Integrating Declarative Models and HMMs for Online Gesture Recognition

Alessandro Carcangiu
Lucio Davide Spano
Department of Mathematics and Computer Science,
University of Cagliari
Cagliari, Italy
alessandro.carcangiu@diee.unica.it
davide.spano@unica.it

ABSTRACT

In the last years, the introduction of new, precise and pervasive tracking devices has contributed to the popularity of gestural interaction. In general, the effectiveness of such interfaces depends on two components: the algorithm used for accurately recognizing the user movements and the guidance provided to users while executing gestures. In this paper, we discuss a work in progress research for connecting these two components and increasing their effectiveness: the recognition algorithm supports the implementation of feedback the and feed-forward mechanisms, providing information on the identified gesture parts in real time, while developers define complex gestures starting from simple primitives.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; • **Computing methodologies** → *Machine learning*.

KEYWORDS

Gestures, Hidden Markov Models, Compositional gesture modelling, Online recognition, Feedback, Feedforward

ACM Reference Format:
Alessandro Carcangiu and Lucio Davide Spano. 2019. Integrating Declarative Models and HMMs for Online Gesture Recognition. In *24th International Conference on Intelligent User Interfaces (IUI '19 Companion), March 17–20, 2019, Marina del Rey, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3308557.3308709

1 INTRODUCTION

Over the years, the literature proposed different solutions to solve the gesture recognition problem. Among them, we can identify two peculiar classes: i) machine learning methods, like Hidden Markov Models (HMM) [5] or neural networks, and ii) compositional techniques, e.g. GestIT [6, 7] and Proton++ [3, 4].

IUI '19 Companion, March 17–20, 2019, Marina del Rey, CA, USA

ACM ISBN 978-1-4503-6673-1/19/03…$15.00
https://doi.org/10.1145/3308557.3308709

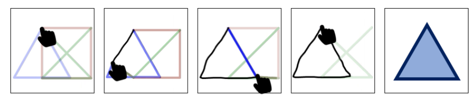

Figure 1: A guidance system for using stroke gestures in a simple geometry drawing application. It shows previous touch positions with a black line (back) and the possible completions (feedforward).

The former offers high accuracy, which is very important for building usable User Interfaces (UIs) As a drawback, the classification phase requires the entire gesture input sequence. This is a problem for creating many types of feedback and feed-forward systems [8] since they would require information during the gesture execution and not only when it finishes. On the other hand, declarative approaches describe gesture through the composition of smaller parts. This is useful for supporting guidance systems and gestural UIs implementation, at the cost of lower accuracy. In Figure 1, we depict an example of a simple gestural drawing application. In order to guide the user in drawing the shapes, the interface requires not only the recognition of the whole shape but also its parts (i.e., its sides).

In order to reduce the gap between machine-learning and compositional methods, we proposed DEICTIC (DEclaratIve and CompositTional Input Classifier) [1, 2]. It is a declarative and compositional approach, which achieves high recognition accuracy and sub-part identification combining HMM classifiers with a declarative gesture model [7]. DEICTIC provides a simple model language for stroke gestures for defining complex gestures through three different geometric primitives (point, line or arc) and a set of temporal operators (sequence, iteration, disabling, choice and parallel) for the composition. Internally, DEICTIC exploits HMMs for recognizing the basic gesture segments (primitives). Each operator corresponds to a connection graph for creating composite HMMs able to recognize complex gestures without any additional training.

2 ONLINE GESTURE RECOGNITION

DEICTIC supports online recognition (i.e. without the acquisition of the whole user movement) if the scale and the position gesture input are known. For instance, it supports online recognition if the gesture bounding box is (roughly) fixed. Unfortunately, such

$X = \boldsymbol{P(0,0) + L(2,-3)} + L(-2,0) + L(2,3)$ $\quad V = \boldsymbol{P(0,0) + L(2,-3)} + L(2,3)$

Figure 2: Two sample uni-stroke gestures (an X and a V) having the first sub-component in common (in red). Below the graphic representation, we show the DEICTIC expressions for modelling each stroke, with the common part in bold.

interaction does not correspond to the general case. The usual solution is a preprocessing step for the gesture data that normalises the reference system for each gesture and increases the recognition accuracy. In this work, we propose to extend DEICTIC for supporting online recognition in the general case. The goal is detecting when a subcomponent is completed, supporting feedback and feedforward mechanisms in real-time, independently from the scale and the position of user input. Differently from the original approach, in this version we employ a tree structure to generate a list of partial-gesture HMMs. The structure avoids the creation of duplicated HMMs, namely two or more HMMs associated to the same sub-expression. Instead, two or more gestures can share a subset of their components.

For instance, the X and V uni-stroke gestures in Figure 2 have the first sub-component in common. We first insert a new tree node for the entire expression that describes the X gesture (Node_A). Then, we split it by removing the last ground term, $L(2,3)$. We associate the resulting expression $P(0,0) + L(2,-3) + L(-2,0)$ to a new node (Node_B) which is a son of Node_A. We recursively apply the same operation to Node_B, and we obtain the Node_C. Then, we start splitting V. When the decomposition algorithm reaches the common part in the strokes ($P(0,0) + L(2,-3)$), we do not create another node, but we link the one we already created for X to the V tree. In this way, the gesture model contains only one node associated with the same sub-expression. Finally, the obtained tree is employed to generate the list of HMMs, one for each node.

We tested the accuracy of the proposed method using an adjusted version of 1$-dataset presented in [9]. It contains 330 repetitions of 16 single stroke gestures, represented as a sequence of points and the related time-stamp. We added for each point the information about gesture sub-component. In this preliminary test, we evaluated only those gestures which not contains arc primitives. We simulated an online dispatching, feeding the HMM with a single frame at a time. Figure 3 reports the accuracy achieved considering the top one, two, three and four likely gestures. It is worth pointing out that the accuracy follows a similar trend in each condition. In particular, we identified two areas, between 30% and 70%, where the mean accuracy decreases. This is because the HMMs require a few points for detecting the next sub-component and this causes a delay in the recognition.

3 CONCLUSIONS AND FUTURE WORK

In this paper, we extended DEICTIC [1, 2] towards the online gesture recognition in the general case, independently from their scale and position. On the one hand, it preserves the composition information on gesture parts, in order to support the development of gesture guidance systems through feedback and feedforward; on the other hand, the overall accuracy is less than 80%, in particular when the user has performed the 30% and 70% of the gesture. In future works, we would like to analyse the achievable accuracy by including those gestures which contain arc primitives. In addition, we plan to determine if other machine learning-based approaches are suitable to be combined with declarative approaches.

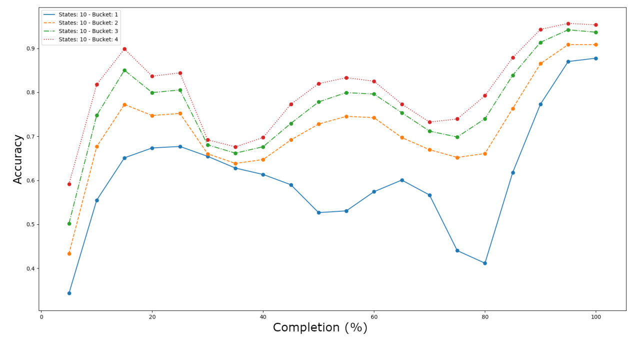

Figure 3: The mean accuracy evolution achieved considering the top one,two,three and four likely gestures. The x-axis represent the mean accuracy while the y-axis represents the gesture completion.

ACKNOWLEDGMENTS

The work has been founded by the D3P2 project (Sardinia regional government, CUP: F72F16002830002) and EmILIE project (Sardinia Regional Government and Fondazione di Sardegna, CUP: F72F16003030002).

REFERENCES

[1] Alessandro Carcangiu, Lucio Davide Spano, Giorgio Fumera, and Fabio Roli. 2017. Gesture modelling and recognition by integrating declarative models and pattern recognition algorithms. In *International Conference on Image Analysis and Processing*. Springer, 84–95.

[2] Alessandro Carcangiu, Lucio Davide Spano, Giorgio Fumera, and Fabio Roli. 2019. DEICTIC: A compositional and declarative gesture description based on hidden markov models. *International Journal of Human-Computer Studies* 122 (2019), 113–132.

[3] Kenrick Kin, B Hartmann, T DeRose, and Maneesh Agrawala. 2012. Proton++ : A Customizable Declarative Multitouch Framework. In *Proceedings of UIST 2012*. ACM Press, Berkeley, California, USA, 477–486.

[4] Kenrick Kin, B Hartmann, T DeRose, and Maneesh Agrawala. 2012. Proton: multitouch gestures as regular expressions. In *Proceedings of CHI 2012*. ACM Press, Austin, Texas, USA, 2885–2894.

[5] Lawrence R Rabiner. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* 77, 2 (1989), 257–286.

[6] Lucio Davide Spano, Antonio Cisternino, and Fabio Paternò. 2012. A Compositional Model for Gesture Definition. In *Proceedings of HCSE 2012*. Springer, 34–52.

[7] Lucio Davide Spano, Antonio Cisternino, Fabio Paternò, and Gianni Fenu. 2013. GestIT: a Declarative and Compositional Framework for Multiplatform Gesture Definition. In *Proceedings of EICS 2013*. ACM, 187–196.

[8] Jo Vermeulen, Kris Luyten, Elise van den Hoven, and Karin Coninx. 2013. Crossing the bridge over Norman's Gulf of Execution: revealing feedforward's true identity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1931–1940.

[9] Jacob O. Wobbrock, Andrew D. Wilson, and Yang Li. 2007. Gestures Without Libraries, Toolkits or Training: A $1 Recognizer for User Interface Prototypes. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)*. ACM, New York, NY, USA, 159–168.