Recognition of Unistroke Gesture Sequences

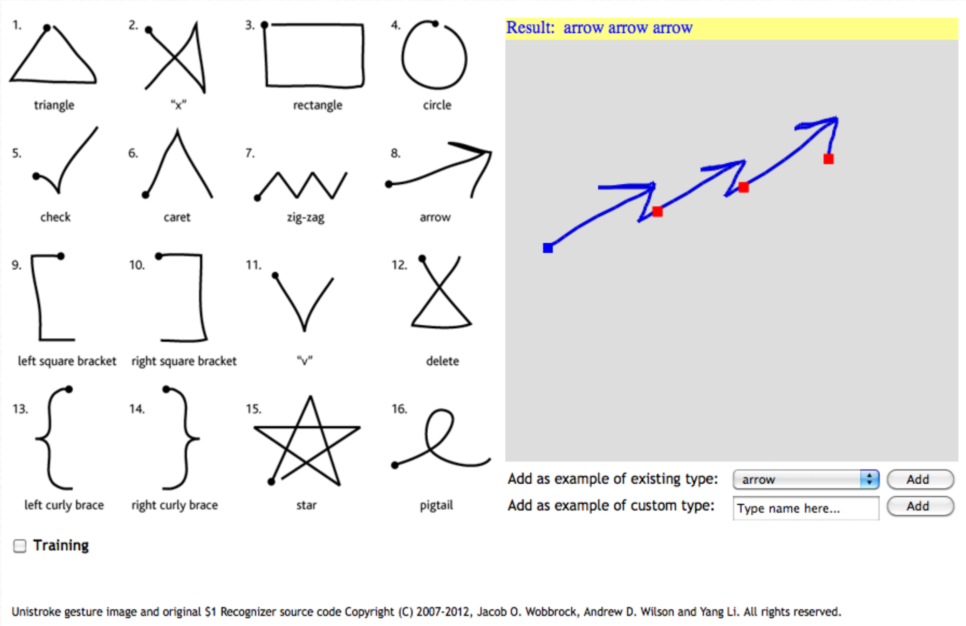
Presented by

Justin Permar

&

Arvind Krishnaa Jagannathan

Problem Statement



To develop a gesture recognition system, which will segment the input sequence consisting of multiple gestures, drawn in one stroke, into the constituent.

Our primary objective in solving this problem is to have a minimal set of training data in order to quickly build a prototype system.

Related Work

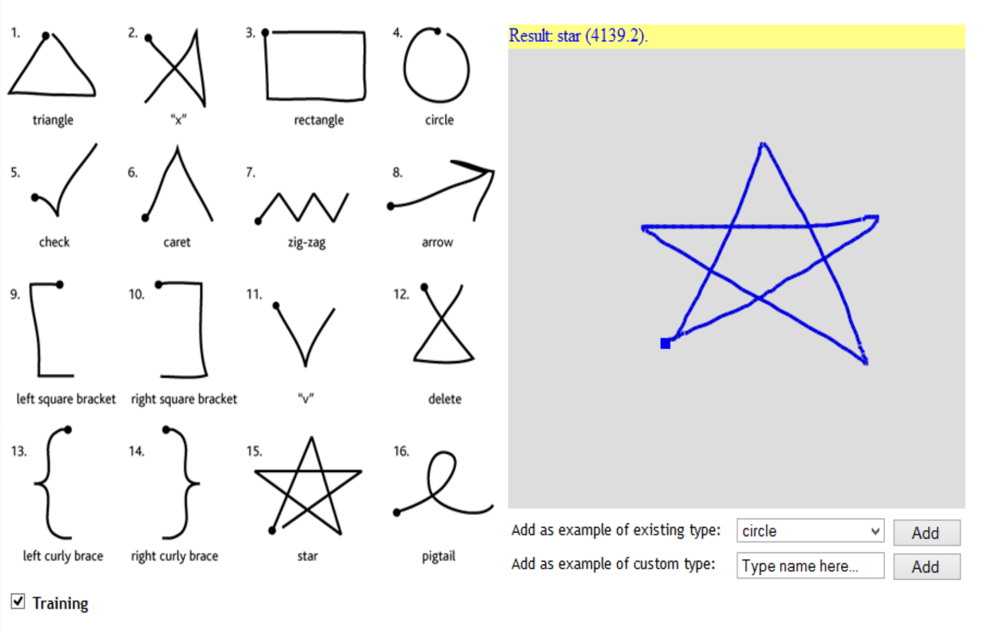
- Recognition of individual gestures in unistroke and multistroke sequences of digits – Yang et al: Training HMMs for continuous sequences
- Individual gesture recognition : Extracting several features from an input and constructing Hidden Markov Models for each gesture – Tanguay
- Recognizing an individual unistroke gesture sequence, using a distance measure from existing templates - $1 Recognizer - Wobbrock et al

Our Approach – Salient Features

- Extend $1 recognizer to identify multiple gestures in a gesture sequence.
- Hardest problem is that of segmentation – Classified in literature as “Inverse Perception Problem”.
- Based on a “Visual Affinity” approach – a type of geometric match.
- Employs Dynamic Time Warping to decide the distance from a given template.
- No restriction on gestures in the input sequence –
  - Arbitrarily long
  - Can contain any pre-defined gesture. New gestures can also be added to the system.

Approach – Training mode

- In training mode, the user draws a template of either an existing or new gesture.
- Only individual gestures are drawn – no need to train the system with gesture sequences.
- If misclassification occurs during training, user can indicate the intended gesture.
- Score next to the shape (“star”) is based upon a distance measure between the input and the indicated gesture.



Approach – Match using Dynamic Time Warping

- Euclidean distance used as the distance metric between the neighboring points of input sequence and each template.

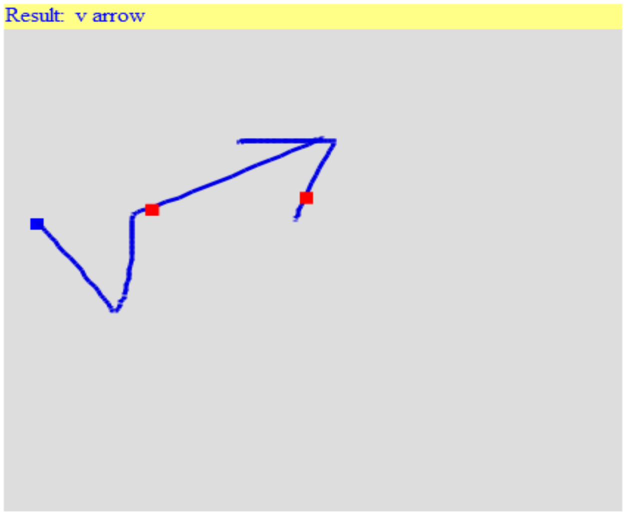
$$d = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2}$$

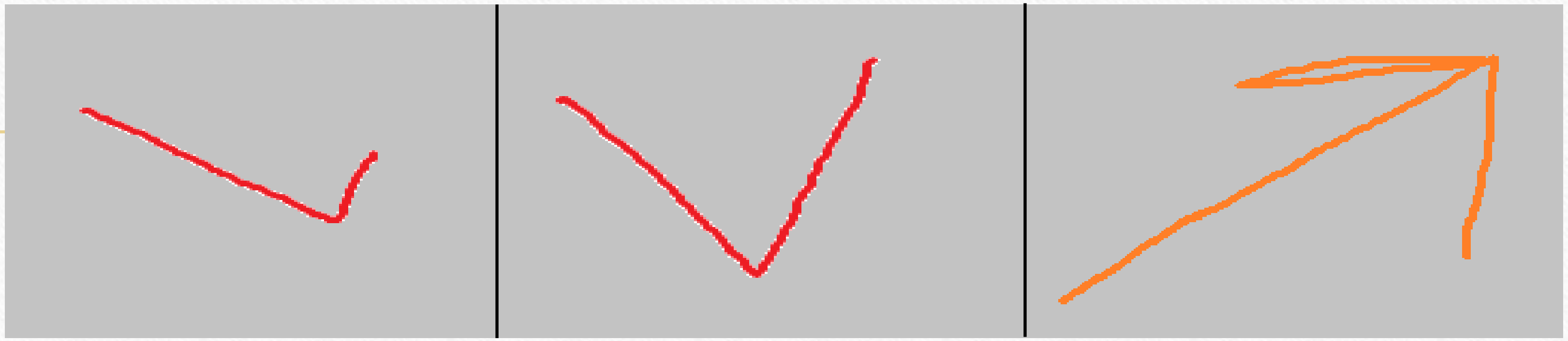
where (x_k, y_k) correspond to the points on the template under consideration

and (x_i, y_i) correspond to the points on the input sequence.

- For a given set of input points, if there is a template which gives a score lower than a fixed threshold, then that gesture is recorded as being part of the input.
- The identified portion of the input is "spliced off" and the matching procedure is repeated for the remaining input.

Approach – Segmentation and Recognition





Candidate input points
No Match

Slide the input set of points
Match with "v" (First dot)

Remaining input after splicing off the matched template

Input Gesture Sequence
Blue Dot – Start of the Sequence
Red Dots – Points at which sequence is segmented

Next set of candidate points from remaining input – No Match

Sliding the set of points
Match with "Arrow" (Final dot)

Results and Interpretation

Sequence Length	Accuracy Rate (%)
1	57.5
2	32.16
3	25.72

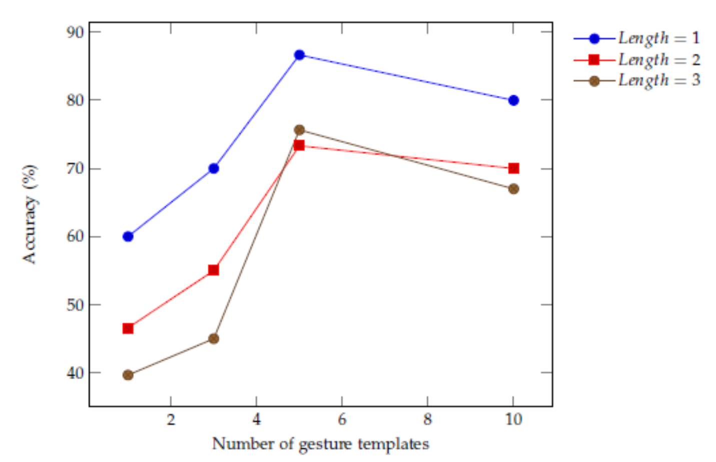
Table:1 Overall Accuracy – Exact Match

Sequence Length	Segmentation Accuracy(%)
1	71.5
2	71.5
3	52.27

Table:2 Segmentation Accuracy

Sequence Length	Relaxed Accuracy (%)
2	56.19
3	56.37

Table:3 Relaxed Accuracy



$$Relaxed\ Accuracy = \frac{Reported\ Gestures\ in\ input}{Total\ Gestures\ in\ input\ sequence}$$

Training Size Vs. Accuracy Rate graph, indicates the following:

1. Even for relatively low training samples $(n = 3)$ fairly good accuracy rates are reported.
2. An optimum recognition rate is achieved at $(n = 5)$
3. Too many templates, cause incorrect recognition due to "confusion" caused by many variations of an individual gesture.

Discussion

- Training only on individual gestures → lose features indicating a transition from one gesture to another (such as a pause)
- Construct HMMs for each individual gesture in the "gesture library" and then chain HMMs to construct arbitrarily long sequences of gestures.
- Compromise training time and computation for improved accuracy.

References

[1] Lyddane, Donald. United states attorneys' bulletin. Technical report, United States Department of Justice Executive Office for United States Attorneys, May 2006.

[2] J. Yang, Y. Xu, and C. S. Chen. Gesture interface: Modeling and learning. In Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference on, pages 1747–1752, 1994.

[3] D. O. Tanguay Jr. Hidden markov models for gesture recognition. Master's thesis, Massachusetts Institute of Technology, 1995.

[4] J.O. Wobbrock, A.D. Wilson, and Y. Li. 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, pages 159–168. ACM, 2007.

[5] Zygmunt Pizlo. Perception viewed as an inverse problem. Vision Research, 41(24):3145–3161, November 2001.