# An In-situ Trainable Gesture Classifier

**Anouk van Diepen**[1] A.V.DIEPEN@STUDENT.TUE.NL
**Marco Cox**[1] M.G.H.COX@TUE.NL
**Bert de Vries**[1,2] BDEVRIES@GNRESOUND.COM
[1]Department of Electrical Engineering, Eindhoven University of Technology, [2]GN Hearing BV, Eindhoven

**Keywords**: gesture recognition, probabilistic modeling, Bayesian inference, empirical Bayes.

## 1. Introduction

Gesture recognition, i.e., the recognition of pre-defined gestures by arm or hand movements, enables a natural extension of the way we currently interact with devices (Horsley, 2016). Commercially available gesture recognition systems are usually pre-trained: the developers specify a set of gestures, and the user is provided with an algorithm that can recognize just these gestures.

To improve the user experience, it is often desirable to allow users to define their own gestures. In that case, the user needs to train the recognition system herself by a set of example gestures. Crucially, this scenario requires learning gestures from just a few training examples in order to avoid overburdening the user.

We present a new in-situ trainable gesture classifier based on a hierarchical probabilistic modeling approach. Casting both learning and recognition as probabilistic inference tasks yields a principled way to design and evaluate algorithm candidates. Moreover, the Bayesian approach facilitates learning of prior knowledge about gestures, which leads to fewer needed examples for training new gestures.

## 2. Probabilistic modeling approach

Under the probabilistic modeling approach, both learning and recognition are problems of probabilistic inference in the same generative model. This generative model is a joint probability distribution that specifies the relations among all (hidden and observed) variables in the model.

Let $y = (y_1, ..., y_T)$ be a time series of measurements corresponding to a single gesture with underlying characteristics $\theta$. The characteristics are unique to gestures of type (class) $k$. We can capture these dependencies by the probability distribution

$$p(y, \theta, k) = \underbrace{p(y|\theta)}_{\substack{\text{dynamical}\\ \text{model}}} \cdot \underbrace{p(\theta|k)}_{\substack{\text{gesture}\\ \text{characteristics}}} \cdot \underbrace{p(k)}_{\substack{\text{gesture}\\ \text{class index}}} . \tag{1}$$

Because the measurement sequence is temporally correlated, it is natural to choose $p(y|\theta)$ to be a hidden Markov model (HMM). HMMs have been successfully applied to gesture classification in the past (Mäntylä et al., 2000). Under this model, $\theta$ represents the set of parameters of the HMM.

During learning, the parameter values $\theta$ of gestures of class $k$ need to be learned from data. We choose to learn this distribution using a two-step approach.

In the first step, a prior for $\theta$ is constructed. This prior distribution can be obtained in various ways. We have chosen to construct one that captures the common characteristics that are shared among *all* gestures. This is done by learning the distribution using dataset $\mathcal{D}$, consisting of one measurement from each gesture class. This can be expressed as

$$p(\theta|\mathcal{D}, k) = \frac{p(\mathcal{D}, \theta, k)}{\int p(\mathcal{D}, \theta, k)\, \mathrm{d}\theta} . \tag{2}$$

In the second step, the parameter distribution is learned for a specific gesture class, using the previously learned $p(\theta|\mathcal{D}, k)$ and a set of measurements $\mathcal{D}_k$ with the same class $k$:

$$p(\theta|\mathcal{D}, \mathcal{D}_k, k) = \frac{p(\mathcal{D}_k|\theta)p(\theta|\mathcal{D}, k)p(k)}{\int p(\mathcal{D}_k, \theta, k|\mathcal{D})\, \mathrm{d}\theta} . \tag{3}$$

In practice, exact evaluation of Eq. 2 and Eq. 3 is intractable for our model due to the integral in the denominator. We use variational Bayesian inference to approximate this distribution (MacKay, 1997), which results in a set of update equations that need to be iterated until convergence.

During recognition, the task of the algorithm is to identify the gesture class with the highest probability of having generated the measurement $y$. This is expressed by

$$p(k|y) = \frac{\int p(y,\theta,k)\,\mathrm{d}\theta}{\sum_k \int p(y,\theta,k)\,\mathrm{d}\theta}\,. \qquad (4)$$

If we assume that each gesture is performed with the same *a priori* probability $p(k)$, then $p(y|k) \propto p(k|y)$. To calculate $p(y|k)$, the method as proposed in Chapter 3 of Beal (2003) is used: the obtained variational posterior distribution of the parameters is replaced by its mean, which allows exact evaluation of $p(y|k)$.

## 3. Experimental validation

We built a gesture database using a Myo sensor bracelet (ThalmicLabs, 2016), which is worn just below the elbow (see Fig. 1). The Myo's inertial measurement unit measures the orientation of the bracelet. This orientation signal is sampled at 6.7 Hz, converted into the direction of the arm, and quantized using 6 quantization directions. The database contains 17 different gesture classes, each performed 20 times by the same user. The duration of the measurements was fixed to 3 seconds.

*Figure 1.* The Myo sensor bracelet used to measure gestures.

As a measure of performance, we use the recognition rate defined as:

$$\text{Recognition rate} = \frac{\#\text{ correctly classified}}{\text{total }\#\text{ of samples}}\,. \qquad (5)$$

The gesture database is split in a training set containing 5 samples of every gesture class, and a test set containing the remaining (15x17=) 255 samples. The recognition rate is evaluated on models trained on 1 through 5 examples. To minimize the influence of the training order, the results are averaged over 5 different permutations of the training set.

To compare our algorithm, we have also evaluated the recognition rate of the same algorithm with uninformative prior distributions and of a 1-Nearest Neighbor (1-NN) algorithm using the same protocol.


*Figure 2.* Recognition rates of the 1-NN algorithm, the proposed algorithm without prior information (HMM), and the proposed algorithm with informed prior distributions (HMM prior).

Figure 2 shows the recognition rates of the algorithms. Both hidden Markov based algorithms have a higher recognition rate than the 1-NN algorithm. For personalization of gesture recognition, we are especially interested in learning gesture classes using a low number of training examples. In particular for one-shot training (from one example only), the hidden Markov model using the learned prior distribution corresponds to the highest recognition rate.

The algorithm was also tested for gestures that are not used to learn the prior distribution. When the prior is constructed with similar gestures, the new gestures are also learned faster than when uninformative priors are used.

There are multiple ways to incorporate these results in a practical gesture recognition system. For example, the prior distribution can be constructed by the developers of the algorithm. Another possibility is to allow users to provide prior distributions themselves. This means that the system will take longer to set up, but when a user wants to learn a specific gesture under in-situ conditions, it will require less training examples.

## References

Beal, M. J. (2003). *Variational algorithms for approximate Bayesian inference.* Doctoral dissertation, University College London.

Horsley, D. (2016). Wave hello to the next interface. *IEEE Spectrum*, *53*, 46–51.

MacKay, D. J. C. (1997). *Ensemble Learning for Hidden Markov Models* (Technical Report).

Mäntylä, V.-M., Mäntyjärvi, J., Seppänen, T., & Tuulari, E. (2000). Hand gesture recognition of a mobile device user. *2000 IEEE International Conference on Multimedia and Expo.* (pp. 281–284).

ThalmicLabs (2016). Myo Gesture Control Armband. Retrieved from https://www.myo.com/.