Vision-based Segmentation of Continuous Mechanomyographic Grasping Sequences for Training Multifunction Prostheses

Natasha Alves and Tom Chau

Abstract— **In designing mechanomyographic (MMG) signal classifiers for prosthetic control, the acquisition of long, continuous streams of MMG signals is typically preferred over the painstaking collection of individual, isolated contractions. However, a major challenge with continuous collection is the subsequent separation of the MMG data stream into segments representing individual contractions. This paper proposes an automatic, vision-based segmentation method for continuously recorded MMG data streams. MMG data acquisition was synchronized with transverse plane video acquisition of functional grip sequences. The automatic segmentation system can track a hand, recognize grips and detect grip transition times as well as extraneous hand movements. The system recognizes two grips with an average accuracy of 97.8±4%, and seven grips with an accuracy of 73±20%. The contraction initiation and termination times agree closely (within 1.3 ± 1 frames) with values obtained manually.**

I. Introduction

When muscles contract they emit low frequency vibrations known as the mechanomyogram (MMG). MMG is the result of the summation effect of propagated motor unit twitches [1], and has several applications in muscle fatigue detection, muscle disease diagnosis [2] and prosthetic upper-limb control [3].

Long streams of MMG recordings are needed to design classifiers to control upper-limb prostheses, and to study the temporal dynamics of repetitive contractions. Individual contractions need to be segmented: to provide exemplars for the classifier in the former, and inter-contraction times in the latter. A method for segmenting MMG data is proposed in this paper. We focus on the application to prosthesis control, but argue that this method could also be used to provide accurate interevent times.

Pilot tests on the feasibility of detecting discernible patterns in MMG for multifunction prosthetic control are usually performed on able-bodied subjects since their intact musculature makes muscle site selection an easy task, and they can perform more classes of repeatable contractions than amputees. Voluminous recordings, in the order of 100 samples per class, are typically required to train classifiers for multifunction control. An automatic data segmentation procedure can make data analysis easier.

Several methods for MMG data segmentation are possible. One approach is to repeatedly stop and start data acquisition for each set of MMG recordings. Since the contraction sequence is always being interrupted, this method is not suitable for studying temporal dynamics in the MMG signal such as the effects of muscle fatigue due to repeated contractions. Moreover, the natural variability in repeated contractions could be artificially reduced as a result of isolating contractions.

A more viable option is to provide the participant with visual or verbal instructions on what contraction to perform while continuous streams of MMG signals are recorded. Since the signals exhibit a spike in amplitude and a burst of activity at the onset of a contraction, events may be detected from the signal itself. This method assumes that easily discernible characteristics can be detected in the signal. However, errors such as producing the wrong contraction and unexpected arm movements cannot be easily detected from the MMG signal itself. These actions contaminate the signal; if undetected, the accuracy of the classifier will be compromised, and if detected, the data collection process has to be restarted.

A third method for segmenting MMG data is to look at external cues of the participants' actions instead of searching for cues internal to the signal. A *vision* based segmentation method is proposed in this paper. This method is useful in providing approximate time segments from where the exact steady state and transient portions of the signal can be extracted. Continuous streams of MMG signals and images of the hand are recorded synchronously; this method streamlines the data acquisition of continuous grip sequences, preserves natural variability, and provides visual information to the recorder. Once the data and its synchronized video have been recorded, manually extracting segments from the data stream by using the video sequence can be extremely time consuming. The following work describes an automated system that can detect grip types, the times of grip initiation and completion, and unwanted hand movements. The end product is segmented and labeled MMG contractions suitable for designing and testing different features and classifiers.

Manuscript received April 3, 2006.

N. Alves is with University of Toronto, Canada (e-mail: natasha.alves@utoronto.ca).

T. Chau is with University of Toronto and Bloorview Kids Rehab, Toronto, Canada

II. Data Collection

Seven subjects were seated at a desk, and objects of different shapes were placed before them. Each object was associated with one of six categories of grips: pinch, palmar, lateral, hook, spherical, and cylindrical, as defined by Schlesinger [4]. The camera was positioned overhead, approximately 1 metre above the worktable, such that the subject's hand and the objects were in the centre of the camera's field of view. Six MMG sensors, manufactured according to the method of Silva et al. [5], were placed on the subjects' forearms. Objects were picked, grasped for 4s, and then released before a different object was picked.

A PXI (NI, PXI-8187) system was employed for data acquisition. Custom data acquisition software, written in LabView, was used to synchronize analog voltage acquisition by a DAQ card (NI, PXI-6052E) with image acquisition from a firewire camera (Sony, DFW-X710). MMG signals were sampled at 1KHz and the camera recorded color images of resolution 320x240 at 7.5 frames/s.

III. Automatic Data Segmentation Method

Once the data were acquired, timing information was automatically extracted from the video to determine the grip type and times of transition between grips. The steps in the data segmentation algorithm are outlined in Figure 1, and are detailed in the following sections.

A. Skin Color Classification

Skin color pixels are separated from non-skin pixels in the YCbCr colour space. Skin color was modeled as a bivariate distribution of the chrominance components, Cb and Cr, with μ_s as the mean vector, and Σ_s as the co-variance matrix [6][7]. μ_s and Σ_s were estimated from a sample of skin color pixels. Each pixel was defined by its feature vector **c** = [Cb Cr]^T [7]. The Mahalanobis distance (d_s) was calculated using:

$$d_s{}^2 = (c - \mu_s)^T \Sigma_s{}^{-1} (c - \mu_s). \qquad (1)$$

A small value of d_s indicates a high probability of skin detection. The skin detection threshold (τ_s) was determined by training the classifier with a sample of skin and non-skin pixels. The aforementioned distance was compared to τ_s to generate a binary skin detection mask (SDM) with elements:

$$SDM(i,j,k) = \begin{cases} 1\,, \textit{ if } \mathrm{d_s(i, j, k)} < \tau_s \\ 0\,, \textit{otherwise} \end{cases} \qquad (2)$$

where d_s(i,j,k) is the Mahalanobis distance of a pixel at spatial location (i,j) in frame k of the video sequence.

B. Hand Detection

The hand detector block uses the SDM to determine which group of skin-color pixels is most likely to be the hand of interest (i.e. forearm from which physiological data is being recorded), and to keep track of the coordinates of the hand.

Morphological closing (dilation, followed by erosion) was performed on the SDM to fill up small holes and connect nearby regions resulting in smoother boundaries of the objects. The refined SDM was then examined for blobs: groups of pixels organized into structures. A blob whose area was within a pre-defined range was considered a candidate hand and its position was recorded; other blobs were ignored. Isolated points and small regions of skin-like pixels were thus eliminated.

Unwanted skin-color regions that meet the area specifications of the hand, such as the subject's other hand or skin-like regions in the background, may be in the camera's field of view and need to be ignored. A Kalman filter was employed to keep track of the hand of interest from one frame to the next.

The Kalman filter estimates the state $(x \in \Re^n)$ of a process by using feedback in the form of noisy measurements $(z \in \Re^m)$. In this case *x* and *z* define the estimated and measured (from the bounding box) positions of the hand respectively. The iterative predictor-corrector process of the Kalman filter estimates the region of interest (ROI) of the hand, defined by its spatial location (*xo,yo*) and the size of its bounding box [*xsize ysize*]:

$$\mathrm{ROI_{hand}} = [\mathrm{xo\ yo\ xsize\ ysize}]. \qquad (3)$$

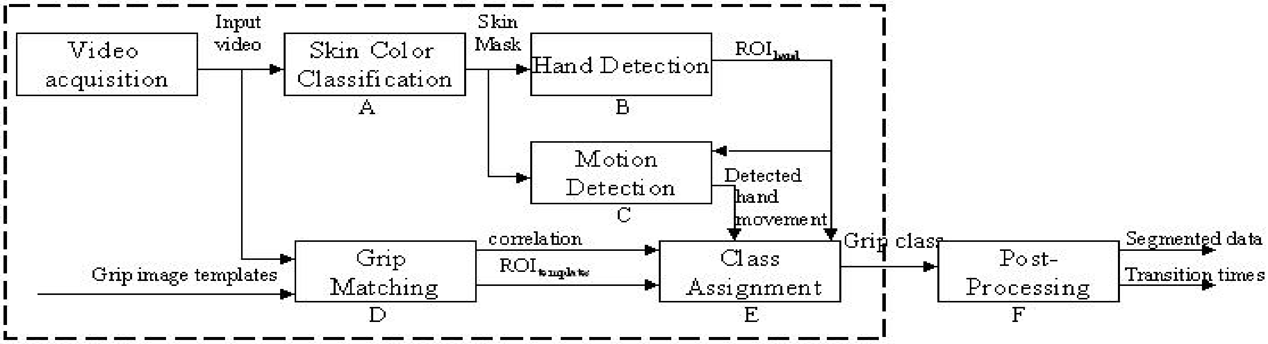

Figure 1: Overview of the automatic data segmentation algorithm

C. Motion detection

Movements of the hand were deteced by using the binary image of the masked hand to calculate the sum of absolute differences (SAD) between the current frame (k) and the previous frame [8][9]. SAD values decrease with increasing similarity between the two images. The SAD is given by:

$$SAD(k) = \sum_{i=xo}^{(xo+xsize)} \sum_{j=yo}^{(yo+ysize)} abs(SDM(i,j,k) - SDM(i,j,k-1)), \qquad (4)$$

where xo, xsize, yo, and ysize define the hand ROI of the current frame. The sensitivity of the motion detector was adjusted by altering the value of the motion threshold (τ_m). Motion is detected (d_m=1) if the SAD is greater than τ_m.

$$d_m = \begin{cases} 1 \text{ , } \mathit{if} \text{ SAD(k)} > \tau_m \\ 0 \text{ , } \mathit{otherwise} \end{cases} \qquad (5)$$

D. Grip Matching

The grip-matching block ranks each grip in order of their likeliness of being performed at that instant. Data acquisition was performed in a constrained environment where participants were asked not to change the orientation of their hand. Two-dimensional cross-correlations between each video frame and the grip templates were sufficient to determine the action of the hand.

The target grip templates were customized to each individual user and consisted of images of the user holding different objects. In order to reduce computational complexity, color images were converted to their equivalent grayscale intensity images and were resized by a factor of 1/3. The normalized cross-correlation (NCC) between frame *k* of the resized intensity video *I*, and grip template h (T_h) of dimensions [Mt_h, Nt_h], was calculated. The location (xo_{Th}, yo_{Th}) with the highest correlation (c) was chosen as the location of the pattern. The co-ordinates of the template were re-scaled by the magnifying factor (shrinking factor^{-1}) to determine the ROI of template h in the video frame.

$$\text{ROI}_{\text{Th}} = [\text{xo}_{\text{Th}} \text{ yo}_{\text{Th}} \text{ Mt}_{\text{h}} \text{ Nt}_{\text{h}}] \cdot \mathit{magnifying\ factor} \qquad (6)$$

It is important to note that the template matching block is restricted in it capabilities since it is not invariant with respect to imaging scale and rotation. If the images are subject to rotation, other pattern matching techniques like feature classification should be used [10]. However, these methods involve voluminous training examples, and were not necessary for this application.

E. Class Assignment

The gesture assignment block combines the outputs of the previous blocks to assign a grip class (class) to each frame (k) of the video sequence. The algorithm for grip assignment is as follows:

If motion is detected (i.e. d_m=1), then class(k)=motion

Else, choose the grip template with the highest correlation coefficient (c), whose ROI overlaps with the ROI of the hand:

1) For each template h, if ROI_{Th} does not overlap with ROI_{hand}, its correlation coefficient is ignored

2) The grip with the highest correlation coefficient is the most likely grip

3) If the maximum correlation is greater than a pre-defined threshold (0.95), the frame is assigned a grip; else, the class is unknown.

F. Post- processing

The class assignment block outputs a grip or motion class for each frame of the video sequence. The output is a string of numbers from 1-9, where 1-6, 7, 8, and 9 correspond to six grips, rest, motion and unknown classes, respectively.

Grip transition windows, which represent the periods of time when the hand is changing its grip, were determined by finding the frame numbers of the start and end of motion. Transition windows that occur in immediate succession, separated by 1-2 frames, were likely to be from the same grip transition and were therefore merged. Since only one class of grip was expected between two successive motion classes, the class of the data segment between transitions was the grip that is detected most frequently in that time interval. If a quick movement (transition) was detected while the subject was gripping the same object, the movement was most likely due to limb motion, and was classified accordingly. All temporal information attained as frame indices were converted to seconds by dividing the frame indices by the frame rate.

IV. Results and Discussion

A. Data Segmentation Results

The video segmentation algorithm obtains transition times from the video sequence by detecting movements of the tracked hand. For validation, selected video sequences were segmented by a human analyst via visual inspection. It must be noted that manual visual inspection is prone to human judgment errors and it is not necessarily accurate. Since images are sampled at 7.5 fps, the maximum accuracy of the image-based methods is 133ms. The accuracy can be increased by sampling images at a higher frame rate. As seen in Table 1, the two time sources are in close agreement (p>0.05) for detecting the times of grip initiation and termination.

Table 1
Agreement between manual and automatic Method

Subject No.	Grip initiation (frames)	Grip termination (frames)
1	1 ± 3	1 ± 2
2	2 ± 3	2 ± 3
3	1 ± 2	0 ± 3
4	0 ± 1	0 ± 0
5	3 ± 5	6 ± 7
6	1 ± 1	0 ± 1
7	1 ± 2	1 ± 2

Median ± Interquartile range, 1 frame=133ms.

Unwanted hand movements are associated with an increase in MMG amplitude. However, by using the MMG signal without any visual feedback it is almost impossible to determine the cause of the increase in signal amplitude. Because the segmentation method relies on visual recordings, periods of unwanted hand movements were effectively detected. When a brief movement separates two similar grasps, the movement is recognized as an unwanted hand movement. Any unexpected movement splits the grasping sequence into two segments.

B. Grip Recognition Results

The target templates were chosen as images of the hand grasping an object. The recognition rate can be significantly improved if more than one template is associated with each grip type; however, the computational complexity also increases. For this reason, one target template was chosen per grip. The accuracies of the system in differentiating between two target classes are shown in Table 2. The input video for each test was continuous sequences of the hand alternating between each grip and the resting position 140 times. Note that the cylindrical grip was tested without the constraint that the object must be in the hand ROI; this is because the cylindrical object occludes the hand thereby hindering proper identification of the hand ROI.

TABLE 2
PERCENTAGE RECOGNITION RATES: 2-CLASS PROBLEM

Subject No.	Cylinder	Lateral	Palmar	Sphere	Tip	Hook
1	95.0 ± 6	87.7 ± 6	100.0 ± 0	100.0 ± 0	99.3 ± 1	98.7 ± 2
2	100.0 ± 0	100.0 ± 0	93.7 ± 5	98.2 ± 4	95.0 ± 5	81.9 ± 6
3	99.2 ± 2	100.0 ± 0	100.0 ± 0	95.5 ± 8	99.3 ± 1	97.4 ± 5
4	100.0 ± 0	100.0 ± 0	100.0 ± 0	100.0 ± 0	99.5 ± 1	95.6 ± 6
5	98.7 ± 2	100.0 ± 0	99.3 ± 1	96.2 ± 8	100.0 ± 0	97.6 ± 5
6	99.2 ± 2	100.0 ± 0	100.0 ± 0	100.0 ± 0	98.1 ± 4	99.3 ± 1
7	100.0 ± 0	98.0 ± 1	98.9 ± 1	100.0 ± 0	96.3 ± 8	93.1 ± 9

Accuracies for detecting between grip and rest; Mean±Std. deviation

Participants performed 60 continuously executed grips from a random selection of seven types of grip. The accuracy of the system in discriminating among the seven grips is shown in Table 3. As expected, the accuracies of the grip recognizer decreases with increasing number of target classes. The majority of the error is due to misclassifications among the lateral grip, palmar grip and rest; this was because the differences in the shape of the hand among these classes are very small. The cylindrical grip is also often misidentified as an ‘unknown’ grip since the hand was occluded by the object.

Each grip is punctuated with two segments of hand movement: grasping and releasing. The video sequence can be spliced into grip segments by removing hand transition frames. It is important to note that the results of the grip recognizer can be easily verified and corrected by viewing a single frame from each spliced video segment. For this reason, the segmentation system can be used for labeling grip types even if grip recognition accuracies are low.

TABLE 3
PERCENTAGE RECOGNITION RATES: 7-CLASS PROBLEM

Subject No.	Cylinder	Lateral	Palmar	Sphere	Tip	Hook	Rest
1	37.5	54.5	50.0	100.0	62.5	80.0	92.3
2	50.0	54.5	66.6	100.0	100.0	80.0	100.0
3	62.5	63.6	16.6	100.0	100.0	80.0	92.3
4	100.0	42.8	40.0	100.0	55.5	60.0	68.4
5	100.0	100.0	66.6	52.6	83.3	40.0	58.3
6	100.0	60.0	100.0	80.0	47.0	46.6	83.8
7	90.0	60.0	46.1	77.7	75.0	80.0	86.6

V. CONCLUSION

The automatic data segmentation system was used in a constrained environment where there was little change in grasp variability and hand orientation. Average recognition rates of 97.8±4% and 73±20% were obtained for two and seven target classes, respectively. The system detects times of grip transitions and hesitations, such as failing to grasp an object, involuntary grasp transitions and other such mistakes. Future work may include the implementation of feature-based pattern recognition to make the system more adaptable for gesture recognition in unconstrained environments.

ACKNOWLEDGMENT

The authors thank NSERC and Bloorview Kids Rehab.

REFERENCES

[1] C Orizio, “Spectral analysis of muscular sound during isometric contraction of biceps brachii.” Journal of Applied Physiology, 68(2): 508–512, 1990

[2] Anonymous Electromyography : physiology, engineering, and noninvasive applications, [Hoboken, NJ]: IEEE/John Wiley & Sons, 2004, pp.305-319

[3] Silva, J., Heim, W., & Chau, T, “A self-contained mechanomyography-driven externally powered prosthesis”, Archives of Physical Medicine and Rehabilitation, 86(10):2066-2070, 2005

[4] J. Napier. The prehensile movements of the human hand. J Bone and Joint Surgery, 38B(4):902.913, Nov 1956.

[5] J. Silva and T. Chau, "Coupled microphone-accelerometer sensor pair for dynamic noise reduction in MMG signal recording," Electronics Letters, vol. 39, pp. 1496-1498, 2003.

[6] D. Chai and K.N. Ngan, "Face segmentation using skin-color map in videophone applications," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 9, pp. 551-564, 1999.

[7] N. Habili, Cheng Chew Lim and A. Moini, "Segmentation of the face and hands in sign language video sequences using color and motion cues," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 14, pp. 1086-1097, 2004.

[8] S.-. Jung, S.-. Shin, H. Baik and M.-. Park, "Efficient multilevel successive elimination algorithms for block matching motion estimation," Vision, Image and Signal Processing, IEE Proceedings-, vol. 149, pp. 73-84, 2002.

[9] S. Vassiliadis, E.A. Hakkennes, J.S.S.M. Wong and G.G. Pechanek, "The sum-absolute-difference motion estimation accelerator," Euromicro Conference Proceedings, pp. 559-566 vol.2, 1998.

[10] R. Brunelli and T. Poggio, "Face recognition: features versus templates," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 15, pp. 1042-1052, 1993.