ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 6, Issue 4, April 2017

CLICK ON AIR

D. Deepa
ME Student, Dept. of ECE
Maharaja Institute of Technology
Coimbatore

R. Dharmalingam
Assistant Professor& Head, Dept. of ECE
Maharaja Institute of Technology
Coimbatore

Abstract: Air-writing differs from conventional handwriting; the latter contains the pen-up-pen-down motion, while the former lacks such a delimited sequence of writing events. We address air-writing recognition problems in a pair of companion papers. In Part I, recognition of characters or words is accomplished based on six-degree-of freedom hand motion data. We address air-writing on two levels: motion characters and motion words. Isolated air-writing characters can be recognized similar to motion gestures although with increased sophistication and variability .For motion word recognition in which letters are connected and superimposed in the same virtual box in space.AhiddenMarkovmodelisusedforairwritingmodelingandrecognition.We show that motion data along dimensions beyond a 2-D trajectory can be beneficially discriminative for air-writing recognition. We investigate the relative effectiveness of various feature dimensions of optical and inertial tracking signals and report the attainable recognition performance correspondingly. The proposed system achieves a word error rate of 0.8% for word-based recognition and 1.9% for letter-based recognition.

***Index Terms*—Air-writing, handwriting recognition, usability study, 6-DOF motion**

I. INTRODUCTION

From a user's perspective, air-writing can be realized in several ways. The first and the most essential is writing of individual isolated letters in an imaginary box in the space, one at a time. The second is the writing of multiple letters across the space from left to right in a style much like writing on a paper. Finally, one can also write several letters, stacked contiguously one over another in the same imaginary box. We call these isolated, connected, and overlapped air-writing, respectively.

The problem of air-writing recognition can be approached progressively. Isolated airwritingcarriestheassumptionthatthe handmotiontorenderaletterhasalreadybeenroughlylocalized in time and in space. Localization of motion rendering may be accomplished by use of a tracker, which can be easily turned ON or OFF, to signify the beginning and ending of a writing activity. The localization is only approximate and not fluctuation-free because most users cannot precisely synchronize the tracker control (ON–OFF) and thetruewritingtrajectory. Thisissimilar to the notorious problem of end-pointing in spoken utterance recognition even with a push-to-talk control.

Between the approximate endpoints, the motion trajectory forms a letter that resembles a unistroke writing. Study of isolated air-writing is essential to provide the technological foundation for subsequent challenges. Beyond isolated letters, recognition of "word" poses two additional challenges: the contiguous writing of letters without segmentation, and the incorporation of sequential constraints between letters. The distinction between connected and overlapped air-writing mainly arises from system usability; the latter requires less limb movement.

Air-writing is a preferred text input method on a motion-based user interface. We evaluate the text input performance of the proposed air-writing system and the use of a virtual keyboard with both subjective and objective metrics.

11 .R ELATED WORK

Hidden Markov models (HMMs) are widely used for online handwriting recognition . Ligature models are proposed to address online recognition of cursive handwriting, in which successive letters are connected without explicit pen-up moves. Motion-based handwriting can also be considered in parallel to motion gestures or sign language. Motion gesture recognition has been studied with different types of motion tracking devices . It was achieved with inertial sensors attached to a glove. Sign language is more sophisticated than motion gestures. Many sign language recognition systems use HMMs with various sensing technologies, such as data gloves and vision-based techniques

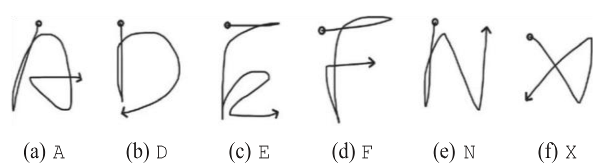
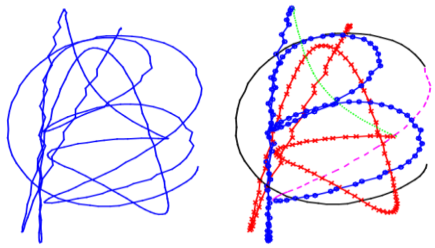


Fig. 1: Illustrations of the unistroke writing of isolated letters. (a) A. (b) D. (c) E. (d) F. (e) N. (f) X.



(a) ABC (b) segmented ABC

Fig. 2:Two-dimensional projected trajectory of a motion word. (a) ABC. (b) Segmented ABC

III. AIR-WRITING WITH SIX-DEGREE-OF-FREEDOM MOTION TRACKING

A. Unique Writing Style

Air-writing is fundamentally different from conventional handwriting on paper or a surface, which provides no haptic feedback. Similar to motion gestures, air-writing is tracked with a continuous stream of sensor data, and the writing is intuitively rendered in the air in unistroke without any pen-up and pendown information. The user envisions a writing box in the space andwrites in this imaginary space without haptic feedback. Air-writing also does not require visual feedback. Air-writing consists of two levels: motion characters and motion words. Motion characters are isolated alphanumeric letters written in one continuous stroke.

One level up, a motion word is formed by connecting motion characters with ligature motions in-between. When there is no haptic or visual feedback, the ordinary left-to-right writing style is difficult to maintain without overlap or shape distortion.Inour preliminary experiment, we discovered that users tend to shrink and overlap the last few letters of a word when the envisioned “writing space” is impacted by limited arm range. Therefore, we ask the user to write every character of a word in a layerby-layer manner, overlapping all letters of the word in the same envisioned virtual box, a writing style we term “overlapped airwriting,” which supersedes the usual connected writing style and appears to be more suitable for air-writing.

B. Six-Degree-of-freedom Motion Tracking and Data Acquisition

We use a hybrid framework for 6-DOF motion tracking: the Worldviz PPT-X4 for optical tracking of the position of the infrared tracker and the Wii Remote Plus (Wiimote) for the inertial measurements of the acceleration and angular speed. The orientationisderived fromafusionof theacceleration and angular speed data. The system tracks a specially designed handheld device and provides both explicit (position and orientation) and implicit (acceleration and angular speed) 6-DOF data sampled at 60 Hz.

IV. AIR-WRITING PROCESSING AND MODELING

A. Feature Processing

From the 6-DOF motion data, we derive five features (observations): position *P* and velocity *V* from optical tracking, orientation *O*, acceleration *A*, and angular speed *W* from inertial tracking. Let$P^o = [p_x, p_y, p_z]^\top$ denote the positions, and $V^o = [\Delta p_x, \Delta p_y, \Delta p_z]^\top$ the rate of change in position. The orientation is represented in quaternion, $O^o = [q_w, q_x, q_y, q_z]^\top$

	avg	std		avg	std		avg	std
A	159.5	37.4	J	60.6	12.0	S	92.7	17.8
B	156.7	37.1	K	136.8	26.4	T	88.6	16.4
C		19.8	L	64.8	15.0	U	73.5	14.1
	77							
.3								
D	118.7	24.9	M	146.4	29.4	V	67.2	11.5
E	190.8	48.6	N	115.7	21.1	W	110.4	19.3
F	132.6	27.4	O	85.1	17.4	X	91.3	16.1
G	149.7	35.1	P	107.6	20.3	Y	105.7	20.5
H	137.5	29.9	Q	119.4	26.4	Z	94.1	18.6
I		10.3	R	134.9	24.5			
	42							
.8								

TABLE I

DURATIONS (IN NUMBER OF SAMPLES) OF MOTION CHARACTERS BY 22 SUBJECTS

B. Air-Writing Modeling

HMM models for motion characters can be readily concatenated to form a motion word with additional connecting ligature motions. Such a modeling methodology is common, known as sub word modeling, in large vocabulary continuous speech recognition to circumvent insufficient training data issues. It is relatively easy to collect sufficient data of each gesture and straightforward to model each gesture directly from its own recordings. However, the vocabulary of air-writing can easily be thousands of words, and it is difficult to collect enough data for every word in the vocabulary. The data sufficiency problem prevents a designer from directly using whole “word” models

The word-level HMM model is built upon these individual character models. A motion word is formed by connecting motion characters with ligature motions. Here, we define the ligature as the motion from the ending point of the preceding character to the starting point of the following character

	Start point		End point
S1	BDEFHKLMNPRTUVWXYZ	E1	BDSX
S2	AIJOQ	E2	ITY
S3	CGS	E3	CEGHKLMQRZ
		E4	JP
		E5	AF
		E6	O
		E7	NUVW

TABLE II CLUSTERS FOR START AND END POINTS OF CHARACTERS

In Table II, we manually clusters characters according to the position of the starting and the ending points. Based on the hard clustering, we generate several general questions, such as: *"Does the previous letter belong to* E1*?"* and *"Does the next letter belong to* S2*?"* To take into account both the endpoint position and the stroke direction, we further divide the hard clustering to create more detailed questions, e.g., *"is the previous letter* B *or* D*(ends at bottom left with a right-to-left)*

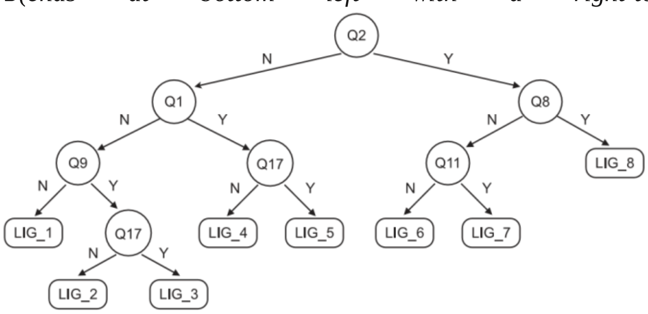


Fig 3:Illustrative decision tree that results in eight clustered ligatures.

In a motion word, we do not have or need the character-level segmentation. With the composite word HMM, character and ligature segmentation (alignment) can be simultaneously accomplished during recognition. Because the motion trajectories of characters and ligatures usually blend together, the ground truth of segmentation may be ambiguous. To have a better understanding of the ligature motions, we manually segmented all the motion words in the 40-word vocabulary recorded by subject M1. The manual segmentation is used for initial estimates of ligature models, which is proven to work better than ones that are initialized with zero means and global variances.

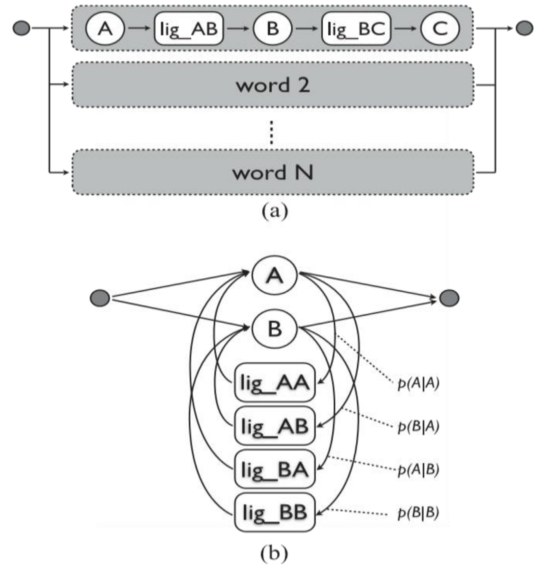


Fig 4:Decoding word networks. (a) Word-based. (b) Letter-based (simplified)

V. MOTION CHARACTER AND MOTION WORD RECOGNITION EVALUATION

We first evaluate motion character recognition with the five basic features (P, V, O, A, W) and different combinations of them, including PV, AWO, and $PVAWO$. The combinations of features actually correspond to different motion tracking devices. PV is the feature set derived purely from optical tracking, and AWO can be considered the full feature set from inertial measurements. PV AWO uses the available data from a hybrid 6-DOF motion tracking system.

features	CER (%) average	std
P	3.72	(3.60)
V	6.12	(2.88)
O	3.81	(5.05)
A	7.97	(7.38)
W	7.92	(3.34)
PV	1.61	(2.16)
AWO	1.84	(2.37)
$PVOAW$	1.05	(1.23)
$\hat{P}$	3.88	(3.55)
$\hat{V}$	6.15	(2.69)
$\hat{P}\hat{V}$	1.61	(2.06)
$\hat{P}\hat{V}OAW$	1.05	(1.33)

TABLE IV
CER OF MOTION CHARACTER RECOGNITION

The letter-based decoding network allows arbitrary decoded letter sequences and can handle OOV words. Another advantage is that letter-based word recognition allows progressive decoding while the user is writing, unlike the wordbased recognition that requires the user to complete a word. The freedom of arbitrary sequences comes at a price of sacrificing the contextual information from the vocabulary

A. Motion Character Recognition

The HMMs of motion characters are trained and tested with isolated characters, and we show the character error rate (CER) of leave-one-out cross validation with different features in Table IV. First, we compare the discriminative power of the basic features. The explicit 6-D features (P, V , and O) outperform the implicit 6-D features (A and W).

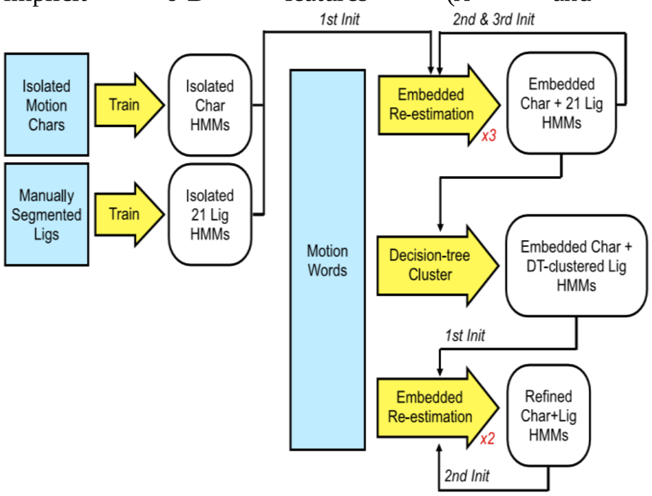


Fig. 5: Flow of the embedded reestimation for character and ligature models

B. Reestimated Character and Ligature Models

From the experiment mentioned above, we obtain the HMMs of isolated motion characters and use them to initialize the characterHMMsformotionwordrecognition.Toconstructtheword model, we also need HMMs for ligatures. In Section IV-B, we propose two approaches to model ligatures: hard clustering and decision tree. First, we extract the ligatures from the manually segmented motion words in the 40-word vocabulary written by subject M1.

C. Word-Based Motion Word Recognition

For word-based word recognition, we use the refined HMMs of character and 21 hard clustered ligatures to build the decoding word network as shown in Fig. 4(a). The word-based word recognition is formulated as a one-out-of-N problem, where N is the vocabulary size. In the word-based decoding network, each path is a word model synthesized from corresponding character and ligature HMMs, and the letter sequences are tightly restricted to the vocabulary.

V1. Results

We show the average writing/typing time and total traverse distance for words of different length in Table VIII. Because airwriting is recognized on a word basis, we report the average number of attempts to correctly input a word. Longer words tend to have higher recognition accuracy and hence need fewer attempts. The average writing time of a two-letter word is 3.9 (= 5.4/1.38) s. For virtual keyboard, we report the average number of extra keystrokes, e.g., a typo and a backspace count as two extra keystrokes.

Question	air-handwriting	virtual keyboard
1. Intuitiveness [5: most intuitive]	4.10	4.75
2. Arm fatigue level [5: no fatigue]	3.05	3.10
3. Vote for inputing a short word (2-3 letters)	16	4
4. Vote for inputing a long word (4+ letters)	11	9
5. Satisfaction of recognition performance [5: most satisfied]	4.25	-

TABLE V
USABILITY RESULT OF AIR-WRITING AND VIRTUAL KEYBOARD (SUBJECTIVE RATING FROM 1 TO 5)

Words-per-minute (WPM) is a common performance metric for text input efficiency. WPM is computed based on correctly inputwordunits,whereonewordunitisfiveletters(keystrokes). The WPM of air-writing and virtual keyboard are 5.43 and 8.42 , respectively. Compared to conventional text input methods , the WPM of pen-based handwriting without recognition is in the range of 15–25, and the WPM range of QWERT typing is 20–40. Although handwriting is not the fastest, it is the most primitive method for text input. Motionbased text input methods are roughly three to five times slower than the conventional ones because relatively large and unconstrained control motions are involved. Our study indicates the speed for these alternative text input methods on a motion-based user interface.

The objective metrics show that air-writing is roughly 1.5 times slower and three times longer in motion footprint than the virtual keyboard. However, we get quite interesting results from the subjective evaluation as shown in Table V. Air-writing is a variation of conventional writing, and virtual keyboard follows the same metaphor of typing on a touchscreen. Both methods are intuitive to users and have neutral scores for the arm fatigue level. Motions in the air involve more muscles than keyboard or touch-based interaction and thus cause more fatigue. Even though the motion footprint of air-writing is three times larger, it does not directly reflect arm fatigue ratings. The arm fatigue level relates to the writing or typing style. For example, airwriting could cause less fatigue for a user who rests the elbow andwriteswiththeupperarmandwristthanauserwhoholdsthe whole arm in the air.

V11. CONCLUSION

Air-writing is unistroke without pen-up/pen-down information. The writing style and motor control are different from ordinary pen-based writing due to lack of haptic and vision feedback. We separate air-writing in two levels: motion characters and motion words. Motion characters are handled similar to motion gestures, and each character is modeled with a HMM. A motion word can be modeled by concatenating character and ligature models. We present two approaches to model ligatures:hardclusteringanddecisiontree.Theformerisproven to be sufficient for word-based word recognition. The latter provides better capability of ligature modeling, which improves the performance of letter-based word recognition. and stringent requirement on the accuracy. On the other hand, letter- The word-based word recognition achieves relatively low WER but is not able to recognize OOV words. The word-based recognizer is suitable for applications that have a limited vocabulary based word recognition has around 10% WER but can handle arbitrary letter sequences and progressive decoding. To substantially improve the letter-based recognition accuracy, the system can provide suggestions with *n*-best decoding and lets the user choose the right one.

A user study investigates input speed, motion footprint, physical strain, and subjective evaluation of two motion-based text input methods: air-writing and virtual keyboard. The results suggest that air-writing is suitable for short and infrequent text input on a motion-based user interface.

V111. REFERENCES

[1]M. Chen, G. AlRegib, and B.-H. Juang, “Air-writing recognition—Part II: Detection and recognition of writing activity in continuous stream of motion data,” *IEEE Trans. Human-Mach. Syst*, to be published.

[2](2015). Leap motion. [Online]. Available: http://www.leapmotion.com

[3]M. Chen, G. AlRegib, and B.-H. Juang, “Feature processing and modeling for6dmotiongesturerecognition,”*IEEETrans.Multimedia*,vol.15,no.3, pp. 561–571, Apr. 2013.

[4]C. H. Blickenstorfer, “Graffiti: Wow!” *Pen Computing Mag.*, vol. 1, pp. 30–31, Jan. 1995.

[5] K.Perlin,“Quikwriting:Continuousstylus-basedtextentry,”in*Proc.11th Ann. ACM Symp. User Interface Softw. Technol.*, 1998, pp. 215–216.

[6] P. Isokoski and R. Raisamo, “Quikwriting as a multi-device text entry method,” in *Proc. 3rd Nordic Conf. Human-Computer Interaction*, 2004 , pp. 105–108.

[7] T. Koltringer,¨ P. Isokoski, and T. Grechenig, “Twostick: Writing with a game controller,” in *Proc. Graph. Interface*, 2007, pp. 103–110.

[8] (2013). Swype—Type fast, swype faster. [Online]. Available: http://www. swype.com/

[9] R. Plamondon and S. Srihari, “Online and off-line handwriting recognition: a comprehensive survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 63 –84, Jan. 2000.

[10] J. Makhoul, T. Starner, R. Schwartz, and G. Chou, “On-line cursive handwriting recognition using hidden markov models and statistical grammars,” in *Proc. Workshop Human Lang. Technol.*, 1994, pp. 432–436.