf.org
[1]Department of Computer Science
University of California at Santa Barbara
CA 93106

Doug Kuschner[2]
dougk@aemf.org

Yuan-fang Wang[1]
yfwang@cs.ucsb.edu

[2]Alfred Mann Foundation
25134 Rye Canyon Loop Suite 200
Santa Clarita, CA 91355

Abstract – **Numerous approaches to sensing limb position for controlling neural prostheses have been proposed, evaluated and even incorporated into commercial products. In general, these technologies have focused on the goals of accuracy, convenience and cost. Here we propose an approach to sensing upper limb posture for a stroke rehabilitation system that does not require any devices attached to the subject. This is achieved through the use of a machine vision approach, which involves focusing a digital video camera on the subject and processing the video stream using a specialized algorithm running on a PC. This algorithm will produce a trigger signal whenever the arm posture conforms to a predefined profile. While the approach itself can be applied to a variety of sensing and control applications, we have demonstrated it by developing and characterizing an algorithm that can accurately sense elbow flexion and extension. The machine vision algorithm performs 3-D recovery of the arm position and calculates the elbow angle accordingly, which we have compared to a commercially available goniometer. It also involves a model based prediction and correction technique that improves the accuracy where the model is trained at the outset of a sensing session. The system uses a commercial off-the-shelf webcam, which is widely available and cost effective. The experiments were done in vivo, and the results have shown that the accuracy of the system is about 90% accurate on average compared to our benchmarking device, and that it has strong potential to facilitate control of neural prostheses.**

I. Introduction

The application of stimulating the muscles in the upper arm to produce elbow flexion and extension has led to the development of various stimulation and control strategies. These mechanisms range in terms of their sophistication. In general, three essential elements are required for the design of a useful neural prosthesis and these are sensing, stimulation and control. Sensing involves some means by which to detect either the command to induce a movement coming from the subject or the result of that movement, i.e. the position of a limb. Control is the pattern by which the sensing inputs are mapped to stimulation commands. It is very common for the sensing subsystem to produce a trigger signal that would either start and stop some stimulation or serve as an input to a finite state machine controller. The sensors would typically produce this trigger signal whenever certain physiological conditions are met.

With regards to an upper arm neural prosthesis, many systems have been developed with varying degrees of success. In these systems, most of the mechanisms for controlling the stimulation have involved only volitional control by the patient. Examples include taking signals from contralateral shoulder movements [1], joystick control [2], respiratory mechanisms [3], wrist movement [4] and even the patient's voice [5]. The development of the above systems has greatly contributed to our understanding of upper extremity control and may be of great benefit to a Spinal Cord Injury (SCI) patient. However, the above systems rely purely on volitional control by the patient. This is not of much use for a stroke patient who would like to use these systems for motor relearning, as that requires a more natural sensing mechanism. Attempts at using residual myoelectric signals as a more natural command source for controlling neural prostheses [6][7] have been made and even though this has many advantages, it requires daily calibration by the hemiplegic patient which limits its practicality.

Against this background we have conceived and implemented a sensing mechanism for an upper extremity neural prosthesis that is based on machine vision technology. Machine vision is the use of computer algorithms for processing video stream data to extract useful information. The system involves the use of an ordinary web-camera (webcam) connected to a mid-range laptop computer to derive the essential trigger points for facilitating a reach-and-grasp exercise for a hemiplegic stroke recovery patient.

One distinct advantage is that it involves no devices connected to the subject's body. This is particularly important when considering the difficulty that hemiplegic subjects have in donning attached sensors and the inherent inaccuracy introduced by their motor limitations. Machine vision can work with or without colored markers attached to the subject's body. It is able to recover the 3-D position information from the camera image. Also, considering it is used to detect a very predictable motion of the arm, it can maintain a model of the movement and that model can be trained for greater accuracy of the sensing system.

The remainder of this paper is organized as follows: Section II, Machine vision based sensing, which gives an overview of the system; Section III, 3-D recovery from 2-D video, which presents the theoretical basis of 3-D information recovery from image sequences obtained by a single camera; Section IV, Model guided tracking and joint angle measurement, which explains how the tracking is performed based on prediction and correction, and how the important parameters required in stimulation control such as joint angle are retrieved; Section V,

Experimental results, which includes the experiments done with comparison to traditional goniometers; Section VI, Conclusion.

II. Machine vision based sensing

Typically, the sensing unit in a Functional Electrical Stimulation (FES) system keeps track of body positions, recognizing certain postures and detecting trigger events. When these triggers are detected, the sensing systems deliver alerts or notification signals to the stimulation controller. One effective way of representing body postures is by a combination of joint angles. Consequently, sensing systems for neural prostheses are often designed to obtain this information.

The most common method of obtaining the joint angle information during use of a FES system is through the use of physical trackers attached to the subject's body. However, with the machine vision based approach, the physical trackers are replaced by optical markers, cameras and a computer. The detected data change from electronic signals to captured video making it remarkably advantageous, as it is less intrusive and very convenient.

By processing the video or image sequences, 2-D positions of the optical markers in the image plane are determined with a certain degree of accuracy. However, due to the possible change in view point, 2-D information is not enough for distinguishing 3-D postures. In order to discriminate ambiguous projections it is necessary to obtain depth information.

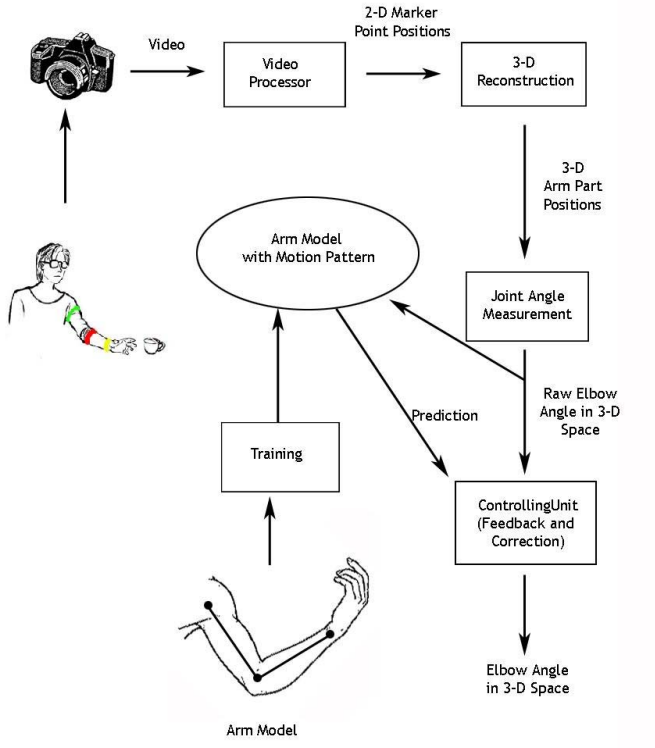

Fig.1.The System Structure Overview

Obtaining this depth information would be most accurately and conveniently done using multiple cameras, i.e. stereovision. However, our current experiment is based on a single-camera approach. Using only one camera for retrieval of depth information requires additional processing. The details of the 3-D recovery operation are explained in Section III.

Although recovery of 3-D coordinates is possible with only one camera, there are many other aspects in implementation that need to be considered. Because of the complicated recovery algorithm, the 3-D recovery procedure is sensitive to the noise in the system that is introduced by the video itself and 2-D position extraction. Furthermore, the algorithm is only applicable to rigid objects and body postures do not really satisfy this constraint. To make the sensing possible, stable and robust, we introduce an arm model to guide and correct the sensing procedure. Using the special pattern of arm movements in our reach-and-grasp exercise, it is possible to train the arm model to keep the motion information. Furthermore, prediction can be made based on this a priori knowledge and observation can also serve as feedback to correct or affect the future prediction. Therefore, the model can be considered as a smoothing and guiding component, which serves the purpose of reducing the noise effect and compensating for the fact that the arm is not rigid.

The system structure and processing procedure can be demonstrated in the flow chart shown in Figure 1.

III. 3-D recovery from 2-D video

The primary function of the system is to measure joint angles based on input from the body position sensors. However, with the machine vision based approach, the system can only acquire the video recording of the body movements, or more specifically in our project, the arm movement. As a result of this 3-D to 2-D mapping, one dimension is lost, thus different body postures can have the same projection on the image plane when viewed from different perspectives, and vice versa. Since the same joint angle may look different in the video from different perspectives, the system must be able to get the actual joint angle value regardless of where the camera is placed. To achieve this goal, 2-D position information of the optical markers in the video is not sufficient. 3-D coordinates must be established which requires depth information that is missing in the 2-D video.

There are various approaches to addressing this problem. For instance, stereovision is the most natural way of extracting depth information. The principle in stereovision is to have at least two cameras focused on the same object. Depth information is then recovered based on the disparity between corresponding points in images taken by these cameras. However, in our application we use a single camera and hence a different mechanism is used to recover the depth information. This approach is based on the fact that motion itself is relative and so as long as the subject is in motion it is possible to transform the information and solve for it as one would solve for a standard stereovision problem.

One first notes that it makes no difference in the video data whether the subject moves or the camera moves. Hence, to measure the depth in the frame at time t_1, we need to use the subsequent frame at time t_2 (the time difference between t_1 and t_2 can be varied as it depends on the sampling rate of the

camera). By dividing the arm into three parts each part can be roughly considered as rigid. With the assumption of rigidity, the transformation of each part from time t_1 to time t_2 can be determined by using the correspondence of eight points between these two images, which is called the eight-point algorithm [9]. The algorithm solves the parameters in the fundamental matrix of the camera system.

Once the subject's motion is established, the problem can be considered in a new way. As mentioned above, with a static subject and a moving camera, we can obtain the same video as with a static camera and a moving subject. Therefore, we invert the subject's motion extracted by the eight-point algorithm and transform it into virtual camera motion. Assuming now that the subject is stationary, we know the relationship between the positions of the original camera and the virtual camera and we also have the images taken by these two cameras respectively. With this approach, the problem is transformed into a standard stereovision problem and the depth information can be retrieved using this method.

Figure 2 demonstrates the principles of this approach.

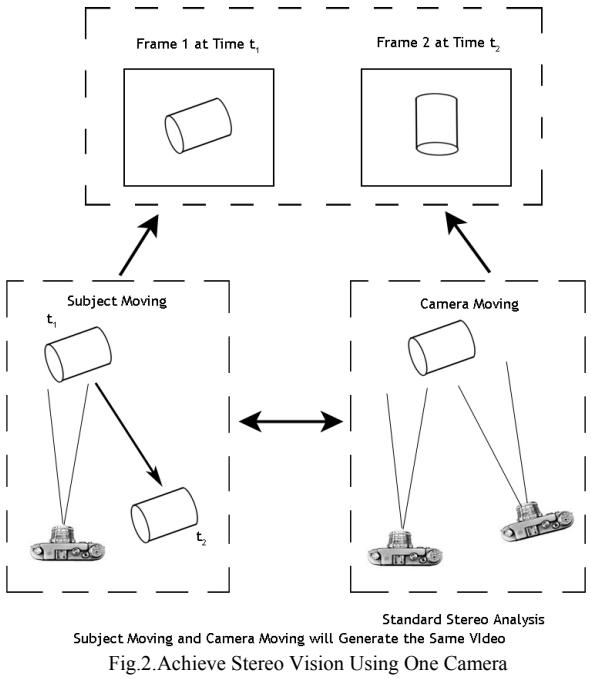

Fig.2.Achieve Stereo Vision Using One Camera

IV. Model Guided Tracking and Joint Angle Measurement

To measure joint angles in 3-D space, the depth information extracted using 3-D reconstruction must be accurate. However, noise in the video capture process is inevitable due to the problems of camera imaging such as distortion and color aberration. Therefore, the images to be analyzed will always contain inaccuracies that make establishing the point correspondences in adjacent frames less reliable. Also, although we divide the arm into three different parts to satisfy the rigid body assumption required by the eight-point algorithm, the human body is still deformable. All of these factors lead to the fact that the system is very sensitive to noise.

To reduce the negative effect of noise and make the system more reliable and robust, we introduce an arm model to save the possible motion pattern information. This model is used to guide the tracking procedure. Since the movement of arms in a reach-and-grasp exercise is highly predictable and repetitive, it is very easy to train the model to record these patterns in advance. During practice the model calculates predictions about the expected joint angles and arm position throughout the movement. These predictions are then used to correct possible errors introduced by the system noise. At the same time, the observed arm position is passed to the controller as feedback, where the predictions given by the model are incorporated with the observation. With this kind of standard prediction and correction procedure, the noise effects are minimized and the system performance is enhanced.

Since the main purpose here is to record the parameters related to the motion pattern, such as joint angles, the arm model is currently built in a simple and intuitive way using sticks and links. This is different from the detailed muscle and bone models widely used in the stimulation stage. However, it is highly likely that these two kinds of models can be integrated and that a more realistic model would help to improve the system performance.

Figure 3 shows the controlling mechanism.

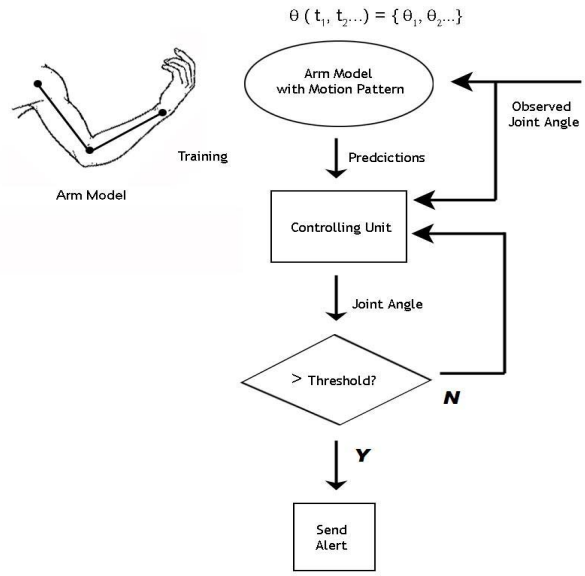

Fig.3.Model Guided Tracking Mechanism

V. Experimental Results

A variety of experiments were done with video recordings of several trials of a reach-and-grasp movement as input to the machine vision algorithm. The software was run on a Pentium IV, 1.70 GHz CPU, 256M RAM computer. With video files recorded by a camcorder or video captured by a webcam, the system was run in real time with its accuracy benchmarked against the Shape Sensor™ S700 which is a commercial goniometer product manufactured by Measurand Inc.

The system allowed for at least 60 degrees per second of angular velocity and worked under various input video resolutions ranging from 320*240 to 720*480.

A. Video Files as Input - Offline

Fifteen video files were recorded by a SAMSUNG SCD67 camcorder with a resolution of 720*480. The files were then compressed to MPEG-4 format, with one key frame per second. The video clips have an average length of two minutes, in which people are performing elbow flexion and extension repeatedly. The following table shows the results of trigger event detection, in which the actual number of trigger events was counted by human operators.

TABLE I
THE ACCURACY PERFORMANCE

	Type of Trigger Events	Number of Trigger Events	Number of Correct Alerts	Accuracy
File 1	Extension	31	31	100.00%
File 2	Extension	25	23	92.00%
File 3	Extension	34	30	88.24%
File 4	Extension	30	29	96.67%
File 5	Extension	31	30	96.77%
File 6	Extension	32	32	100.00%
File 7	Extension	27	26	96.30%
File 8	Extension	32	30	93.75%
File 9	Extension	34	33	97.06%
Total/Avg.		276	264	95.65%
File 10	Flexion	29	26	89.66%
File 11	Flexion	30	28	93.33%
File 12	Flexion	34	29	85.30%
File 13	Flexion	29	28	96.55%
File 14	Flexion	32	29	90.63%
File 15	Flexion	30	28	93.33%
Total/Avg.		184	168	91.30%

B. Video Captured through a Webcam as Input – Real Time

Benchmarking of our system was done by comparing it to the Shape Sensor™ S700 manufactured by Measurand Inc. The input source was a Logitech webcam, which offered video with a resolution of 640*480. Four key degrees of elbow angles were selected, which were 45, 90, 135 and 180 degrees. The experiments were done over a total time of 30 minutes, which included about 3,000 times the arm reached the above key elbow angles. Instead of giving all the data (which is not necessary), the following table shows the joint angles detected by our system and the benchmarking device on average:

The computation required for angle detection is simplified when the arm is fully extended, as this is a pre-defined trigger event. The accuracy of the angular measurement is reduced for an arm that is not fully extended but is still quite high as shown in Table II.

TABLE II
BENCHMARKING AGAINST SHAPE SENSOR™ S700

Shape Sensor™ S700	45º	90º	135º	180º
Vision Sensor (Avg.)	37.6º	78.3º	122.7º	178.2º
Vision Sensor (Deviation.)	8.5º	9.2º	6.4º	4.9º
Accuracy	83.6%	87.0%	90.9%	98.9%

VI. CONCLUSION

In this paper we propose a machine vision approach for upper arm position sensing. The mechanism uses the video captured by a single camera as the sensor input and the measurement of joint angles is done by analyzing the image information. Experimental results indicate that our approach is comparable to industry standard accuracy with the added advantage that our approach does not require the hemiplegic patient to be connected to any restrictive devices. As mentioned in Section III greater accuracy could be obtained in the 3-D reconstruction by using a stereo-vision approach with multiple cameras In future work, we will consider the use of two cameras and hence standard stereo analysis will be applied directly with expected improvements in the accuracy of the measurements. With the easy availability of commercial products that perform stereo processing in real time we expect no significant technical difficulties with the future work.

Integration of the vision component into control systems is a widely accepted approach in many other application areas, such as robotics and human-computer interface design based on mixed reality. The use of vision based approaches in sensing applications for FES may hence be a possible system design choice in the future given that image analysis and processing are already widely used in biomedical applications.

REFERENCES

[1] Hart RL, Kilgore KL, Peckham PH (1998) A comparison between control methods for implanted FES hand-grasp systems. IEEE Trans Rehab Eng TRE-6:208-218

[2] Rudel D, Bajd T, Reberšek S, Vodovnik L (1984) FES assisted manipulation in quadriplegic patients. In: Popović D (ed.) Advances in external control of human extremities VIII. ETAN, pp 273-282

[3] Handa Y, Ohkubo K, Hoshiyama N (1989) A portable multi-channel functional electrical stimulation system for restoration of motor function of the paralyzed extremities. Automedica 11(1-3):221:232

[4] Prochazka A, Gauthier M, Wieler M, Kenwell Z (1997) The Bionic glove: an electrical stimulator garment that provides controlled grasp and hand opening in quadriplegia. Arch Phys Med Rehabil 78:608-614

[5] Nathan RH (1989) an FNS-based system for generating upper limb function in the C4 quadriplegic. Med Biol Eng Comp 27:549-556

[6] Thorsen R, Ferrarin M, Spadone R, Frigo C (1998) An approach using wrist extension as control of FES for restoration of hand function in tetraplegics. In: Proc 6th Vienna Workshop on Functional Electrostimulation

[7] Saxena S, Nikolić S, Popović D (1995) An EMG controlled FES system for grasping in tetraplegics. J Rehabil Res Dev 32:17-23

[8] David A. Forsyth, Jean Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, 2003, pp.216-219