Improving the Recognition of Eating Gestures Using Intergesture Sequential Dependencies

Raul I. Ramos-Garcia, Eric R. Muth, John N. Gowdy, and Adam W. Hoover

***Abstract*—This paper considers the problem of recognizing eating gestures by tracking wrist motion. Eating gestures are activities commonly undertaken during the consumption of a meal, such as sipping a drink of liquid or using utensils to cut food. Each of these gestures causes a pattern of wrist motion that can be tracked to automatically identify the activity. Previous works have studied this problem at the level of a single gesture. In this paper, we demonstrate that individual gestures have sequential dependence. To study this, three types of classifiers were built: 1) a K-nearest neighbor classifier which uses no sequential context, 2) a hidden Markov model (HMM) which captures the sequential context of subgesture motions, and 3) HMMs that model intergesture sequential dependencies. We built first-order to sixth-order HMMs to evaluate the usefulness of increasing amounts of sequential dependence to aid recognition. On a dataset of 25 meals, we found that the baseline accuracies for the KNN and the subgesture HMM classifiers were 75.8% and 84.3%, respectively. Using HMMs that model intergesture sequential dependencies, we were able to increase accuracy to up to 96.5%. These results demonstrate that sequential dependencies exist between eating gestures and that they can be exploited to improve recognition accuracy.**

***Index Terms*—Activity recognition, gesture recognition, hidden Markov models (HMM), mHealth.**

I. Introduction

THIS paper considers the problem of recognizing eating gestures by tracking wrist motion. Eating gestures are activities commonly undertaken during the consumption of a meal, such as taking a bite of food, sipping a drink of liquid, or using utensils to cut food. Each of these gestures causes a pattern of wrist motion that can be tracked to automatically identify the activity. Previous works have studied this problem at the level of a single gesture [1]–[3]. In this paper, we demonstrate that individual gestures have sequential dependence. For example, the action of using utensils to cut food is typically followed by the action of taking a bite of food, which is typically followed by the action of resting the wrist during mastication of the food. We have developed hidden Markov models (HMMs) that model these sequential dependencies, and demonstrate that they improve recognition accuracy compared to simpler classifiers.

Manuscript received January 29, 2014; revised April 11, 2014; accepted May 29, 2014. Date of publication June 5, 2014; date of current version May 7, 2015. This work was supported by the NIH under Grant 1R41DK091141-A1.

R. I. Ramos-Garcia, J. N. Gowdy, and A. W. Hoover are with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634 USA (e-mail: gramos@clemson.edu; jgowdy@clemson.edu; ahoover@clemson.edu).

E. R. Muth is with the Department of Psychology, Clemson University, Clemson, SC 29634 USA (e-mail: muth@clemson.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2014.2329137

This paper is motivated by recent advances in body sensing and mobile health technology that have created new opportunities for empowering people to take a more active role in managing their health [4]. Our group has been researching methods to build a watch-like tool that tracks wrist motion to help automatically monitor dietary intake [5]–[8]. Obesity now afflicts one in three adults and one in six children in the United States [9], [10] and has been recognized as a major health problem in need of new tools [11]–[13]. Self-monitoring of dietary intake, body weight, and physical activity have been consistently found to be associated with successful weight loss and maintenance [14]. However, food diaries and other tools currently used for monitoring dietary intake require users to manually estimate and record energy intake, making them prone to error and difficult to use for long periods of time [15]. Body-worn sensors offer the opportunity to automatically track dietary intake [16]–[18]. In our previous work, we demonstrated methods for detecting periods of eating (i.e., meals, snacks) by tracking wrist motion continuously all day [6], [8], and for detecting and counting bites taken during a meal [5], [7]. In the current paper, we consider the problem of recognizing multiple types of eating gestures during eating and specifically the use of sequential context to improve recognition accuracy.

Activity recognition using body-worn sensors generally takes the approach of gesture spotting. In this approach, each gesture type is modeled by an HMM with states representing subsequences of the motion of the gesture. This approach has been used to recognize hand gestures for sign language [19], [20], hand gestures for interacting with robots [21], [22], and the use of different types of hand tools [23]. The approach has also been used to recognize typical daily activities like walking, running and climbing stairs [22], [24], to detect when the elderly fall down [25], [26], and to detect tremors due to Parkinson's disease [27]. For the problem of monitoring dietary intake, the approach has been demonstrated on gestures related to eating foods [2], [3] and drinking liquids [1]. However, these works assume that individual gestures (or activities) are independent of the preceding or succeeding gestures. Thus, while the HMMs capture the sequential dependence of the subcomponents of a gesture, for example, back-and-forth motion during hand waving, they do not model what might be likely to happen after the gesture, for example, returning to the natural rest position after hand waving. The typical method to implement gesture spotting is to start and end all individual gesture HMMs at a common null state [28], [29]. However, this misses the opportunity to model sequential dependencies between gestures. In this paper, we build HMMs that extend the dependence modeling from subgesture components to gesture-to-gesture sequencing. We

TABLE I
FOOD LIST PER SUBJECT

Sub.	Foods
1	Stir fry vegetables, water
2	Chicken sandwich, french fries, tea
3	Refried beans, shoestring french fries, popcorn chicken, taco, tea
4	Popcorn chicken, shoestring french fries, banana, water
5	Stew beef, seasoned dry limas, steamed California blend vegetables, white rice, water
6	African spiced sweet potato, blackened tilapia, seasoned corn, sauteed tomatoes, and zucchini, sweet tea
7	Bread, salad, wild rice, sprite zero
8	Ice cream, cupcake, custom fruit bowl, water
9	Eggplant and broccoli pizza, hamburger, pepperoni pizza, shoestring french fries, sweet tea
10	Black beans and rice, sauteed pollock, kiwi juice
11	Garlic bread sticks, grilled Italian sausage with onions and peppers, rotini with marinara, water
12	Cupcake, veggie Indian currry, brownie, lemonade
13	Eggplant parmesan, salad bar, sweetzza cinnamon pecan, sweetzza apple, unsweet tea
14	Portobello sandwich, salad, sweet tea
15	Pasta tour of Italy, powerade
16	Buffalo tenders, salad bar, cranberry juice water mix
17	BBQ brisket and kaiser roll, cereal corn pops, hash sweet potato and bacon, apple juice and water
18	Peanut butter chocolate fudge, salad bar, spinach and cheese quiche, water
19	Homestyle chicken sandwich, salad bar, sweet tea
20	Fish, mac and cheese, sweet tea
21	Pasta tour of italy, salad bar, bread, blueberry crobbler, brownie, sweet tea
22	Pepperoni pizza, spice pork and vegetable, sweetzza chocolate peanut butter, diet coke
23	Bread sticks, desert pizza, ziti, mellow yellow
24	Asian vegetables, salad, wasabi potatoes, coke
25	Shoestring french fries, homestyle chicken sandwich, salad bar, diet coke

accomplish this through the use of higher order HMMs that can be reduced to first order for processing. Our experiments demonstrate this approach on the recognition of eating gestures, but the same approach could be applied to any type of activity recognition where gesture-to-gesture levels of sequential dependence exist.

II. METHODS

A. Data

For this study, 25 participants were recorded eating a meal in the Harcombe Dining Hall at Clemson University [30]. Participants were free to choose any available foods and beverages so that eating gestures would be uncontrolled. Table I lists the foods and beverages chosen. Fig. 1 shows our instrumented table that was capable of recording data from up to four participants simultaneously [31]. Four digital video cameras in the ceiling (approximately 5 m height) were used to record each participant's mouth, torso, and tray during meal consumption. A custom wrist-worn device containing MEMS accelerometers (AccX, AccY, AccZ) and gyroscopes (Yaw, Pitch, Roll) [32], [33] was used to record the wrist motion of each participant at 15 Hz. Cameras and wrist motion trackers were wired to the same computers and used timestamps for synchronization. All the data were smoothed using a Gaussian-weighted window of width 1 s and standard deviation of $\frac{2}{3}$ s.

Fig. 1. Table instrumented for data collection. Each participant wore a custom tethered device to track wrist motion.

For purposes of recognition, we defined four eating gestures: *rest*, *utensiling*, *bite*, and *drink* [34]. All other actions, whether they involve eating or noneating activities, (e.g., gesturing while talking, cleaning with a napkin, waving at a friend, etc.) are referred to as *other*. We hand-labeled the meal recordings using the custom program shown in Fig. 2. Labels were determined by manually observing the intent of the eater in the synchronized video. Unlabeled segments of duration less than 4 s were considered transition gaps between gestures and were ignored. Table II lists the total number, average time, and cumulative duration of gestures in the dataset. An interrater reliability study was conducted to determine the stability of the gesture definitions and hand labelings, finding 90.7% total time agreement [34].

B. Classifiers Overview

Before introducing our new classifier, we describe two simpler classifiers used to establish baseline accuracy. This provides a reference for evaluating the improvement obtained by our method. Fig. 3 illustrates the windows of time over which features are calculated for each classifier. The first classifier is a K-nearest neighbor (KNN) and calculates features across an entire gesture. The second classifier is an HMM and calculates features across subgesture periods of time. This is the classic gesture spotting approach. It is assumed that the second classifier will perform better than the first due to its use of sequential context within a gesture. The third classifier is also an HMM but it calculates features across an entire gesture, modeling intergesture sequential dependencies. Using the first two classifiers, each gesture is recognized independently, while the third classifier uses the context of one or more previous gestures to improve recognition accuracy. The following sections provide the details for each classifier.

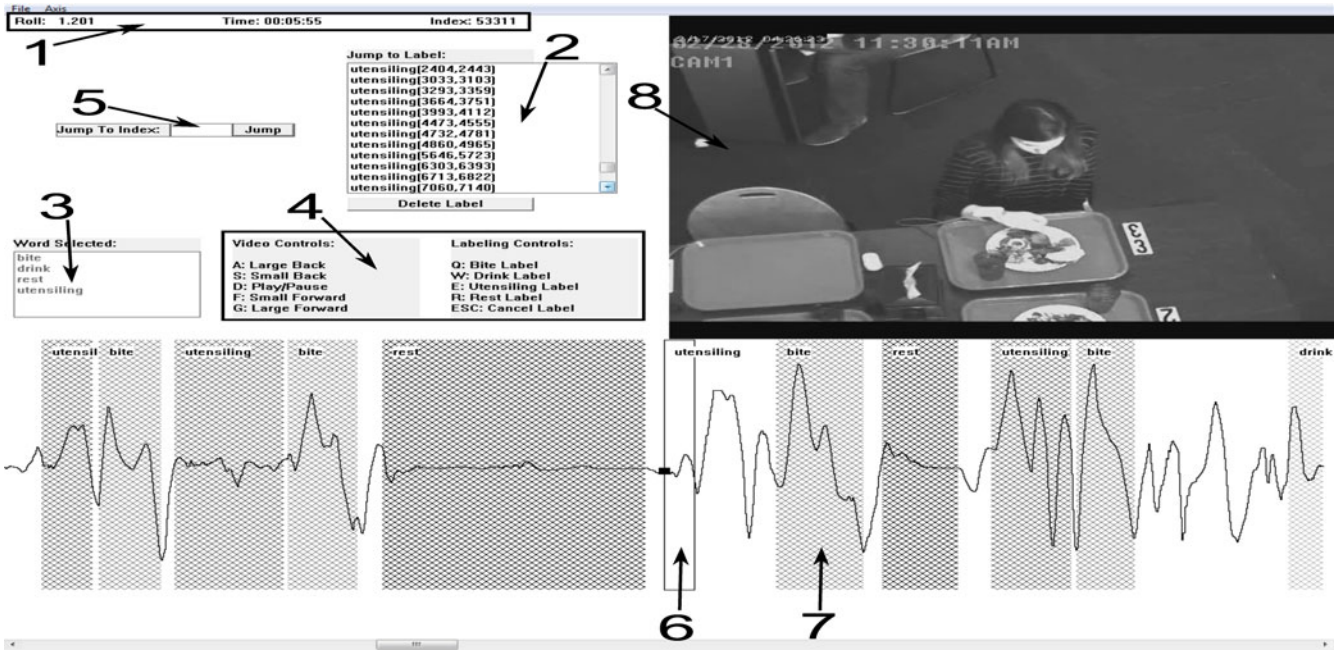

Fig. 2. Custom program was created for manual labeling. Tool elements: (1) indicates the current accelerometer or gyroscope signal, the current recording time, and the current sample/frame index; (2) list and navigation of labeled gestures; (3) indicator of the current label; (4) instructions for the tool's usage; (5) navigation to specific sample/frame index; (6) shows an example of creating a segment using current label; (7) filled boxes show completed segments; (8) video is synchronized with the wrist motion data displayed at (6).

TABLE II
OCCURRENCES, AVERAGE TIME PER GESTURE, AND CUMULATIVE TIME OF GESTURES

Gesture	Occurrences	Avg. Time (s)	Total Time (min)
Rest	582	8.3	80.6
Utensiling	700	7.3	84.7
Bite	1039	3.0	52.3
Drink	155	7.2	18.6
Other	310	8.6	44.5

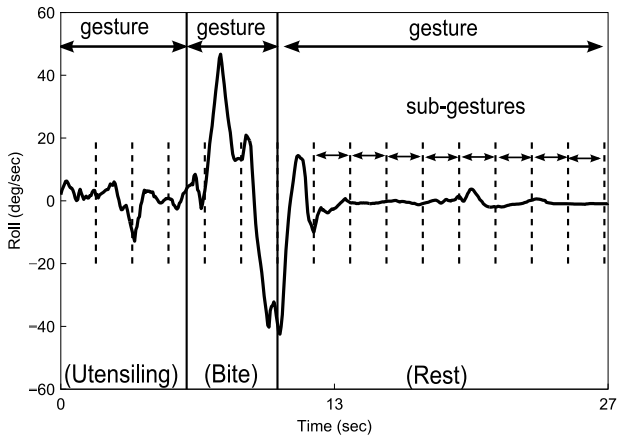

Fig. 3. Periods of time used to calculate features for the classifiers.

C. K-Nearest Neighbor

For our first baseline, we used a KNN classifier [35]. Each vector in a training dataset is associated with a class label. The process of classification places an unlabeled sample x in the feature space among the training data. The classification searches for the K closest labeled samples to x. The number of occurrences of each label are calculated among the K closest samples. The label with the largest number of occurrences is assigned to x. Features are normalized by computing the Z-norm. Euclidean distance was used for feature comparison.

We calculated the following features across the duration of a gesture: the total motion in each of the six axes, the 15 correlations between all pairs of axes, and the ratio of rotational motion to linear motion (called manipulation in [8]). This feature set was reduced to those most useful for classification using a feature forward selection method [36]. This is an iterative process that begins by selecting the single feature that by itself provides the highest classification accuracy. Subsequently, all the remaining features are paired with the first selected feature to find the pair that provides the highest classification accuracy. This process continues until adding another feature results in a negligible increase in classification accuracy. This feature selection process was performed for $K = 1, 3, \ldots, 19$, using 18 meals for training and seven meals for testing. This process yielded $K = 7$ and the 10 features shown in Table III.

D. Subgesture HMMs

For our second baseline, we used HMMs to model the temporal sequencing of the subcomponents of each gesture. States represent gesture fragments (see Fig. 3). Five HMMs were built, one for each of the four defined eating gestures and one for the other category. The HMMs were designed with a left-to-right

TABLE III
KNN FEATURES

No.	KNN Features
1	Manipulation
2	Amount of motion of AccX
3	Amount of motion of AccY
4	Amount of motion of AccZ
5	Amount of motion of Pitch
6	Correlation(AccX, AccZ)
7	Correlation(AccY, AccZ)
8	Correlation(Yaw, Pitch)
9	Correlation(Yaw, Roll)
10	Correlation(AccX,Yaw)

architecture with skip using an HMM toolbox for MATLAB[1]. For observables, a set of features were computed using a sliding window of 0.5 s with 50% overlap. The features included the mean, standard deviation, and slope for each of the six axes. Emission probabilities were calculated by modeling features as continuous observations using Gaussian mixture models (GMMs), where the model order M is the number of Gaussians. We calculated GMMs using the expectation maximization algorithm [37]. The GMM for subgesture γ is given by (1), where c_{γ_m}, μ_{γ_m}, and Σ_{γ_m} represent the mixture weight, mean, and covariance matrix of the mth Gaussian, respectively. In this paper, only diagonal covariance matrices were used.

$$G_\gamma = \{c_{\gamma_m}, \mu_{\gamma_m}, \Sigma_{\gamma_m}\}, \text{where } m = 1, \ldots, M. \quad (1)$$

For model parameter selection, we tested several numbers of states (3 to 15) and Gaussians (1 to 10) using a five-fold cross validation, selecting 13 states and five Gaussians as the best parameters for each gesture. For recognition, the Viterbi algorithm was used to calculate the log probability of each of the five subgesture HMMs given the observed feature sequence. The HMM with the highest log probability determined the classification of the gesture.

E. Gesture-to-Gesture HMMs

The purpose of this classifier is to capture sequential context between gestures; in other words, to use the history of one or more preceding gestures to improve the recognition of the current gesture. States represent whole gestures. Fig. 4 shows an HMM that uses a history of one gesture, which we denote as HMM1. Each state corresponds to one of the five gestures (rest, utensiling, bite, drink, and other). This HMM is ergodic, meaning every state can be reached by any other state.

To incorporate additional history, we first build a conditional higher-order HMM and then convert it to an equivalent first-order HMM [38]. We denote HMMn as an nth-order HMM that uses the previous n gestures for sequential context. For example, HMM2 has the capability to recognize state q_t based on the previous states q_{t-1} and q_{t-2}. Fig. 5 shows an example second-order HMM for only two of our gestures (for clarity), bite (B), and utensiling (U). The temporal sequencing is left-to-right. For example, a_{UBU} indicates a state transition where the previous two states were U then B, and the next state is U. Fig. 6 shows the equivalent first-order HMM. Each state models a sequence of two gestures. Logically, transitions between some states are impossible. For example, the state BU cannot transition to BB because the former's most recently recognized gesture is U, which does not match the memory of the latter. As another example, Fig. 7 shows the same part of HMM3 after it has been reduced to first-order. In general, any HMM of order n (HMMn) can be converted into a first-order equivalent.

For observables, we used the five log probabilities obtained from the subgesture HMMs described in Section II-D. Emission probabilities were calculated by modeling these observables using GMMs as in (1). We selected the number of Gaussians

[1]http://www.cs.ubc.ca/ murphyk/Software/HMM/hmm.html

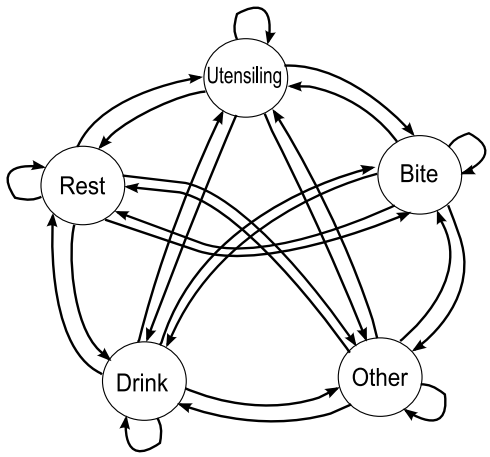

Fig. 4. HMM1 uses a history of 1 gesture to help recognize the current gesture.

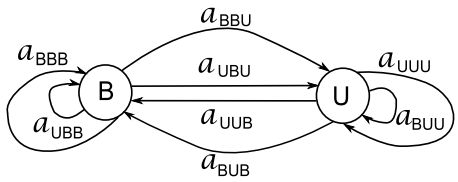

Fig. 5. Example second-order HMM of bite (B) and utensiling (U). A transition probability a is conditional and has three subscripts indicating the gesture history at times $t-2$, $t-1$, and t.

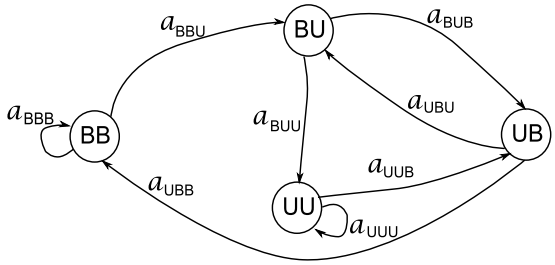

Fig. 6. Equivalent first-order HMM of Fig. 5. The state notation (e.g. BB) shows the memory (left-to-right) of the most recently seen gestures. A transition probability a has three subscripts indicating the gesture history at times $t-2$, $t-1$, and t.

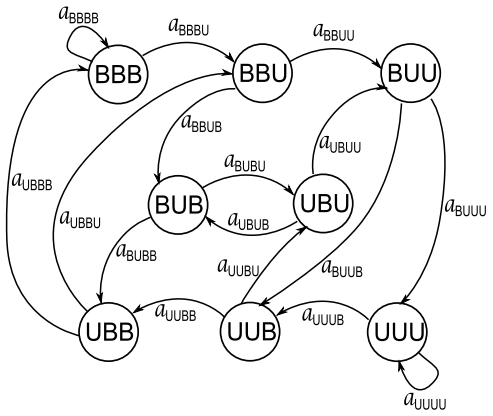

Fig. 7. Equivalent first-order HMM from a third-order HMM for bite (B) and utensiling (U). The state and transition notation reads left-to-right and defines the memory of the most recently seen gestures.

by performing a five-fold cross validation for $M = 1, \ldots, 20$ in HMM1, choosing $M = 7$.

The transition probabilities $a_{\alpha\beta\ldots\phi\omega}$ are equivalent to $P(\{q_t = \omega\}|\{q_{t-1} = \phi, \ldots, q_{t-n-1} = \beta, q_{t-n} = \alpha\})$, where $\alpha, \beta, \ldots, \phi, \omega \in \{\text{rest, utensiling, bite, drink, other}\}$ and $\alpha\beta\ldots\phi\omega$ represent the states visited (from left to right) at times $t-n$, $t-n-1$, ..., $t-1$, and t, respectively. Thus, for a given set of data, these can be calculated as

$$a_{\alpha\beta\ldots\phi\omega} = \frac{\text{Total \# of transitions from gesture sequence } \alpha\beta\ldots\phi \text{ to gesture } \omega}{\text{Total \# of } \alpha\beta\ldots\phi \text{ gesture sequences}}. \quad (2)$$

Similarly, the prior probabilities can be calculated from a given set of data as

$$\pi_{\alpha\beta\ldots\phi} = \frac{\text{Total \# of } \alpha\beta\ldots\phi \text{ gesture sequences}}{\text{Total \# of n-gesture sequences}}. \quad (3)$$

The Viterbi decoding algorithm outputs the most probable n-gesture $\alpha\beta\ldots.\phi$ state sequence Q, but only ϕ from each state is retained in the output. The other portions of each state are the running memory and are redundant.

A drawback of this approach is that the number of states and transition probability matrix grow exponentially as the order grows. In our case, as we increase the model order, the total number of states are 5^n and the transition probability matrix contains 5^{2n} elements. Although not all transitions are logically possible due to the constraint of the history of gestures, there are no fully-empty columns or rows in the transition matrix. Due to software limitations, we were only able to test this approach for our data up to $n = 6$.

III. Results

The transition probabilities found for HMM1 are shown in Table IV. The amount of sequential dependence between gestures can be seen in entries with values larger or smaller than 0.2. For example, the likelihood of transitioning from utensiling to bite is 81.1%. The tables for HMM2 to HMM6 are too large to display easily.

TABLE IV
Transition Probabilities

From\To	Rest	Utensiling	Bite	Drink	Other
Rest	0.072	0.338	0.364	0.093	0.133
Utensiling	0.134	0.007	0.811	0.007	0.040
Bite	0.309	0.364	0.141	0.022	0.165
Drink	0.253	0.226	0.137	0.151	0.233
Other	0.285	0.256	0.292	0.167	0.000

TABLE V
Total Recognition Accuracy

Method	Accuracy (%)
KNN	75.8
Subgesture HMM	84.3
HMM1	87.7
HMM2	88.0
HMM3	89.6
HMM4	92.2
HMM5	94.6
HMM6	96.5

TABLE VI
Recognition Accuracy for Each Gesture

Method	Rest (%)	Utensiling (%)	Bite (%)	Drink (%)	Other (%)
KNN	88.8	76.8	84.3	71.5	31.7
Subgesture HMM	91.7	83.2	86.9	86.5	56.0
HMM1	93.5	87.5	93.0	75.1	63.1
HMM2	93.8	87.9	92.8	75.5	64.6
HMM3	94.3	89.5	92.8	77.5	69.0
HMM4	96.8	92.6	95.1	79.2	75.5
HMM5	97.7	94.4	96.2	82.7	82.8
HMM6	99.3	97.3	98.5	82.0	88.4

All classifiers were trained and tested using leave-one-out cross validation. Training the subgesture and gesture-to-gesture HMMs consisted of calculating GMM values (1). Training the KNN consisted of populating the feature neighbor space with the training data for purposes of calculating nearest neighbors for the test data.

For all classifiers, accuracy was measured as the total percentage of time in all the meals that gestures were labeled correctly. Table V presents a summary of the accuracy achieved by each of the classifiers. For the baseline classifiers, the subgesture HMMs performs better than the KNN with an 8.5% improvement. Our gesture-to-gesture classifier that incorporates gesture history shows further improvement, from 3.4% to 12.2% across HMM1 to HMM6. For example, HMM3 uses a history of three gestures and achieved 89.6% accuracy, while HMM6 uses a history of six gestures and achieved 96.5% accuracy.

Table VI shows the results broken down by gesture type. It can be seen that gesture history improves the recognition accuracy of every gesture type except drinks. Table VII shows the

TABLE VII
HMM6 CONFUSION MATRIX

Actual\Classifier	Rest (%)	Utensiling (%)	Bite (%)	Drink (%)	Other (%)
Rest	99.3	0.3	0.1	0.0	0.3
Utensiling	1.2	97.3	0.3	0.0	1.3
Bite	0.5	0.2	98.5	0.0	0.8
Drink	8.5	1.9	3.0	82.0	4.6
Other	2.0	7.5	0.9	1.2	88.4

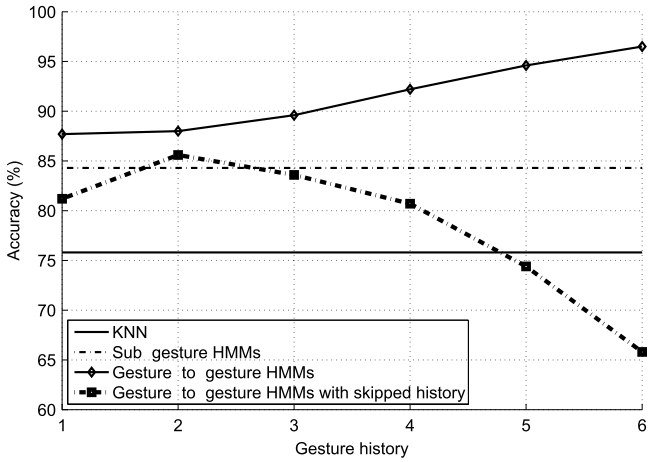

Fig. 8. Classifier accuracy with increasing gesture history and increased skip in history.

confusion matrix for HMM6 for each gesture. The most confusion occurs between drink and rest gestures, and utensiling and other gestures. This is likely due to the similarities of motions between these gesture types. Drink gestures can have pauses that can be confused with brief periods of rest, and both utensiling and other gesture types encompass a wide variety of motions.

We conducted an additional test of our classifier to determine if the most recent gesture history is the most important. This was done also to provide assurance that our results are not an effect of over-fitting. We constructed an HMMn-skip for each value of n, where the gesture history captured in HMMn skipped over the most recent n gestures and instead used the history of the n gestures preceding them. For example, HMM1-skip models history by skipping over the most recent gesture and instead uses the identity of the one preceding it; HMM2-skip models the history by skipping over the two most recent gestures and instead uses the sequence of the two preceding them; etc. Formally, (2) was modified to skip n gestures between $\alpha\beta\ldots\phi$ and ω. Fig. 8 shows the accuracies of all the classifiers as the amount of history and skipped history is increased. The KNN and subgesture HMMs have constant accuracy because they do not incorporate gesture-level history. Our gesture-to-gesture HMM increases in accuracy as the history of gestures is increased. The HMMs with skip perform much worse. This indicates that the most recent gesture history is in fact the most relevant.

IV. DISCUSSION

This paper demonstrates that the recognition of eating gestures can be improved through knowledge of the sequential dependence of individual gestures. We believe this is due to the patterned nature of activities during eating. For example, a common pattern is to use utensils to prepare a bite of food, consume the bite of food, and then rest hands while masticating and swallowing. Another common pattern is to intersperse drinks with food bites. Our HMMs capture this "language of eating" and use it to improve recognition accuracy. The results in Fig. 8 show evidence supporting this conclusion. A skipped history of two gestures shows improvement over a skipped history of one gesture, but additional skipped history shows a continually decreasing accuracy. This is likely due to HMM2-skip capturing some of the cyclical phrasing of gesture patterns, such as utensil-bite-utensil-bite, where the skip of two matches the repetition in the phrasing. We assume that other HMMn-skip classifiers capture less phrasing and hence show increasingly worse results. A phrase-level study of the sequential context during eating is a topic for future work.

Methods for activity recognition that are developed in a controlled setting will potentially be brittle in a natural setting. This problem affects all studies where some instrumentation is necessary to record behavior. In our case, our data was collected in a cafeteria which is as natural a setting as possible where wrist motion trackers and discreetly positioned video cameras could be used to record eating. Participants selected their own foods from everything available in the cafeteria and consumed them however they wished, with no instructions on how or what to eat. This is arguably more natural than asking participants to conduct a sequence of scripted gestures in a lab or asking participants to eat a small set of controlled foods (as in [3]). Although our dataset was acquired from 25 people each eating a meal, the total gestures recorded numbered 2786 (see Table II). We believe this captures sufficient variety in pace and style of each gesture type that the improvement we found in recognition is not brittle.

Other limitations of this study include the number of activities modeled and the use of manually segmented data. We found that 15.9% of the total time during meals is comprised of other activities, such as wiping hands on a napkin or gesturing while talking. Increasing the number of activities may affect overall classifier accuracy. However, we still expect that modeling sequential dependencies would produce an improvement in accuracy. For our experiments, we used manually segmented data in order to determine the impact of sequential dependence modeling on classifier accuracy independent of possible segmentation errors. In the future, we plan to explore automated segmentation methods. It may be that sequential dependencies can be exploited to improve automated segmentation as well as classification. Finally, it should be noted that other classifiers besides HMMs can model sequential context such as conditional random fields and dynamic Bayesian networks. The comparison of these classifiers on the problem of eating gesture recognition is a topic of future work.

The method presented in this paper could be used to improve the accuracy of automated methods that track wrist motion to monitor the number of drinks [1] or bites [6] taken by a person. Our method could also potentially be used to analyze eating habits of individuals that may correlate with variations in energy intake. For example, slowing the pace of eating has been found to be associated with decreased intake during a meal [39]. It may be that other activity patterns have similar associations. In this case, a tool that automatically measures these patterns could prove useful in diagnosis and behavior treatment. These topics are subjects for future work.

REFERENCES

[1] O. Amft, D. Bannach, G. Pirkl, M. Kreil, and P. Lukowicz, "Towards wearable sensing-based assessment of fluid intake," in *Proc. 8th IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Apr. 2010, pp. 298–303.

[2] O. Amft and G. Tröster, "Recognition of dietary activity events using on-body sensors," *Artif. Intell. Med.*, vol. 42, no. 2, pp. 121–136, 2008.

[3] H. Junker, O. Amft, P. Lukowicz, and G. Tröster, "Gesture spotting with body-worn inertial sensors to detect user activities," *Pattern Recog.*, vol. 41, no. 6, pp. 2010–2024, 2008.

[4] S. Kumar, W. Nilsen, M. Pavel, and M. Srivastava,, "Mobile health: Revolutionizing healthcare through transdisciplinary research," *Computer*, vol. 46, no. 1, pp. 28–35, 2013.

[5] Y. Dong, A. Hoover, and E. Muth, "A device for detecting and counting bites of food taken by a person during eating," in *Proc. IEEE Int. Conf. Bioinformat. Biomed.*, 2009, pp. 265–268.

[6] Y. Dong, A. Hoover, E. Muth, and J. Scisco, "A new method for measuring meal intake in humans via automated wrist motion tracking," *Appl. Psychophysiol. Biofeedback*, vol. 37, no. 3, pp. 205–215, 2012.

[7] Y. Dong, A. Hoover, J. Scisco, and E. Muth, "Detecting eating using a wrist mounted device during normal daily activities," in *Proc. 9th Int. Conf. Embedded Syst. Appl.*, 2011, pp. 3–9.

[8] Y. Dong, J. Scisco, M. Wilson, E. Muth, and A. Hoover, "Detecting periods of eating during free-living by tracking wrist motion," *IEEE J. Biomed. Health Informat.*, in press.

[9] K. Flegal, M. Carroll, B. Kit, and C. Ogden, "Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999–2010," *J. Amer. Med. Assoc.*, vol. 307, no. 5, pp. 491–497, 2012.

[10] C. L. Ogden, M. D. Carroll, B. K. Kit, and K. M. Flegal, *Prevalence of obesity in the United States, 2009–2010*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2012.

[11] A. Ershow, A. Ortega, J. Baldwin, and J. Hill, "Engineering approaches to energy balance and obesity: Opportunities for novel collaborations and research: Report of a joint National Science Foundation and National Institutes of Health workshop," *J. Diabetes Sci. Technol.*, vol. 1, no. 1, pp. 96–105, 2007.

[12] B. McCabe-Sellers, "Advancing the art and science of dietary assessment through technology," *J. Amer. Dietetic Assoc.*, vol. 110, no. 1, pp. 52–54, 2010.

[13] F. E. Thompson, A. F. Subar, C. M. Loria, J. L. Reedy, and T. Baranowski, "Need for technological innovation in dietary assessment," *J. Amer. Dietetic Assoc.*, vol. 110, no. 1, pp. 48–51, 2010.

[14] L. E. Burke, J. Wang, and M. A. Sevick, "Self-monitoring in weight loss: A systematic review of the literature," *J. Amer. Dietetic Assoc.*, vol. 111, no. 1, pp. 92–102, 2011.

[15] F. Thompson, and A. Subar, *Dietary Assessment Methodology*. 2nd ed. New York, NY, USA: Academic Press/Elsevier, 2008.

[16] O. Amft, and G. Tröster, "On-body sensing solutions for automatic dietary monitoring," *IEEE Pervasive Comput.*, vol. 8, no. 2, pp. 62–70, Apr./Jun. 2009.

[17] S. Päßler, M. Wolff, and W.-J. Fischer, "Food intake monitoring: an acoustical approach to automated food intake activity detection and classification of consumed food," *Physiol. Meas.*, vol. 33, no. 6, p. 1073, 2012.

[18] E. Sazonov, and S. Schuckers, "The energetics of obesity: A review: Monitoring energy intake and energy expenditure in humans," *IEEE Eng. Med. Biol. Mag*, vol. 29, no. 1, pp. 31–35, Jan./Feb. 2010.

[19] Y. Li, X. Chen, X. Zhang, K. Wang, and Z. Wang, "A sign-component-based framework for chinese sign language recognition using accelerometer and sEMG data," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2695–2704, Oct. 2012.

[20] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Trans. Syst. Man Cybern., Part A: Syst. Humans*, vol. 41, no. 6, pp. 1064–1076, Nov. 2011.

[21] S. Kim, G. Park, S. Yim, S. Choi, and S. Choi, "Gesture-recognizing hand-held interface with vibrotactile feedback for 3d interaction," *IEEE Trans. Consum. Electron.*, vol. 55, no. 3, pp. 1169–1177, Aug. 2009.

[22] C. Zhu, and W. Sheng, "Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living," *IEEE Trans. Syst. Man Cybern. Part A: Syst. Humans*, vol. 41, no. 3, pp. 569–573, May 2011.

[23] J. Ward, P. Lukowicz, G. Troster, and T. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers," *IEEE Trans. Pattern Anal. Mach Intell.*, vol. 28, no. 10, pp. 1553–1567, Oct. 2006.

[24] D. Trabelsi, S. Mohammed, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "An unsupervised approach for automatic activity recognition based on hidden markov model regression," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 829–835, Jul. 2013.

[25] J. Cheng, X. Chen, and M. Shen, "A framework for daily activity monitoring and fall detection based on surface electromyography and accelerometer signals," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 1, pp. 38–45, Jul. 2013.

[26] L. Tong, Q. Song, Y. Ge, and M. Liu, "Hmm-based human fall detection and prediction method using tri-axial accelerometer," *IEEE Sensors J.*, vol. 13, no. 5, pp. 1849–1856, May 2013.

[27] G. Rigas, A. Tzallas, M. Tsipouras, P. Bougia, E. Tripoliti, D. Baga, D. Fotiadis, S. Tsouli, and S. Konitsiotis, "Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 478–487, May 2012.

[28] E. Guenterberg, H. Ghasemzadeh, and R. Jafari, "Automatic segmentation and recognition in body sensor networks using a hidden markov model," *ACM Trans. Embedded Comput. Syst.*, vol. 11, no. S2, pp. 46:1–46:19, 2012.

[29] D. Jurafsky, J. H. Martin, A. Kehler, K. Vander Linden, and N. Ward, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Cambridge, MA, USA: MIT Press, vol. 2, 2000.

[30] Z. Huang, "An assessment of the accuracy of an automated bite counting method in a cafeteria setting," Master's Thesis, Dept. Electr. Comput. Eng., Clemson University, Clemson, SC, USA, 2013.

[31] J. Salley, "Accuracy of a bite-count based calorie estimate compared to human estimates with and without calorie information available," Master's Thesis, Dept. Psychol., Clemson University, Clemson, SC, USA, 2013.

[32] STMicroelectronics. Lis344alh: Mems inertial sensor. (2013). [Online]. Available: http://www.st.com/stonline/products/literature/ds/14337.pdf

[33] STMicroelectronics. Lpr410al: Mems motion sensor. (2013). [Online]. Available: http://www.st.com/st-web-ui/static/active/en/resource/technical/document/datasheet/CD00254123.pdf

[34] R. I. Ramos-Garcia and A. W. Hoover, "A study of temporal action sequencing during consumption of a meal," in *Proc. ACM Int. Conf. Bioinformat. Comput. Biol. Biomed. Informat.*, 2013, pp. 68–75.

[35] L. Rabiner, S. Levinson, A. Rosenberg, and J. Wilpon, "Speaker-independent recognition of isolated words using clustering techniques," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 27, no. 4, pp. 336–349, Jun. 1981.

[36] D. W. Aha and R. L. Bankert, "A comparative evaluation of sequential feature selection algorithms," in *Learning from Data: Artificial Intelligence and Statistics V (ser. Lecture Notes in Statistics)*., New York, NY, USA: Springer-Verlag, 1996, ch. 4, pp. 199–206.

[37] D. Reynolds, and R. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 1, pp. 72–83, Jan. 1995.

[38] J.-F. Mari, J.-P. Haton, and A. Kriouile, "Automatic word recognition based on second-order hidden Markov models," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 1, pp. 22–25, Jan. 1997.

[39] J. Scisco, E. Muth, Y. Dong, and A. Hoover, "Slowing bite-rate reduces caloric consumption; an application of the bite counter device," *J. Amer. Dietetic Assoc.*, vol. 111, pp. 1231–1235, Aug. 2011.

Authors' photographs and biographies not available at the time of publication.