AN APPROACH TO HUMAN ACTIVITY CLUSTERING USING INERTIAL MEASUREMENT DATA

Milan Gnjatović, PhD[1]
Vojkan Nikolić, PhD
Dušan Joksimović, PhD
University of Criminal Investigation and Police Studies, Belgrade, Serbia

Nemanja Maček, PhD
School of Electrical and Computer Engineering of Applied Studies, Belgrade, Serbia
Faculty of Computer Science, Megatrend University, Belgrade, Serbia

Nebojša Budimirović
Preschool Teacher Training College, Šabac, Serbia

Abstract: This paper reports on a pilot study of an approach to human activity clustering using inertial measurement data. At the signal level, we particularly consider the angular velocity and instantaneous acceleration data obtained from a three-axis inertial measurement unit placed on the right arm of the human subject. At the methodology level, the approach consists of three components: symbol-based modeling of spatiotemporal signals, *an adaptation of the* Levenshtein distance, and a graph-based clustering algorithm. A prototype system implementing the proposed approach is evaluated on recordings of six human subjects involved in four task-oriented activities with significant intercluster similarity. The clustering results are assessed using the Rand index (RI=0.921), the precision rate (P=1), the recall rate (R=0.636), and the balanced F-measure (F=0.778).

Keywords: human activity clustering, inertial measurement data, adapted Levenshtein distance, graph-based clustering.

INTRODUCTION

Automatic human activity recognition is regarded as an important and challenging task in security, military and police applications. For an overview of research in the field, the reader may consult Jo-

1 milan.gnjatovic@kpu.edu.rs

banputra, Bavishi, and Doshi (2019), Hussain, Sheng and Zhang (2019), and Avci, Bosch, Marin-Perianu, Marin-Perianu and Havinga (2010). One line of research in the field concerns the processing of external sensor data (e.g. videos) which is computationally expensive and of limited pervasiveness. The second research line concerns the processing of wearable sensor data (e.g. inertial measurement sensor data) which is less computationally demanding and more appropriate for tactical scenarios. This contribution follows the latter research line.

This paper introduces a pilot study on human activity clustering based on inertial data. At the methodology level, the proposed approach integrates three methodological aspects: (i) symbol-based modeling of spatiotemporal signals, (ii) *an adaptation of the* Levenshtein distance, and (iii) a graph-based clustering algorithm. The first two aspects build upon the method for on-line signature authentication introduced by Schimke, Vielhauer and Dittmann (2004). The third aspect is inspired by the method for digital image segmentation introduced by Felzenszwalb and Huttenlocher (2004).

In addition, the approach is introduced with sufficient detail to allow for a computational implementation. A prototype system implementing the proposed approach is evaluated on recordings of human subjects involved in four task-oriented activities with significant intercluster similarity, selected from the Carnegie Mellon University Multimodal Activity (CMU-MMAC) Database (De la Torre et al. 2009).

THE APPROACH

Symbol-based Modeling of Spatiotemporal Signals

Let a be an activity captured by a set of n parameters $\{p_1, p_2, \ldots, p_n\}$ sampled at $k \geq 2$ equidistant time points, i.e.:

$$p = \begin{pmatrix} p[1,1] & p[1,2] & \cdots & p[1,n] \\ p[2,1] & p[2,2] & \cdots & p[2,n] \\ \cdots & \cdots & \cdots & \cdots \\ p[k,1] & p[k,2] & \cdots & p[k,n] \end{pmatrix} \quad (1)$$

where $p[i,j]$ represents the value of parameter p_j at time point t_i. In addition, let S be a set containing $2n+1$ symbols:

$$S = \left\{ s_1^{\max}, s_1^{\min}, s_2^{\max}, s_2^{\min}, \ldots, s_n^{\max}, s_n^{\min}, \varepsilon \right\} \quad (2)$$

where ε is the empty symbol. Matrix p is first mapped into an intermediate sequence of strings over set S, calculated as (cf. Schimke et al., 2004):

$$\begin{pmatrix} f(p[1,1]) \oplus f(p[1,2]) \oplus \ldots \oplus f(p[1,n]) \\ f(p[2,1]) \oplus f(p[2,2]) \oplus \ldots \oplus f(p[2,n]) \\ \cdots \\ f(p[k,1]) \oplus f(p[k,2]) \oplus \ldots \oplus f(p[k,n]) \end{pmatrix} = (S_1, S_2, \ldots, S_k)^T \quad (3)$$

where $\oplus$ stands for symbol concatenation, and:

$$f(p[i,j]) = \begin{cases} s_j^{\max}, & \text{if } e_{\max}(p,i,j) = \text{T}, \\ s_j^{\min}, & \text{if } e_{\min}(p,i,j) = \text{T}, \\ \varepsilon, & \text{otherwise}, \end{cases} \tag{4}$$

$$e_{\max}(p,i,j) = \begin{cases} p[i,j] > p[i+1,j] & \text{if } i = 1, \\ p[i,j] > p[i-1,j] & \text{if } i = n, \\ (p[i-1,j] < p[i,j]) \wedge (p[i,j] > p[i+1,j]) & \text{if } 1 < i < n, \end{cases} \tag{5}$$

$$e_{\min}(p,i,j) = \begin{cases} p[i,j] < p[i+1,j] & \text{if } i = 1, \\ p[i,j] < p[i-1,j] & \text{if } i = n, \\ (p[i-1,j] > p[i,j]) \wedge (p[i,j] < p[i+1,j]) & \text{if } 1 < i < n. \end{cases} \tag{6}$$

It is easy to show that the following holds:

$$(\forall p,i,j) \neg (e_{\max}(p,i,j) \wedge e_{\min}(p,i,j)) \tag{7}$$

The final sequence of strings that represents activity a is obtained by omitting all empty strings from sequence (3):

$$a = (S_1, S_2, \ldots, S_{\hat{k}})^T, \tag{8}$$

where $\hat{k} \leq k$.

Adapted Levenshtein Distance

The widely acknowledged Levenshtein distance (Jurafsky and Martin, 2009; Levenshtein, 1966) between two strings S_a and S_b is defined as the minimum number of single-character editing operations (deletion, insertion and substitution) needed to transform one string into the other, and can be calculated as:

$$\lambda(S_a, S_b) = D(|S_a|, |S_b|) \tag{9}$$

where:

$$\begin{aligned} &(\forall\, 0 \leq i \leq |S_a|)\ D(i,0) = i, \\ &(\forall\, 0 \leq j \leq |S_b|)\ D(0,j) = j, \end{aligned} \tag{10}$$

and

$$(\forall\, 1 \leq i \leq |S_a|, 1 \leq j \leq |S_b|)\ D(i,j) = \min \left\{ \begin{array}{l} D(i-1,j)+1, \\ D(i,j-1)+1, \\ D(i-1,j-1) + \begin{cases} 1, & \text{if } S_a(i-1) \neq S_b(j-1) \\ 0, & \text{if } S_a(i-1) = S_b(j-1) \end{cases} \end{array} \right\} \tag{11}$$

The adapted Levenshtein distance used in this approach quantifies the similarity between two sequences of strings, i.e. two activities (cf. Eq. (8)). We build upon the work of Schimke et al. (2004) and consider the *weighted string editing* operations:

The cost of insertion or deletion of a string is equal to its length.

The cost of substitution of one string by another is equal to the standard Levenshtein distance (i.e. λ, cf. Eq. (9)-(11)) between them.

The adapted Levenshtein distance between two activities, i.e. two sequences of strings, a_1 and a_2 is defined as:

$$\hat{\lambda}(a_1, a_2) = \frac{\hat{D}(|a_1|, |a_2|)}{|a_1| + |a_2|}, \tag{12}$$

where:

$$\begin{aligned} &\hat{D}(0,0) = 0, \\ &(\forall\, 1 \le i \le |a_1|)\ \hat{D}(i,0) = \hat{D}(i-1,0) + |a_1(i-1)|, \\ &(\forall\, 1 \le j \le |a_2|)\ \hat{D}(0,j) = \hat{D}(0,j-1) + |a_2(j-1)|, \end{aligned} \tag{13}$$

and

$$(\forall\, 1 \le i \le |a_1|, 1 \le j < |a_2|)\ \hat{D}(i,j) = \min\begin{Bmatrix} \hat{D}(i-1,j) + |a_1(i-1)|, \\ \hat{D}(i,j-1) + |a_2(j-1)|, \\ \hat{D}(i-1,j-1) + \lambda(a_1(i-1), a_2(j-1)) \end{Bmatrix}. \tag{14}$$

Graph-based Clustering

Let $A = (a_1, a_2, \ldots a_m)$ be a sequence of m activities that need to be clustered. The graph-based clustering algorithm used in this approach is inspired by the image segmentation algorithm introduced by Felzenszwalb and Huttenlocher (2004), and can be described through the following steps.

In the starting clustering, each activity a_i is assigned to its own cluster $c(a_i)$, i.e.:

$$(\forall\, 1 \le i \le m)\ c(a_i) = i\,. \tag{15}$$

Let $\wp(A) = (a_{i_1}, a_{j_1})\ (a_{i_2}, a_{j_2}) \ldots, (a_{i_w}, a_{j_w})$, where $w = \frac{m(m-1)}{2}$ be a sequence of all unordered pairs of activities in A ordered by non-decreasing adapted Levenshtein distance between the activities contained in each pair.

The ordered sequence $\wp(A)$ is iterated through as follows. For a given pair (a_i, a_j), if activities a_i and a_j are assigned to different clusters and the adapted Levenshtein distance between them is *less than a given threshold*[2] *value* τ, *the corresponding clusters are merged. This step can be written in pseudocode as:*

2 The threshold value is an input real-number parameter whose selection will be discussed in other paper.

$$
\begin{aligned}
&(for\ 1 \le q \le w)\\
&\quad f\ \ (\hat{\lambda}(a_{i_q}, a_{j_q}) < \tau) \wedge (c(a_{i_q}) \neq c(a_{j_q})))\\
&\qquad (for\ 1 \le r \le m)\\
&\qquad\quad f\ \ (c(a_r) = c(a_{j_q}))\ \ c(a_r) \leftarrow c(a_{i_q})
\end{aligned}
\tag{16}
$$

When the loop given above is finished, the final clustering is determined by values $c(a_i)$ for $1 \le i \le m$.

EVALUATION

Database

To evaluate the introduced approach in a realistic setting, we resort to the Carnegie Mellon University Multimodal Activity (CMU-MMAC) Database (De la Torre et al. 2009). It contains recordings of human subjects cooking different recipes in a kitchen environment. One of the modalities contained in this database is captured with three-axis inertial measurement units (MicroStrain's 3DM-GX1) containing an accelerometer, gyroscope, and magnetometer which allow for measuring absolute orientation, angular velocity and instantaneous acceleration. The signals are gyro-stabilized and recorded at the rate of 60 Hz.

We selected six human subjects (S50, S51, S52, S53, S54, S55) each of which was recorded while cooking four different recipes (brownies, scrambled eggs, pizza and sandwich). One of these twenty-four subject-activity pairs (i.e. subject S52 preparing brownies) contained errors and was excluded. All the selected subjects are right-handed, so we decided to consider the inertial measurement signals obtained from the inertial measurement unit placed on their right arms. At the signal level, we particularly consider the instantaneous acceleration and angular velocity in the three axes (i.e. six parameters in total, cf. the first six columns of Table 1). As an illustration, the parameter values sampled in five successive time points representing a very small fragment of subject-activity pair (S50, Brownie) are given in Table 1. The list of subject-activity pairs is given in the first column of Table 2. Each subject-activity pair is considered as an activity captured by a set of six parameters, cf. Eq. (1).

Table 1 *Illustration of the input inertial measurement data.*

Acceleration			Angular velocity			Count	System time
a_x	a_y	a_z	Roll	Pitch	Yaw		
0.406751	0.820338	-0.414228	0.098211	0.003138	0.135864	2824	16:39:56:000
0.415082	0.833796	-0.417432	0.102291	0	0.146533	2825	16:39:56:008
0.419355	0.842555	-0.415937	0.107938	0.023533	0.149984	2826	16:39:56:016
0.421491	0.846828	-0.406537	0.121431	0.030436	0.132727	2827	16:39:56:024
0.4232	0.843196	-0.396924	0.139002	0.034515	0.144964	2828	16:39:56:032

PREPROCESSING

The selected data are preprocessed in two respects. First, parts of the input data that are not relevant for the activity performed by the subjects were excluded from the consideration. The start and end system times of the activity-relevant segments for each subject-activity pair are given in Table 2. And second, for the purpose of efficiency, the data are further automatically down-sampled to 1 Hz.

Table 2 *Subject-activity pairs and activity-relevant segments.*

Subject, activity	Start Count	Start system time	End Count	End system time
S50, Brownie	2824	16:39:56:000	51077	16:46:22:000
S50, Eggs	2953	15:42:40:000	36453	15:47:08:000
S50, Pizza	3312	15:24:22:000	61074	15:32:04:000
S50, Sandwich	3059	16:27:46:000	24934	16:30:41:000
S51, Brownie	3113	10:39:47:000	45112	10:45:23:000
S51, Eggs	6874	10:01:16:000	34500	10:04:57:000
S51, Pizza	4151	09:32:28:005	64901	09:40:34:005
S51, Sandwich	5056	10:25:04:003	21804	10:27:18:003
S52, Eggs	2627	15:01:01:000	28752	15:04:30:000
S52, Pizza	5168	14:49:46:000	41543	14:54:37:000
S52, Sandwich	2590	15:18:54:006	17090	15:20:50:006
S53, Brownie	2487	10:31:06:003	39738	10:36:04:003
S53, Eggs	2318	10:05:07:007	28568	10:08:37:007
S53, Pizza	1785	09:42:13:005	58917	09:49:50:005
S53, Sandwich	2472	10:25:11:000	18347	10:27:18:000
S54, Brownie	3785	11:56:08:000	52785	12:02:40:000
S54, Eggs	1715	11:29:03:003	31340	11:33:00:003
S54, Pizza	3064	11:15:27:000	62064	11:23:19:000
S54, Sandwich	2921	11:47:33:000	24672	11:50:27:000
S55, Brownie	5205	13:20:55:000	45081	13:26:14:000
S55, Eggs	2934	12:47:49:000	32309	12:51:44:000
S55, Pizza	3544	12:33:04:000	72297	12:42:14:000
S55, Sandwich	3407	13:11:43:006	23657	13:14:25:006

RESULTS

The proposed approach is implemented in a prototype system, and the clustering results are given in Table 3. When discussing these results, one should keep in mind that the number of expected clusters was not provided a priori. Twenty-three subject-activity pairs are grouped in eight clusters, which may appear as significantly more than the four ground truth clusters. However, eighteen of twenty-three pairs were correctly clustered in four clusters, and there were no false positives. To trade-off the clustering quality against the number of identified cluster, we consider the Rand index (RI=0.921), the precision rate (P=1), the recall rate (R=0.636), and the balanced F-measure (F=0.778), as summarized in Table 4. While this is a pilot study, the results are encouraging.

The results presented in the left part of Table 4 are derived directly from Table 3. We make a short remark. Let a_i and a_j be two distinct subject-activity pairs arbitrarily selected from the given set of 23 subject-activity pairs. The number of such two-element combinations is equal to 253 (i.e. $\binom{23}{2}$, cf. the sum of numbers in the second column of Table 4), and for each of them the prototype system evaluates whether or not a_i and a_j belong to the same cluster.

Table 3 *Clustering results.*

Subject, activity	Clusters							
	C1	C2	C3	C4	C5	C6	C7	C8
S50, Brownie	x							
S51, Brownie	x							
S53, Brownie		x						
S54, Brownie	x							
S55, Brownie	x							
S50, Eggs			x					
S51, Eggs			x					
S52, Eggs			x					
S53, Eggs			x					
S54, Eggs			x					
S55, Eggs			x					
S50, Pizza				x				
S51, Pizza				x				
S52, Pizza					x			
S53, Pizza				x				
S54, Pizza				x				
S55, Pizza				x				
S50, Sandwich							x	
S51, Sandwich								x
S52, Sandwich						x		
S53, Sandwich						x		
S54, Sandwich							x	
S55, Sandwich							x	

Table 4 Evaluation results.

True positives	TP=35	The Rand index	RI=0.921
True negatives	TN=198	Precision	P=1
False positives	FP=0	Recall	R=0.636
False negatives	FN=20	Balanced F-measure	F=0.778

DISCUSSION AND CONCLUSION

In addition to the evaluation results reported above, it is important to note that the symbol-based modeling of spatiotemporal signals and the *adapted* Levenshtein distance considered in this paper provide a basis for addressing some of the desiderata in the field:

1. Generalizability: The proposed approach is general to the extent that it includes no explicit expectations or previous knowledge on the activities to the clustered. The only assumption on the input sensor data is that they can be represented by an arbitrary number of synchronized sequences of numeric values.

2. Discrimination capacity: Approaches to automatic human activity recognition are typically intended to differentiate between human behaviors of relatively small similarity, including fundamental behaviors (e.g. walking, sitting, standing, running, etc.), body postures (e.g. arm downwards, arm upwards, etc.) and task-oriented behaviors (e.g. showering, dressing, leaving house, etc.), cf. Jobanputra et al. (2019). In contrast to them, the proposed approach provides a basis for comparing task-oriented human behaviors of significant intercluster similarity, without prior activity segmentation.

3. Real-time performance: Almost all approaches reported in the literature are based on two-stage process: gathering of sensory information, and off-line data processing. Thus, on-line human activity recognition is identified as one of the important open research questions in the field (cf. Avci et al., 2010). The approach introduced in this paper allows for real-time representation and comparison of human activities.

On the other hand, the further challenges of the proposed approach are related to research questions of: (i) more detailed and domain-targeted testing, (ii) exploring new features derived from the original features (e.g. velocity and relative position can be derived from the acceleration data, etc.), and (iii) selecting an appropriate threshold value. These questions will be taken up in future work.

ACKNOWLEDGMENTS

The presented study was partially funded within the framework of the ERA.Net RUS Plus program (research grant ID 99). The Carnegie Mellon University Multimodal Activity (CMU-MMAC). Database used in this paper was obtained from kitchen.cs.cmu.edu, and the data collection was funded in part by the National Science Foundation under Grant No. EEEC-0540865.

REFERENCES

1. Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R., Havinga, P. (*2010*) Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey, *23rd International Conference on Architecture of Computing Systems 2010*, Hannover, Germany, pp. 1-10.
2. De la Torre, F., Hodgins, J., Montano, J., Valcarcel, S., Forcada, R., Macey, J. (2009) *Guide to the Carnegie Mellon University Multimodal Activity (CMU-MMAC) Database*, Tech. report CMU-RI-TR-08-22, Robotics Institute, Carnegie Mellon University.
3. Felzenszwalb, P.F., Huttenlocher, D.P. (2004) Efficient Graph-Based Image Segmentation, *International Journal of Computer Vision*, 59, pp. 167–181.
4. Hussain, Z., Sheng, M., Zhang, W.E. (2019) Different Approaches for Human Activity Recognition: A Survey, arXiv, 1906.05074, Downloaded July 16th 2020, https://arxiv.org/abs/1906.05074.
5. Jurafsky, D., Martin, J.H. (2009) *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*, 2nd edition, Prentice-Hall.

6. Jobanputra, C., Bavishi, J., Doshi, N. (2019) Human Activity Recognition: A Survey, *Procedia Computer Science*, 15, pp. 698-703.

7. Levenshtein, V.I. (1966) Binary codes capable of correcting deletions, insertions, and reversals, *Cybernetics and Control Theory*, 10(8), pp. 707-710 (Original in Doklady Akademii Nauk SSSR 163(4): 845-848, 1965).

8. Schimke, S., Vielhauer, C., Dittmann, J. (2004) Using adapted Levenshtein distance for on-line signature authentication, *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*, Cambridge, 2004, pp. 931-934, Vol.2.