Qualification of Arm Gestures using Hidden Markov Models

G. Eliezer Quintana, L. Enrique Sucar and Gildardo Azcárate
Instituto Nacional de Astrofísica, Óptica y Electrónica
Luis Enrique Erro #1, Tonantzintla, Puebla, Mexico
eliezerq@hotmail.com, esucar@inaoep.mx, yayo_tec@hotmail.com

Ron Leder
Universidad Nacional Autónoma de México
Ciudad Universitaria, Mexico City, Mexico
rleder@ieee.org

Abstract

We propose the use of hidden Markov models (HMMs) to qualify arm gestures. A HMM is trained based on the reference or correct *gesture. Then, samples of the gesture that we want to score are used to train a second HMM. Both HMMs are compared, and a measure of their similarity is used to qualify the gesture. We used 3 different metrics to compare HMMs: Levinson, Kullback–Leibler and Porikli. For this, a visual system was developed to track a person's arm, which serves as input to the models that qualify the gestures. We applied this method to qualify the arm movements of stroke patients under rehabilitation. We analyzed three therapeutic gestures: flexion, circular and abduction. A HMM is trained to represent the movement of a healthy person for each gesture, which is compared with the HMMs obtained for each patient. The results are compared with the scales that are used in therapy. From the analysis of several experiments, the Porikli metric was the best to qualify the three gestures, in terms of the motricity index.*

1. Introduction

Recently there has been increasing interest in visual recognition of human gestures for different applications, such as intelligent interfaces and human–robot interaction. Most work focuses on gesture classification, that is, in classifying a visual sequence as one of a predefined set of types of gestures [1, 5]. However, for certain applications, it is required to qualify certain gestures; that is, to compare a human motion against a reference pattern, and to give a measure on how close is the observed gesture to the reference. For instance, in sports or dance training, we would like to know how well did the trainee performed compared to the desired movement. We are particularly interested in rehabilitation after stroke, to evaluate how well did a patient do certain movement, to quantify her progress and provide quantitative feedback.

To qualify arm gestures we propose the use of hidden Markov models (HMMs) [11]. A HMM is trained based on the reference or *correct* gesture. Then, samples of the gesture that we want to qualify are used to train a second HMM. Both HMMs are compared, and a measure of their similarity is used to qualify the gesture. We tried different metrics to compare HMMs, in particular the Levinson, Kullback–Leibler and Porikli metrics.

We evaluated experimentally the gesture qualification system with different patients under rehabilitation after stroke. We consider three typical arm movements used in therapy: flexion, circular and abduction; and trained HMMs for healthy persons, and for several patients at different stages in the rehabilitation process. The results are compared with the scales that are used in therapy, in particular the motricity index and the Fugl–Meyer scale. From the analysis of several experiments, the Porikli metric was the best to qualify the three gestures, in particular in terms of the motricity index.

2. Metrics

A hand gesture can be described by the trajectory of the hand in a sequence of images; so in principle we can compare gestures by comparing trajectories. We first describe some measures to compare directly different trajectories, and analyze why these are not appropriate for gesture qualification. We then present model–based metrics, in particular those based on HMMs, which are the basis for this work.

978-1-4244-2154-1/08/$25.00 ©2008 IEEE

2.1. Trajectory Metrics

In computer vision, a trajectory is a sequence of coordinates of an object in a temporal sequence of images. In 2–D a trajectory can be represented as [9]:

$$T = \{p_n\} = \{(x_1, y_1, t_1), \ldots, (x_N, y_N, t_N)\} \tag{1}$$

where x_i, y_i are the coordinates of the center of mass of the object, t_i is the temporal index, and N is the number of frames or trajectory duration. This definition can be extended to 3–D trajectories by incorporating the z coordinate.

When we have two or more different trajectories, a natural question is: How similar are these trajectories? For comparing trajectories there are metrics based on distances and based on characteristics [9]. The simplest metric to compare two trajectories, T_1 and T_2, is the average distance between the coordinates of each one, defined as [9]:

$$D_m(T_1, T_2) = \frac{1}{N} \sum_{n=1}^{N} d_n^2 \tag{2}$$

where d_n corresponds to the displacement between positions defined in terms of the Cartesian distance:

$$d_n = [(x_n^1 - x_n^2)^2 - (y_n^1 - y_n^2)^2]^{\frac{1}{2}} \tag{3}$$

where x_n^j and y_n^j are the coordinates of object j at time n of trajectory T_j. This metric makes 3 crucial assumptions: (i) the duration of both trajectories is the same: $N^1 = N^2 = N$, (ii) the trajectories are synchronized: $t_n^1 = t_n^2$; that is, the samples from both trajectories are taken at the same times, and (iii) the sampling rates are constant. Given the previous assumptions, the distance metric is not adequate for comparing gestures which in general have different durations and can be sampled at different rates. Some variations of the distance metric have been proposed which consider second order statistics such as the median, variance, maximum and minimum distances [9]; however these have similar limitations.

An alternative to compare trajectories is Dynamic Time Warping (DTW) [14], a technique to measure similarity between two sequences that can vary in time or speed. DTW allows to find an *optimal* alignment between two sequences given certain restrictions, and obtain a similarity measure independent of certain variations in time. However, the main problem of DTW is its computational complexity, since it is of order N^2, where N is the number of observations in each sequence.

2.2. Model–Based Metrics

An alternative to compare trajectories considering variations in duration and sampling are model–based metrics. Instead of comparing directly the motions of two objects, we train a model based on the trajectories and compare the models. Given that HMMs are commonly used for gesture recognition, they are a natural choice for measuring the similarity between gestures.

An HMM [11] represents a dynamic processes with two stochastic elements: a hidden state and a set of observation symbols which depend on the state. In the discrete case is defined by:

1. N discrete states.
2. M observation symbols.
3. A transition matrix $A = \{a_{ij}\}$ which denotes the transitions probabilities from state i to state j.
4. An observation matrix $B = \{b_{jk}\}$ that represents the probability of observing symbol k at state j.
5. An initial probability vector $\pi = \{\pi_i\}$ that denotes the initial probability of state i.

An HMM is represented in a compact way as: $\lambda = A, B, \pi$. The model parameters are obtained from training samples (sequences of observation symbols) using Baum Welch's algorithm [11]. To estimate the probability of a model given an observation sequence, $P(O \mid \lambda)$, the *forward* algorithm is used [11].

Several metrics have been proposed to compare HMMs, next we describe the most common ones.

2.2.1 Levinson

The Levinson metric [8] was the first one proposed to compare discrete HMMs. This measure is based on comparing the observation probability matrices of both models using Euclidian distance:

$$D_e(\lambda_1, \lambda_2) = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \sum_{k=1}^{M} \left\| b_{jk}^1 - b_{jk}^2 \right\|^2} \tag{4}$$

where b_{jk}^i are the observation probabilities for model i. It is assumed that the number of states in both models is the same: $N^1 = N^2$. An alternative measure considers the sum of minimum distances:

$$D_{mec}(\lambda_1, \lambda_2) = \sqrt{\frac{1}{N} \sum_{j=1}^{N} min_j \sum_{k=1}^{M} \left\| b_{jk}^1 - b_{jk}^2 \right\|^2}$$

These metrics do not take into account the temporal structure of the Markov chain, so there are cases in which two HMMs, λ_1 and λ_2, have a Levinson distance close to zero but the probability measures $P(O_1|\lambda_1)$ and $P(O_2|\lambda_2)$ are completely different [6].

2.2.2 Kullback-Leibler

Juang and Rabiner [10] propose an alternative metric to compare two HMMs based on the Kullback-Leibler distance between two probability distributions. This distance measure between two HMMs, λ_1 and λ_2, is defined as:

$$D_k(\lambda_1, \lambda_2) = \frac{1}{T} \left| \log(P[O^1|\lambda_1]) - \log(P[O^1|\lambda_2]) \right| \quad (5)$$

where O^1 is an observation sequence generated by model λ_1 and T is the size of the observation sequence. This measure can be interpreted as how well does the model λ_2 represents the observations generated by model λ_1, with respect to the correspondence of the same observations to the generating model, λ_1 [11].

We notice that this measure is not symmetric. An alternative symmetric measure is the following [10]:

$$D_{sk}(\lambda_1, \lambda_2) = \frac{D_k(\lambda_1, \lambda_2) + D_k(\lambda_2, \lambda_1)}{2} \quad (6)$$

which corresponds to the average of the two non–symmetric measures. This metric has convergence problems. In an experimental evaluation it was obtained that it converges as the size of the observation sequence, T, is incremented [10]. In general, the metric $D_k(\lambda_1, \lambda_2)$ is sensitive to the parameters of the model: the number of states, N, and the number of observation symbols, M.

2.2.3 Porikli

Porikli [9] proposes another metric to compare HMMs in terms of a *cross distance*, defined as follows:

$$\begin{aligned} D_{porikli}(\lambda_1, \lambda_2) = |P(O_1|\lambda_1) + P(O_2|\lambda_2) - \\ P(O_1|\lambda_2) - P(O_2|\lambda_1)| \end{aligned} \quad (7)$$

which evaluates the correspondence of two observation sequences to the two models. O_i is an observation sequence generated by model i, and $P(O_i|\lambda_i)$ is the probability of the observation sequence given the model. The terms $P(O_1|\lambda_1)$ and $P(O_2|\lambda_2)$ indicate the probability of each sequence given their corresponding model; while the cross terms $P(O_1|\lambda_2)$, $P(O_2|\lambda_1)$ measure the probability of the sequences to be generated by the other model. In other words, if the trajectories are the same the direct and cross terms cancel each other and the measure tends to zero; otherwise the distance increases.

3. Methodology

Before comparing the arm gestures, we need to capture and characterize the motion of the arm. For this we develop a stereo visual tracking system [2] that is described next.

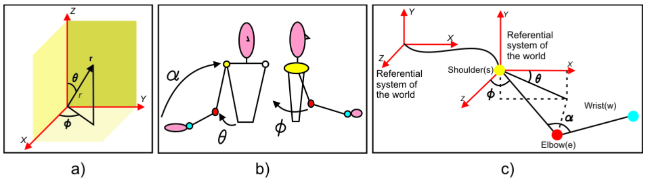

Figure 1. Motion representation: (a) spherical coordinates within a Cartesian system, (b) the 3 angles (α, θ and ϕ) used to represent the motion of the arm, (c) the angles with respect to the reference system.

3.1. Visual System

A color marker is attached to the main joints in the arm of a person: wrist, elbow and shoulder. Each marker is of a different color and all are tracked in the image sequence using the Camshift algorithm [4]. Two cameras positioned in orthogonal directions are used, which are previously calibrated so that the 3D coordinates of each marker can be estimated. For each camera, the centroids (x_i, y_i) of each marker per frame are obtained; and then the 3D coordinates, (X_i, Y_i, Z_i) are estimated using the method proposed by [7]. This consists of projecting the lines of vision of the two cameras to the centroids of the object, and then obtaining their intersection. Once the positions of the 3 joints is obtained, a simplified virtual wire model of the arm is built, with a line from the wrist to the elbow and a second line from the elbow to the shoulder.

To quantify the motion of the arm we use the 3 relative angles of the wire model (see figure 1), as these are invariant with respect to diferents distances of the person to the cameras and the dimensions of the arm. The 3 angles considered are:

- α, the angle between the arm and the forearm (the elbow).
- θ and ϕ, the spherical coordinates of the elbow relative to the shoulder.

The spherical coordinates of the elbow are obtained using the following relations [3]:

$$\rho = \sqrt{X_d^2 + Y_d^2 + Z_d^2} \quad (8)$$

$$\theta = \tan^{-1}\left(\frac{Y_d}{X_d}\right) \quad (9)$$

$$\phi = \cos^{-1}\left(\frac{Z_d}{\rho}\right) \quad (10)$$

where X_d, Y_d, Z_d are the cartesian coordinates of the elbow relative to the shoulder position. The 3 angles are normalized by dividing them by 2π radians. The visual tracking system is illustrated in figure 2.

Figure 2. Visual tracking system. Top: view from each camera, with the 3 markers and wire model in each image. Bottom: graph of the three angles vs. time for a sample gesture.

3.2. Gesture Modeling and Qualification

As mentioned before, we use HMMs to model and compare the gestures. As we are using discrete models, we have to discretize the observations to obtain a finite set of observations sequences. For this we used the *K-means* algorithm, so the 3 normalized angles are transformed to a finite set of symbols (in the experiments we used K = 16, 32, 64, 128 and 256).

To qualify a gesture (in our case a therapeutic gesture) we compare it to a reference or standard gesture. Several repetitions of the reference gesture are captured and used to train a reference HMM, λ_R. Then, the gesture to be evaluated is also captured several times and another HMM is trained, λ_T. Finally the gestures are compared by using the 3 different metrics described before, to obtain a distance between the two HMMs, $D(\lambda_R, \lambda_T)$.

4. Experiments and Results

We present the application and experimental evaluation of our methodology in the qualification of therapeutic gestures for stroke patients. First we briefly introduce the problem, and then we present the experiments and results.

4.1. Rehabilitation after Stroke

Each year in the U.S. over 600,000 people survive a stroke [13], and worldwide stroke affects millions of people. One of the sequela of this disease is hemiplegia. Because the effects of hemiplegia take a long time of therapy to cure and this is very expensive, it is devastating in underdeveloped countries and for people with limited economic resources. Today, new therapeutics options are required to

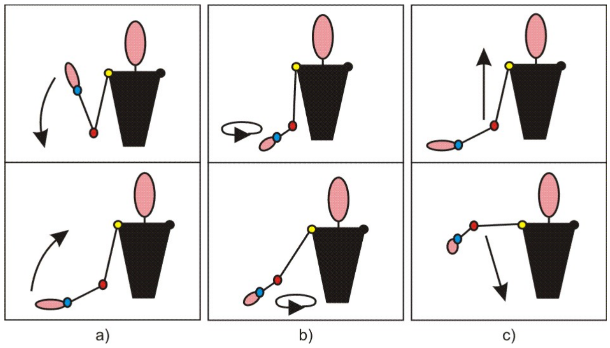

Figure 3. Therapeutic gestures used in the experiments: (a) flexion–extension, (b) circular motion, (c) abduction.

avoid the need of the patient being in the hospital for a long period of time. For this reason, automatic methods are needed that can evaluate the progress of the rehabilitation and provide feedback to the patient without the need of an always present therapist.

Rehabilitation consists of a series of exercises for the upper and lower extremities usually guided by a therapist, designed so that the patient can recover the movements required to perform daily activities. To evaluate the progress of the patient, the therapist uses different clinically objective scales, two common ones are the Fugl–Meyer scale and the motricity index [15].

The Fugl–Meyer scale consists of several aspects that evaluate the motion and pain of the upper and lower extremities, each aspect is assigned a value in the range 0 to 2. We considered the 22 aspects for the upper extremities, so a healthy person will have the full scale of 44. The score assigned to a patient under therapy gives an indication of the degree of advance in the recovery.

The motricity index evaluates the range of motions in the upper extremities: abduction, flexion, and pressing. Each aspect is evaluated by observation in a scale from 0 to 33. A healthy person will be assigned the maximum score of 100.

Recently new therapeutic options are being developed so that the patients can continue the therapy at home, without the need of a therapist present all the time [12]. In this case there is the need of a system that can evaluate the progress of the patient and give her feedback. We have applied our gesture qualification methodology as a way to provide objective, automatic feedback to the patient. For this we consider 3 different arm gestures that are used by therapists in the rehabilitation process: (a) flexion–extension, (b) circular motion, (c) abduction. These are illustrated in figure 3.

4.2. Evaluation Methodology

To evaluate the gesture qualification method, we analyzed the 3 therapeutic gestures for 10 stroke patients who were at different stages in the rehabilitation process. For each gesture, we recorded a reference pattern as performed by the therapist; and then we recorded the patients performing several repetitions of each gesture. Each gesture (patients and reference) was visually analyzed and used to train a HMM. The HMM for each patient and for each of the 3 gestures is compared to the reference HMM using the 3 metrics described earlier: Levinson, Kullback–Leiber and Porikli. Then each metric is compared to the two measures clinical used by the therapists: Fugl–Meyer and motricity index. We also recorded a healthy person performing each of the 3 gestures for comparison. Next we present the results of the experiments.

4.3. Results

The experiments were done at the National Institute for Neurology (INNN) in Mexico City. Ten patients participated in the study, who were guided by a therapist to perform each of the 3 selected gestures, 100 repetitions each, and were recorded using a stereo system. We also recorded the therapist and a healthy person performing each gesture. So in total we have 12 recordings per gesture, and we obtained a HMM for each one. Each HMM was compared to the reference one (therapist) using the 3 metrics. The therapist also graded each patient using the Fugl–Meyer scale and the motricity index.

We compared the 3 metrics with the two therapeutic scales for each gesture. The results are summarized in figure 4 for the flexion gesture and figure 5 for the abduction gesture (the results for the circular gesture are similar and are omitted due to space limitations). For comparison the measures are normalized: 0–100 for the motricity index, 0-44 for the Fugl–Meyer and 0-(value maximum) for the gesture score. In all the graphs the measures are ordered from the higher to the lower grade according to the motricity index.

We can observe that the best agreement is for the Porikli metric and the motricity index. We also observed a good correlation between the Levinson metric and the Fugl–Meyer scale. To quantify these comparisons, we obtained the average Euclidian distance between the metrics and scales, normalized, as follows:

$$d_e^2 = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|m_i - e_i\|^2} \qquad (11)$$

where m_i is the HMM metric for patient i and e_i is the therapeutic measure for patient i. N is the number of samples.

Table 1 shows the average distance for each metric for the 3 gestures, with respect to the motricity index and he Fugl–Meyer scale. The table confirms our observations; the minimum distance is for the Porikli metric with the motricity index, followed by the Levinson for the Fugl–Meyer scale. We also compared the probability of each

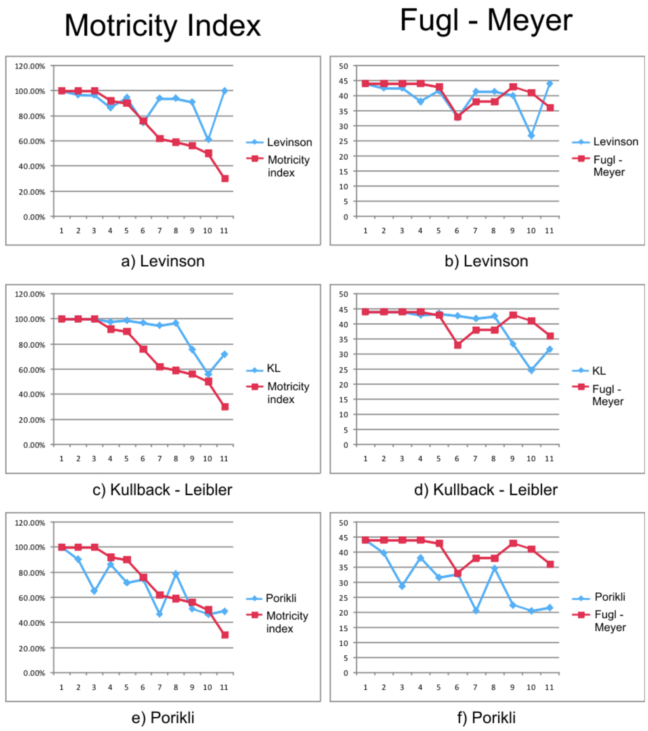

Figure 4. Comparison of the 3 metrics (Levinson, Kullback–Leiber and Porikli) with the 2 therapeutic scales (Motricity index and Fugl–Meyer) for the flexion–extension gesture.

observation sequence given the reference model (therapist), $P(O|\lambda)$. We observe from the table that the evaluation using the standard measure, $P(O \mid \lambda)$, is not as good as the metrics. Thus, it seems that these metrics provide a better alternative to qualify gestures using HMMs.

This results are promising, as for the two clinical scales we obtain a quantitative metric with an average difference of less the 20%. We consider that with these results our gesture qualification method can be the basis for a system to give objective feedback to patients under rehabilitation after stroke.

5. Conclusions

For certain applications, such as in rehabilitation, it is important to qualify certain gestures; that is, to compare them against a reference pattern, and to give a measure on how close is the observed gesture to the reference. In this

Metrics				
	Levinson	K - L	Porikli	$P(0\|\lambda)$
Motricity Index	0.2423	0.2170	0.1762	0.4727
Fugl–Meyer	0.1857	0.3209	0.3465	0.6547

Table 1. Average distance between the HMM metrics and the observations probability vs. the therapeutic scales for the 3 gestures.

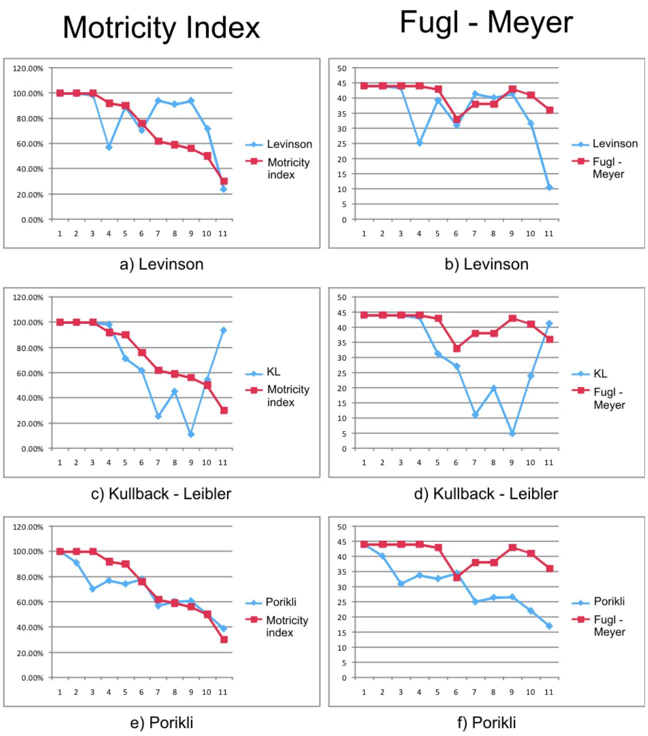

Figure 5. Comparison of the 3 metrics (Levinson, Kullback–Leiber and Porikli) with the 2 therapeutic scales (Motricity index and Fugl–Meyer) for the abduction gesture.

paper we have proposed a method to qualify arm gestures based on HMMs. An HMM is trained based on the reference or *correct* gesture. Then, samples of the gesture that we want to qualify are used to train a second HMM. Both HMMs are compared, and a measure of their similarity is used to qualify the gesture. We used 3 different metrics: Levinson, Kullback–Leibler and Porikli.

We have applied and validated our method for providing objective feedback to patients under rehabilitation after stroke. We compared the 3 metrics against two common scales used by therapist, and we obtained a good correlation, in particular for the Porikli scale with respect to the motricity index. Also, all the metrics have a superior performance than estimating the probability of the observations given the model. These results show that the proposed method is useful, and could be the basis for a system that provides objective feedback to the patients under rehabilitation. We plan to integrate the feedback mechanism to a low-cost system for rehabilitation after stroke, and to test our method in other domains such as in athletic training.

Acknowledgments

This research was supported in part by CONACYT/SALUD–70074 and a grant from the Department of Education NIDRR grant number H133E070013. (These contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.)

References

[1] H. Aviles, L. Sucar, and C. Mendoza. Visual recognition of similar gestures. In *International Conference on Pattern Recognition*, pages 1100–1103. IEEE, 2006.

[2] G. Azcarate. *Seguimiento visual para la rehabilitación*. Master of science thesis, ITESM - Campus Cuernavaca, 2006.

[3] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice-Hall, 1982.

[4] G. Bradski. Computer vision face tracking as a component of a perceptual user interface. *Workshop on Applications of Computer Vision*, 1:214–219, 1998.

[5] J. Deng and H. Tsui. An hmm-based approach for gesture segmentation and recognition. *IEEE, Trans. On PAMI*, 19:679–682, 2000.

[6] M. Falkhausen, H. Reininger, and Wolf. Calculation of distance measures between hidden markov models. Technical report, Institut fr Angewandte Physik, Johann Wolfgang Goethe Universitt, FRG, 1995.

[7] P. B. Gonzalez and J. M. C. nas. Seguimiento tridimensional usando dos cámaras. Technical report, Universidad Rey Juan Carlos, España, Octubre 2004. (In Spanish).

[8] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi. An introduction to the application of the theory of probabilistic function of a markov process to automatic speech recognition. *B. S. T. J.*, 62:1036–1074, 1983.

[9] F. Porikli. Trajectory distance metric using hidden markov model based representation. Technical report, Mitsubishi Electronic Research Laboratories, 2005.

[10] B. H. J. A. R. Rabiner. A probabilistic distance measure for hidden markov models. *AT&T Technical Journal*, 64:391–408, 1985.

[11] L. R. Rabiner. A tutorial on hidden markov models an selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.

[12] Reinkensmeyer, Pang, Nessler, and Painter. Java therapy: Web-based robotic rehabilitation. 2002.

[13] D. J. Reinkensmeyer, L. E. Kahn, M. Averbuch, A. McKenna-Cole, B. D. Schmit, and W. Z. Rymer. Understanding and treating arm movement impairment after chronic brain injury: progress with the arm guide. *J. Rehabil. Re. Develop.*, 37(6):653–662, 2000.

[14] S. Salvador and P. Chan. Fastdtw: Toward accurate dynamic time warping in linear time and space. In *KDD Workshop on Mining Temporal and Sequential Data*, pages 70–80, 2004.

[15] Sanford, Moreland, Swanson, Stratford, and Gowland. Reliability of the fugl - meyer assessment of testing motor performance in patients following stroke. *Physical therapy*, 73:447 – 454, 1993.