TEMPORAL INTERFRAME PATTERN ANALYSIS FOR STATIC AND DYNAMIC HAND GESTURE RECOGNITION

Kaoning Hu[1], Lijun Yin[2], Tianyang Wang[3]

[1] Texas A&M University-Commerce, [2] SUNY Binghamton, [3] Austin Peay State University

INTRODUCTION

- In the recent decade, the performance of hand gesture recognition has been significantly improved. However, there is still a non-negligible gap between lab experiments and real applications. Therefore, **every 1% of improvement is highly desired.**
- In this paper, we propose a three-level scheme to utilize the **temporal interframe pattern** on the recognition of **both static and dynamic hand gestures.**
- A **dynamic hand gesture** of finger movements can be considered as a **temporal sequence of static hand gestures.**
- The static hand gestures must also have temporal patterns, because **a human would not express a static hand gesture out of context.** Therefore we are inspired to exploit the **temporal patterns** to correct the error in the recognition of **static hand gestures.**

TEMPORAL PATTERN ANALYSIS

- The error in the recognition would cause the labeled hand gesture sequence to show **different temporal patterns** from the correct instances.
- However, a human user would not change the hand gesture unreasonably.
- The accuracy will be improved by the proper analysis of the temporal patterns of the consecutive frames.

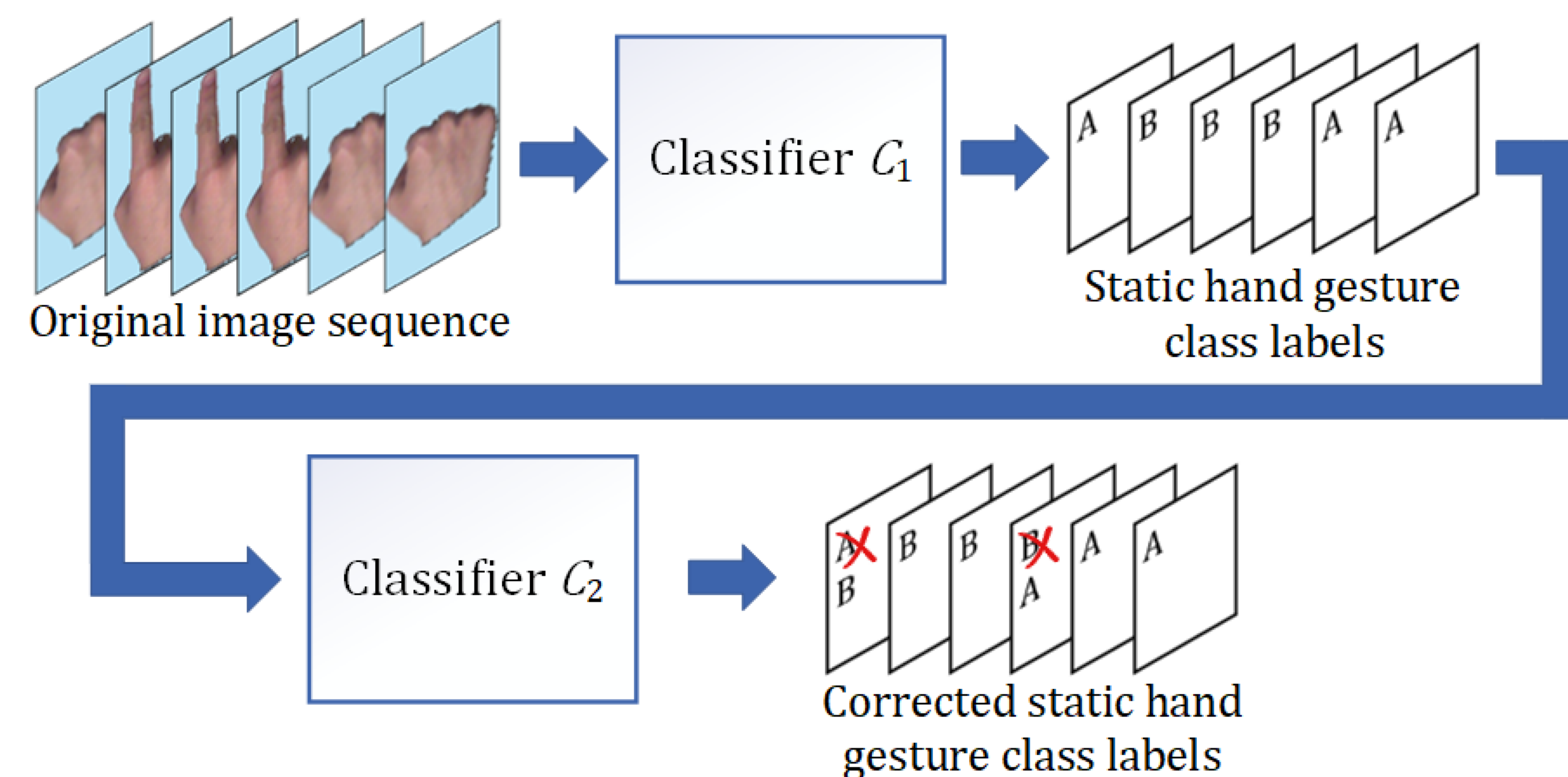A real sequence when the hand gesture changed from "point" to "grab":

(point) (point) (point) (grab) (grab)

A faulty sequence when a finger was not detected due to some error:

(point) (point) (grab) (point) (point)
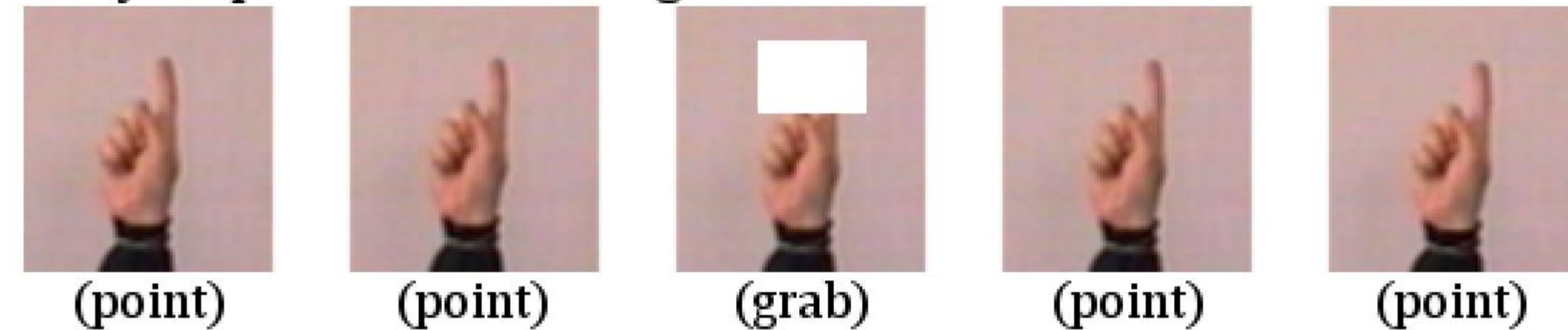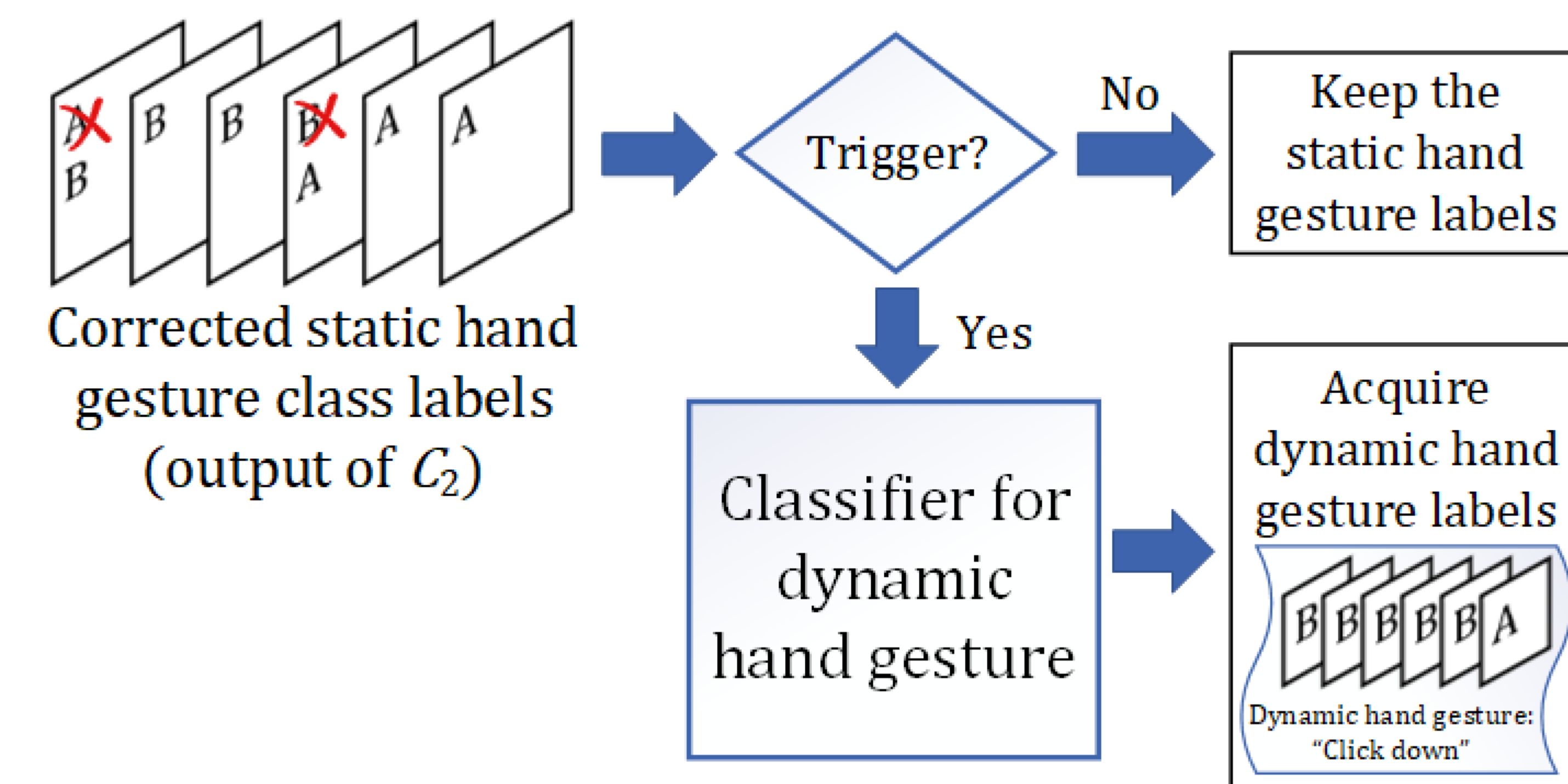
A SECOND-LEVEL CLASSIFIER

- If we input the label of the hand gesture of one frame and the labels of its previous frames to a classifier as the training data, the classifier will **learn the different patterns between incorrectly labeled frames and the real transition of the hand gestures.**
- The first classifier C_1 gives a class label to every frame of the image sequence, whereas the second classifier C_2 uses the labels generated by the first classifier to detect and correct the mistakes.

Original image sequence → Classifier C_1 → Static hand gesture class labels
A β β β A A

Classifier C_2 → Corrected static hand gesture class labels
A β β β A A

A THIRD-LEVEL CLASSIFIER

- A temporal sequence may be a single dynamic hand gesture, or a video sequence of several static hand gestures.
- We only trigger a classifier for dynamic hand gesture when the class label of static hand gesture of the current frame is different from the previous frame, or when the coordinates of the hand suddenly change.
- As demonstrated by Hu and Yin, using a sequence of labels as the input data for dynamic hand gesture recognition worked better than using the features directly extracted from the original image sequence. Therefore, the temporal sequence of labels generated by C_2 is used for dynamic hand gesture recognition.

THE THREE-LEVEL CLASSIFIER

Corrected static hand gesture class labels (output of C_2) → Trigger? — No → Keep the static hand gesture labels

Trigger? — Yes → Classifier for dynamic hand gesture → Acquire dynamic hand gesture labels

Dynamic hand gesture: "Click down"

- Any classifier that is relatively robust can be used as C_1.
- To reduce the detection latency, the class label of the current frame and its 4 previous frames as the input of C_2.
- Bayesian network is used as the classifier of C_2, and has the best performance.
- Hidden Markov Model (HMM) classifier is used as C_3 (classifier for dynamic hand gesture).

EVALUATION

- We have a dataset with 16 static hand gestures and 7 dynamic hand gestures. 3438 samples in total from 17 people.
- Without C_2, the accuracy of C_1 is 91.9%.
- C_2 improves the accuracy of **static hand gesture recognition** to **98.1%. (+6.2%)**
- 97.5% of the dynamic hand gestures were distinguished from static hand gestures.
- The HMM classifier C_3 achieves an accuracy of 98.3% across the seven dynamic hand gestures. This makes the overall accuracy of **dynamic hand gesture recognition** reach to **95.8%.**