Quantitative Analyses on Effects from Constraints in Air-Writing

Songbin XU^{†a)}, Nonmember, Yang XUE[†], Member, and Yuqing CHEN[†], Nonmember

SUMMARY Very few existing works about inertial sensor based airwriting focused on writing constraints' effects on recognition performance. We proposed a LSTM-based system and made several quantitative analyses under different constraints settings against CHMM, DTW-AP and CNN. The proposed system shows its advantages in accuracy, real-time performance and flexibility.

key words: inertial sensor, air-writing recognition, quantitative analyses, LSTM-RNN

1. Introduction

The history of writing is exactly the history of mankind [1]. Writing has been existed since the naissance of characters, and has been widely extended as an essential interaction modality so far. Nowadays, handwriting on a touchscreen provides a complimentary text input modality for general human computer interaction (HCI). This type of HCI approach is so nature and simple that people does not have to spend much time learning and memorizing. However, with the rapid development of wearable and ubiquitous computing systems, the size of touchscreen is getting smaller, such as the smart watches, on which handwriting becomes difficult and un-convenient. Addressing this issue, the airwriting is proposed, which sets users free from writing with traditional input devices, especially the touchscreens and the keyboards. The term "air-writing" refers to writing meaningful isolated characters freely in the air, and in this paper we focus on the inertial sensor based air-writing, which is executed with a hand-held or wearable device with inertial sensor built-in. This kind of air-writing has attracted people's attention for three key reasons. Firstly, the inertial sensors are becoming more accurate and cheaper, consequently more common in wearable and hand-held devices. Secondly, the user needn't worry about the changeable environment of light and sound, since inertial sensors still work well even in dark and noisy scenes [2]. Thirdly, such a system allows the user to write wherever and whenever he likes, unlike in optical situation the user has to stay in a narrow area and keep an appropriate angle to remain the camera effective.

Originally, people defines "air-writing" the trajectory

a) E-mail: xu.songbin@mail.scut.edu.cn

of hand movements, as another case for pen-based writing, but the real spatial trajectory is not directly accessible. Thus, the researchers applied inertial navigation algorithm for trajectory reconstruction [3], [4]. However, sensor drift and noise lead to serious deviations which accumulate over time due to the multiple integrations [5]. To avoid this issue, the optical sensor was hired to provide the spatial information of motion. Chen et al. [6] used a hybrid of opticalinertial sensors for motion tracking and achieved high accuracy, but additional devices may impose behavioral burdens on users. Some researchers proposed to abandon optical sensor. Amma et al. [7] developed a data glove with inertial sensor for air-writing, showing that without trajectory reconstruction, air-writing recognition is still possible, but wearing a glove may still be an undesirable burden [6]. It's better to use something handy and carry-on to write, so that people would feel natural when air-writing is required. Hand-held and wearable devices such as watch, phone and game controller, with inertial sensors built-in, are exactly suitable carrier of air-writing.

Many existing works [2]–[12] have achieved outstanding performance in air-writing recognition, some of which are summarized in Table 1. The algorithms are not limited to traditional methods like DTW [8] and CHMM [9], [10] but also neural networks [11], [12]. Most researchers introduced specific rules during data collection, such as space limitation, stroke orders, sampling rates, etc. However, very few of them laid stress on explaining these constraints' effects on system performance. Why are the writing constraints necessary in air-writing? To answer this question, we firstly proposed an air-writing recognition system using LSTM-RNN, the end-to-end trainable system accepts input of any durations requiring no extra feature engineering, then we made several quantitative analyses on three datasets under different constraints settings. We drew a conclusion that proper writing constraints improve classifier's performance, and we also discovered that the LSTM-RNN outperforms traditional methods on three aspects, namely, higher accuracy, better and more stable real-time performance, and great flexibility and scalability in both group and universal usage.

2. Airwriting Style

In this paper, air-writing refers to writing isolated Arabic numerals and English letters in the air using hand-held or wearable devices embedded with inertial sensors. Then, the

Manuscript received March 3, 2018.

Manuscript revised September 27, 2018.

Manuscript publicized January 28, 2019.

[†]The authors are with the South China University of Technology, Guangzhou, 510640 China.

DOI: 10.1587/transinf.2018EDL8042

Ref.	Classes	Algorithms	Writers	Constraints	Datasets	Sampled at	Sensors	Accuracy
[3]	10 Arabic numbers	reconstruct trajectory	15	stroke orders,	24 per writer	N/A	2 accelerometers	≈96~100%
	3 extra gestures	bayesian networks		writing space	per gesture		3 gyroscopes	
[4]	10 Arabic numbers	reconstruct trajectory	1	stroke orders,	NI/A	N/A	2 accelerometers	N/A
	26 uppercase letters			writing space	IN/PA		3 gyroscopes	
[6]	26 uppercase letters	left-to-right HMM	22	box-writing,	10 per writer	60Hz	Wiimote + Leap	92.0~98.2%
				stroke orders	per gesture			
[6]	40 words	left-to-right HMM	22	Same as above	5,400	60Hz	Wiimote + Leap	85.07%
	10 Arabic numbers	left-to-right HMM	1	keep wrist fixed		819.2Hz	1 accelerometer	76.7~100%
[7]				and limit size	500		1 gyroscope	
				of characters			- 85Fe	
[7]	26 uppercase letters	left-to-right HMM	10	Same as above	25 per writer	819.2Hz	1 accelerometer	81.9~95.3%
					per gesture		1 gyroscope	
[7]	100 words	repositioning HMM	10	Same as above	N/A	819.2Hz	1 accelerometer	42.0~96.0%
10							1 gyroscope	

 Table 1
 Summary of representative past related work on air-writing recognition using inertial sensors.

writing procedure is as follows. When the user writes in the air, the accelerometer senses the directional shifts and the gyroscope records the rotating of his wrist, which are stored in form of discrete vectors sequence and constitute meaningful strokes. Each motion character consists of several simple strokes following a sequential order, which results in tight interrelation and inter-dependency of air-writing. The signal waveform in air-writing is obviously different from those in traditional hand-writing. Firstly, the waveform represents motion transformation, rather than the real spatial trajectory. Secondly, the strokes are contiguously connected, but the demarcations among adjacent strokes are so ambiguous and fuzzy that they can hardly be distinguished. Last but not least, a writer writes on an imaginary plane in the air without any visual or haptic feedback, rather than writing on a physical surface which provides writing guidance.

Different writing habits lead to various writing scales, strength, speed, and stroke orders. Such differences affect the waveform's amplitude, pattern and order, and make it difficult to utilize raw signals for recognition. It is possible to decrease the effect of habits discrepancy by introducing strict rules and constraints of writing, even though they may not be user-friendly and convenient. In order to evaluate how air-writing constraints effect, we researched on three databases of isolated air-writing characters recorded under different levels of writing rules. The rules include the stroke orders of each character and the spatial range of writing in the air. We created two databases, one was collected under no handwriting constraints, called "NC" (No Constraints) database, the other was recorded under fixed stroke orders, and we call it "SOC" (Stroke Order Constraint) database. We also introduced a public database [6] for reference, which limits the box-writing style and the stroke orders, and we name it "SSC" (Space and Stroke Constraint) database.

Details of three databases are listed in Table 2 where "Acc", "Gyo" respectively represents tri-axial acceleration signal and angular velocity, and the characters consist of 10 Arabic numerals (Num), 26 uppercase (Upp) and 26 lowercase (Low) English letters.

3. The Proposed Air-Writing Recognition System

We propose to utilize LSTM-RNN to construct our air-

Table 2Detail settings of three databases.

6					
	NC	SOC	SSC		
Amount	1,130	14,530	8,571		
Classes	Num (10)	Num+Upp (36)	Num+Upp +Low (62)		
Data component	3d Acc	6d Acc+Gyo	6d Acc+Gyo		
Sampling frequency	100Hz	15Hz	50Hz		
Writers	40	49	22		
Constraints	none	stroke orders	stroke orders, box-writing		
Equipment	Wii remote	HTC G17	Wii remote		

writing recognition system, because we have analyzed airwriting signal's particularity, and we require LSTM with its remarkable memory for temporal modelling. The proposed system is illustrated in Fig. 1. Simple but necessary preprocessing includes moving average filtering to handle high frequency noise, with a window length of 3 or 5, as well as Z-score normalization to remove gravity and alleviate effect of writing scale and strength. The network can learn the pattern from original waveform, so manual feature selection and extraction are not required. Once trained, it is able to directly map an unlabelled sequence of inertial signals to a specific character, in other words, it is end-to-end trainable. In addition, due to LSTM's nature, the system accepts input sequences in arbitrary lengths.

3.1 Feature Learning and Recognition Based on LSTM

LSTM is the core of the proposed system, including an input layer, a recurrent hidden layer and an output layer. Inputs to LSTM are a sequence of vectors through time-steps t such as $x = \{x_1, x_2, ..., x_T\}$ rendered as a ribbon shaped matrix, where each $x_t = [A_x, A_y, A_z, G_x, G_y, G_z]$ represents tri-axial acceleration and angular velocity. The hidden layer receives the hidden activations from last time-step and the current inputs from the input layer, and a LSTM block defines "memory" in form of activations like $h_t = f(x_t, h_{t-1})$, where f is a complex function of non-linearity and varies for different hidden units. The hidden state integrates necessary information over past time-steps, and helps the network to decide its future behavior and make accurate predictions at the output layer. We also averaged all the hidden states vectors before transferring them to the Softmax.



Fig. 1 Illustrations of the proposed air-writing recognition system.

3.2 LSTM Training

We reshaped the databases into several mini-batches of batch-size 64, and implemented Back Propagation Through Time for neural network training. We used ADADELTA to adaptively adjust the learning rate for gradient descent. We also introduced dropout with kept probability=0.5 and an early stopping strategy to alleviate over-fitting.

4. Experiments and Quantitative Analyses

We compare LSTM with 3 recent methods on air-writing recognition, namely, dynamic temporal wrapping with affinity propagation clustering (DTW-AP), continuous hidden Markov model (CHMM) and convolutional neural network (CNN), by adapting them to fit our databases. The LSTM is a basic LSTM structure with forget gate and peephole connections, and the CNN has the same structure to [13].

To obtain a more intuitive comparison, we separated SOC and SSC into mutually disjoint subsets, respectively the "Num", the "Upper" and the "Lower". On each subset we evaluated a model with 5-folds cross validation in both mixed-writer and writer-independent way. All the codes were implemented by ourselves, without introducing any public integrated toolkits.

4.1 Evaluation of Writing Constraints

Writing constraints do help improving recognition performance. As Table 3 shown, in mixed-writer case models on SSC (most constrained) always outperform SOC and NC (free). However, in writer-independent, they perform a little worse on SSC than SOC. Due to the lack of training data of lowercase letter and Arabic numerals, it seems that neural network traps itself in specific writing styles it has already learned. Therefore, we recommend to add proper air-writing

 Table 3
 Average error of 5-folds cross validation on mixed-writer.

	-			
	DTW-AP	CHMM	CNN	LSTM
NC All	16.55 ± 2.55	27.96 ± 1.20	18.05 ± 1.92	9.40 ± 1.87
SOC Num	10.06 ± 0.71	4.70 ± 0.98	3.81±0.73	3.29 ± 1.04
SOC Upp	20.29 ± 0.71	10.89 ± 0.67	6.99±1.16	4.60 ± 0.45
SOC All	23.00 ± 0.46	12.48 ± 0.83	9.76±1.06	6.04 ± 0.40
SSC Num	3.33±1.56	1.00 ± 1.09	1.00 ± 0.37	2.67 ± 1.49
SSC Upp	7.81±0.85	2.71±0.66	1.20 ± 0.15	1.67 ± 0.50
SSC Low	5.31±1.33	1.78 ± 0.72	1.70 ± 0.54	3.20 ± 0.57
SSC All	30.51 ± 0.87	4.09 ± 0.47	5.06 ± 0.52	5.25 ± 0.31

Table 4 Average error of 5-folds cross validation on writer-independent.

DTW-AP	CHMM	CNN	LSTM
43.75±6.01	34.36 ± 8.74	25.56 ± 8.65	17.15 ± 5.68
56.18 ± 5.46	7.68 ± 3.56	8.47±1.83	6.85±1.61
47.91 ± 9.45	15.51 ± 4.22	15.87 ± 2.13	13.04 ± 2.41
N/A	17.71±4.69	18.62 ± 2.56	14.84 ± 2.44
43.50±13.1	$3.30{\pm}4.09$	10.20 ± 5.72	12.20 ± 2.39
59.47 ± 5.41	8.96 ± 1.54	4.41±1.65	5.66 ± 1.12
57.16 ± 9.61	23.62 ± 5.25	16.44 ± 5.00	22.30 ± 6.42
N/A	37.31±2.91	17.64 ± 3.60	19.91±3.71
	DTW-AP 43.75±6.01 56.18±5.46 47.91±9.45 N/A 43.50±13.1 59.47±5.41 57.16±9.61 N/A	DTW-AP CHMM 43.75±6.01 34.36±8.74 56.18±5.46 7.68±3.56 47.91±9.45 15.51±4.22 N/A 17.71±4.69 43.50±13.1 3.30±4.09 59.47±5.41 8.96±1.54 57.16±9.61 23.62±5.25 N/A 37.31±2.91	DTW-AP CHMM CNN 43.75±6.01 34.36±8.74 25.56±8.65 56.18±5.46 7.68±3.56 8.47±1.83 47.91±9.45 15.51±4.22 15.87±2.13 N/A 17.71±4.69 18.62±2.56 43.50±13.1 3.30±4.09 10.20±5.72 59.47±5.41 8.96±1.54 4.41±1.65 57.16±9.61 23.62±5.25 16.44±5.00 N/A 37.31±2.91 17.64±3.60

limitations as a guidance for writing and an assistant for recognition. For example, a set of carefully designed stroke orders is easy to learn and accept for the public.

4.2 Evaluation of Recognition Performance

In horizontal contrast, the proposed system based on LSTM outperforms the others on all subsets of NC and SOC, except SSC, on which LSTM falls slightly behind CNN but still acceptable. On the other hand, the DTW-AP performs great when the character set is small, but becomes less effective when the number of class increases, until hardly reliable on SSC with 62 classes. CHMM's performance is remarkable and even outperforms CNN and LSTM on the SSC's "All", but it performs the worst on NC.

Table 5Average time consumption (mili-seconds).

	Sample Length (points)	CHMM	LSTM
NC All	204.5±42.3	26.97	64.76
SOC Num	27.31±10.6	2.43	9.03
SOC Upp	26.42±10.3	5.90	8.65
SOC All	26.67±10.4	8.40	8.74
SSC Num	109.4±35.1	33.92	34.92
SSC Upp	110.0 ± 42.9	79.76	34.90
SSC Low	85.16±25.2	48.79	27.49
SSC All	105.7±41.0	164.80	33.12

4.3 Evaluation of Real-Time Performance

We implemented recognition tests for each dataset and their corresponding trained models. For a real-world air-writing application, the integrated time consumption should include the time it costs for data pre-processing plus the network's forward computation. Considering an application scene on Android, we implemented them in JAVA running on PC with a quad-core CPU (Core i7-4790, 3.60GHz). As Table 5 shows, LSTM's delay mainly depends on the average duration of samples, relevant to sampling ratio. The proposed system's time consumption remains stable as the character classes increase, as a comparison, the CHMM was affected badly by the vocabulary size. This fact makes the proposed system more adaptive and flexible in real-world applications.

4.4 Evaluation of Scalability

Writer-independent tasks are more difficult than mixedwriter ones. As expected, the 4 models' performances on writer-independent decrease to different degrees. DTW-AP becomes totally unreliable, CHMM loses its accuracy on SSC, CNN is the best on SSC but failed on others, while the proposed system still outperforms, except that it drops a little on SSC. Therefore, the proposed system is more reliable in not only specific groups but also general public usage.

It's worth noting that the proposed system is flexible in its architecture. LSTM blocks can be replaced with basic RNN or GRU units, bi-directional structure is also allowed. In addition, neuron numbers of each layer, activation functions and initialization also vary. Selection of the above is mostly dependent on the problem and the nature of data. For example, LSTM is better than RNN when data is deep in time, GRU is simpler and faster in training than LSTM, etc.

5. Conclusions and Future Work

In this paper, we made quantitative analyses on effects from constraints of Air-Writing. Specifically, we introduced three databases of different writing constraints settings, and then we proposed a system based on LSTM-RNN for contrastive experiments against three other common models. Results show that with a proper level of air-writing constraints, the system can realize more reliable recognition, thus, writing constraints are necessary and important. As for algorithm, we recommend LSTM, due to its advantages including higher accuracy, better and more stable real-time performance, great flexibility and scalability in both group and universal usage.

The recognition task on isolated air-writing character is fundamental to further research. A next step is to extend this system to words and sentences recognition, another direction would be Chinese character recognition. More potential applications require further investigations.

Acknowledgments

This work was supported in part by National Natural Science Funds (no. 61771199), Guangdong Science and Technology Research Plan (no. 2015B010131004, 2017A010101027) and the Fundamental Research Funds for the Central Universities of China (no. D2171560).

References

- G. Sampson, "Typology and the study of writing systems," Linguistic Typology, vol.20, no.3, pp.561–567, 2016.
- [2] A. Akl, C. Feng, and S. Valaee, "A novel accelerometer-based gesture recognition system," IEEE Trans. Signal Process., vol.59, no.12, pp.6197–6205, 2011.
- [3] S.-J. Cho, J.K. Oh, W.-C. Bang, W. Chang, E. Choi, Y. Jing, J. Cho, and D.Y. Kim, "Magic wand: A hand-drawn gesture input device in 3-D space with inertial sensors," 9th International Workshop on Frontiers in Handwriting Recognition, pp.106–111, 2004.
- [4] W.-C. Bang, W. Chang, K.-H. Kang, E.-S. Choi, A. Potanin, and D.-Y. Kim, "Self-contained spatial input device for wearable computers," 7th IEEE International Symposium on Wearable Computers, pp.26–34, 2003.
- [5] W.H. Baird, "An introduction to inertial navigation," American Journal of Physics, vol.77, no.9, pp.844–847, 2009.
- [6] M. Chen, G. AlRegib, and B.-H. Juang, "Air-writing recognition— Part I: Modeling and recognition of characters, words, and connecting motions," IEEE Trans. Human-Mach. Syst., vol.46, no.3, pp.403–413, 2016.
- [7] C. Amma, M. Georgi, and T. Schultz, "Airwriting: Hands-free mobile text input by spotting and continuous recognition of 3D-space handwriting with inertial sensors," 16th International Symposium on Wearable Computers, pp.52–59, 2012.
- [8] Y.-L. Hsu, C.-L. Chu, Y.-J. Tsai, and J.-S. Wang, "An inertial pen with dynamic time warping recognizer for handwriting and gesture recognition," IEEE Sensors Journal, vol.15, no.1, pp.154–163, 2015.
- [9] D. Arsenault and A.D. Whitehead, "Gesture recognition using Markov systems and wearable wireless inertial sensors," IEEE Trans. Consum. Electron., vol.61, no.4, pp.429–437, 2015.
- [10] S. Xu and Y. Xue, "Air-writing characters modelling and recognition on modified CHMM," IEEE International Conference on Systems, Man, and Cybernetics, pp.001510–001513, 2016.
- [11] S. Duffner, S. Berlemont, G. Lefebvre, and C. Garcia, "3D gesture classification with convolutional neural networks," IEEE International Conference on Acoustics, Speech and Signal Processing, pp.5432–5436, 2014.
- [12] G. Lefebvre, S. Berlemont, F. Mamalet, and C. Garcia, "Inertial gesture recognition with BLSTM-RNN," Artificial Neural Networks, Springer Series in Bio-/Neuroinformatics, vol.4, pp.393–410, Springer International Publishing, 2015.
- [13] Y. Chen and Y. Xue, "A deep learning approach to human activity recognition based on single accelerometer," IEEE International Conference on System, Man, and Cybernetics, 2015.