Datasets for Motion-Capture-Based Hand Gesture Recognition

Andrew Gardner
abg010@gmail.com

Louisiana Tech University

May 7, 2017

This document describes the collection, features, and organization of three motion capture datasets for hand posture and gesture recognition. In each case, the Vicon motion capture system described in Section 1.1 was used to collected the data. Each dataset along with its accompanying file format is described in its own section.

1 General Remarks

Some general remarks on the data follow.

1.1 Problem Setting

This subsection describes the laboratory in which the research was conducted as well as the glove constructed to serve as the source of data for all algorithms analysis and development.

1.1.1 Laboratory

Ten Vicon MX T40 (4 megapixel) motion capture cameras available in the Micro-Aerial Vehicle and Sensor Networks (MAVSeN) Laboratory at Louisiana Tech University act as the source of data. The MAVSeN lab conducts research and development in small-scale vehicle design, cooperative intelligent sensing, and control algorithms for unmanned air and ground vehicles (see Figure 1). As the figure partially shows, the cameras are arranged roughly on the boundary of a rectangular area. The cameras are capable of recording at multiple framerates, with 50 Hz and 100 Hz being the options used in the majority of situations including data capture and interactive tests.

Figure 1: The MAVSeN laboratory near the time data was gathered. A non-reflective padded covering was placed on the floor after the photo was taken.

1.1.2 Data Source

The collection of data is facilitated by the Vicon Tracker application, which provides a graphical user interface to configure camera settings and define rigid patterns. Vicon Tracker does not support skeletons. Vicon DataStream SDK [1] enables programmatic access to streaming data from Vicon Tracker via C++ and C# libraries. This data can then be written to a file or reacted to in a real-time or near real-time fashion.

A glove with fifteen markers attached is used as the source of data for posture and gesture recognition, both for the generation of datasets and for the practical evaluation of developed algorithms. Four of the markers form a rigid pattern on the back of the hand to serve as identification of the hand's position and orientation and to create a local coordinate system for the remaining eleven markers. Figure 2 shows a picture of the glove with all markers visible.

The remaining eleven markers are unlabeled; they do not form part of a rigid pattern nor skeleton. A rigid pattern is infeasible because the markers are not related in any manner that could be described as rigid. All distances and angles between these markers are flexible. For a similar reason, a skeleton is also infeasible since in theory the skeleton needs fixed segment lengths between certain markers. In reality, even if some distortion is allowed in the segment lengths, a skeleton is still infeasible, or at the least impractical, due to the variance in the lengths but more so due to the inherently high rates of marker occlusion. Visibility of fingers can be blocked by other fingers or the hand itself depending upon the hand's pose and orientation. For example, the fingertips are occluded when making a fist and multiple markers may become occluded simply when the user's hand is relaxed at their side and pointing downwards. An effective skeletal model also requires a denser marker set than ours in order to capture

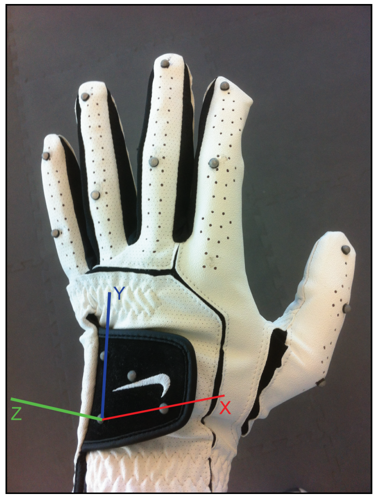

Figure 2: The glove used as the data source for all datasets. The axes of the local coordinate system based upon the rigid pattern are shown.

the 20+ degrees of freedom of the hand [2, 3] and eliminate ambiguity between similar poses. A denser marker set is not very practical in our laboratory (but also in general for large capture spaces) due to limited resolution as the cameras have a hard time discerning individual markers that are too close together.

1.2 Marker Placement and Local Coordinates

A rigid pattern of markers on the back of the glove is used to establish a local coordinate system for the hand, and 11 other markers are attached to the thumb and fingers of the glove. Three markers are attached to the thumb with one above the thumbnail and the other two on the interphalangeal and metacarpophalangeal joints (i.e. the knuckles). Two markers are attached to each finger with one above the fingernail and the other on the proximal interphalangeal joints (see Figure 2 for a detailed view).

The pattern of markers visible in Figure 2 on the back of the glove plays an important role in establishing a local coordinate system for posture and gesture recognition. Four markers comprise the pattern and are given the labels “Origin,” “XMarker,” “YMarker,” and “Extra.” Four is the minimum number of markers required to define a pattern in Vicon Tracker, although only three must be visible in order for the pattern to be detected. The axes of the local coordinate system are determined according to the pseudocode in Figure 3, which assumes that the origin is not occluded and tries to recover if any of the other markers are not visible.

procedure getLocalCoordinateAxes

Given: Origin $\mathbf{o}$, XMarker $\mathbf{x}$, YMarker $\mathbf{y}$, Extra $\mathbf{e}$
Output: local x-axis $\mathbf{x}^\star$, y-axis $\mathbf{y}^\star$, z-axis $\mathbf{z}^\star$

if XMarker is not occluded & YMarker is not occluded **then**
  $\mathbf{x}^\star = \mathbf{x} - \mathbf{o}$
  $\mathbf{y}^\star = \mathbf{y} - \mathbf{o}$
  $\mathbf{z}^\star = \mathbf{x}^\star \times \mathbf{y}^\star$
else if YMarker is not occluded **then**
  $\mathbf{y}^\star = \mathbf{y} - \mathbf{o}$
  $\mathbf{z}^\star = (\mathbf{e} - \mathbf{o}) \times \mathbf{y}^\star$
  $\mathbf{x}^\star = \mathbf{y}^\star \times \mathbf{z}^\star$
else if XMarker is not occluded **then**
  $\mathbf{x}^\star = \mathbf{x} - \mathbf{o}$
  $\mathbf{z}^\star = \mathbf{x}^\star \times (\mathbf{e} - \mathbf{o})$
  $\mathbf{y}^\star = \mathbf{z}^\star \times \mathbf{x}^\star$
end if
$\mathbf{x}^\star = \mathbf{x}^\star / \|\mathbf{x}^\star\|$
$\mathbf{y}^\star = \mathbf{y}^\star / \|\mathbf{y}^\star\|$
$\mathbf{z}^\star = \mathbf{z}^\star / \|\mathbf{z}^\star\|$

Figure 3: Pseudocode for calculating axes of the hand's local coordinate system using labeled markers.

2 Labeled Marker Dataset

This section describes the dataset of labeled markers and its associated file format.

2.1 Data Collection and Description

In contrast to the posture and gesture datasets, a single user donated this data. The purpose of this dataset is to provide the range of motion for each part of the hand/glove to which a marker is attached. This dataset is naturally limited in that it cannot apply to all potential users, but it may still serve as a basis for future algorithm development.

In order to be absolutely certain that no confusion between markers was possible, only a single unlabeled marker was attached to the glove at a time during capture. The user performed a full range of motion with each marker.

The data described here is already preprocessed. First, all markers were transformed to the local coordinate system of the record containing them using the axes given by the algorithm in Figure 3. Any record that could not be transformed was dropped. Second, each transformed marker with a norm greater than 200 millimeters was pruned. Finally, any record that contained more than one marker was dropped. Figure 4 provides a plot the processed data.

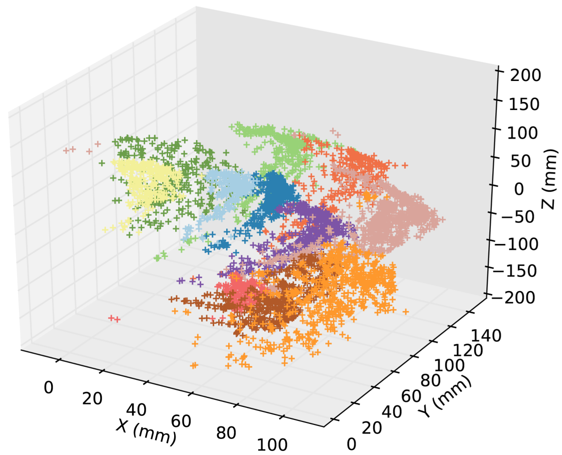

Figure 4: The labeled marker dataset after processing (i.e. in local coordinates). Some outliers for certain classes are visible.

2.2 File Format

Data is provided as a comma separated value (CSV) file. A header row provides the name of each attribute. There are no missing values. Each record corresponds to the position of a single labeled marker. The attributes are defined in the following list and are enumerated by their names.

- 'Class': Integer. The class ID of the given record. Ranges from 1 to 11 with

$$\begin{aligned}
1 &\mapsto \text{Pinky Finger (Joint)},\\
2 &\mapsto \text{Pinky Finger (Nail)},\\
3 &\mapsto \text{Ring Finger (Joint)},\\
4 &\mapsto \text{Ring Finger (Nail)},\\
5 &\mapsto \text{Middle Finger (Joint)},\\
6 &\mapsto \text{Middle Finger (Nail)},\\
7 &\mapsto \text{Pointer Finger (Joint)},\\
8 &\mapsto \text{Pointer Finger (Nail)},\\
9 &\mapsto \text{Thumb (Metacarpophalangeal Joint)},\\
10 &\mapsto \text{Thumb (Interphalangeal Joint)},\\
11 &\mapsto \text{Thumb (Nail)}.
\end{aligned}$$

- 'X': Float. The x-coordinate of the marker.

- ‘Y’: Float. The y-coordinate of the marker.
- ‘Z’: Float. The z-coordinate of the marker.

3 Posture Dataset

This section describes the posture dataset used throughout the dissertation and its associated file format. Figure 5 provides illustrations of instances within the dataset.

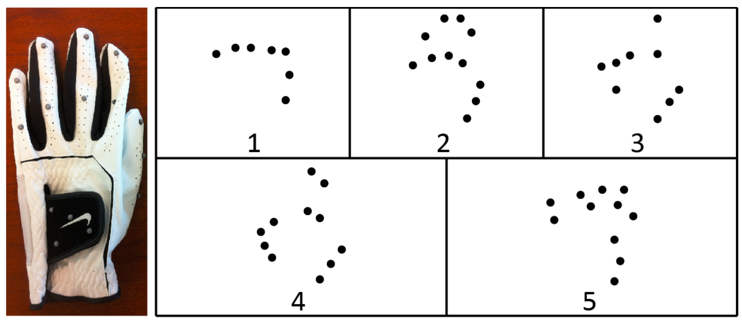

Figure 5: The glove used to capture data along with a sample from each class of posture projected onto the local XY plane. The classes are fist (1), stop (2), point with one finger (3), point with two fingers (4), and grab (5).

3.1 Data Collection and Description

We recorded 12 users performing five hand postures with markers attached to a left-handed glove (Figure 5).

The 11 markers not part of the rigid pattern were unlabeled; their positions were not explicitly tracked. Consequently, there is no *a priori* correspondence between the markers of two given records. In addition, due to the resolution of the capture volume and self-occlusion due to the orientation and configuration of the hand and fingers, many records have missing markers. Extraneous markers were also possible due to artifacts in the Vicon software’s marker reconstruction/recording process and other objects in the capture volume. As a result, the number of visible markers in a record varies considerably.

The data described here is already partially preprocessed in the following manner. The data was transformed and pruned in the same manner as the Labeled Marker Dataset. Any record that could not be transformed or contained fewer than three markers was removed. The processed data has at most 12 markers per record and at least three, which implies that at least one record has an extraneous marker. See the next subsection for more information on the attributes and file format. Unprocessed data in global coordinate is also available, but is not used anywhere in the dissertation and therefore an associated file format is not described.

Due to the manner in which data was captured, it is likely that for a given record and user there exists a near duplicate record originating from the same user. We recommend therefore to evaluate classification algorithms on a leave-one-user-out basis wherein each user is iteratively left out from training and used as a test set. One then tests the generalization of the algorithm to new users. The 'User' attribute is provided to accommodate this strategy.

This dataset may be used for a variety of tasks, the most obvious of which is posture recognition via classification. One may also attempt user identification. Alternatively, one may perform clustering (constrained or unconstrained) to discover marker distributions either as an attempt to predict marker identities or obtain statistical descriptions/visualizations of the postures.

3.2 File Format

Data is provided as a CSV file. A header row provides the name of each attribute. An initial dummy record composed entirely of zeros should be ignored (this record was included for compatibility with WEKA [4]). A question mark '?' is used to indicate a missing value. A record corresponds to a single instant or frame as recorded by the camera system. Descriptions of each attribute are provided in the following list organized by attribute name.

- 'Class': Integer. The class ID of the given record. Ranges from 1 to 5 with

$$\begin{aligned}
&1 \mapsto \text{Fist (with thumb out)},\\
&2 \mapsto \text{Stop (hand flat)},\\
&3 \mapsto \text{Point1 (point with pointer finger)},\\
&4 \mapsto \text{Point2 (point with pointer and middle fingers)},\\
&5 \mapsto \text{Grab (fingers curled as if to grab)}.
\end{aligned}$$

- 'User': Integer. The ID of the user that contributed the record. No meaning other than as an identifier.
- 'Xi': Float. The x-coordinate of the i-th unlabeled marker position. 'i' ranges from 0 to 11.
- 'Yi': Float. The y-coordinate of the i-th unlabeled marker position. 'i' ranges from 0 to 11.
- 'Zi': Float. The z-coordinate of the i-th unlabeled marker position. 'i' ranges from 0 to 11.

Each record is a set. The i-th marker of a given record does not necessarily correspond to the i-th marker of a different record. One may randomly permute the visible (i.e. not missing) markers of a given record without changing the set that the record represents. For the sake of convenience, all visible markers of a given record are given a lower index than any missing marker. A class is not guaranteed to have even a single record with all markers visible.

4 Gesture Dataset

This section describes the gesture dataset and its associated file format.

4.1 Data Collection and Description

The same 12 users of the posture dataset reprised their roles for this dataset. Each user repeated each of six gestures for approximately 30 times.

Since the pattern is not always visible and has noisy or even incorrect observations, a filter should be used to smooth the measurements of the labeled markers. Since there are many ways one could define a filter for this purpose, no processing has been performed on the data as it could bias subsequent results. As a result of no pruning or local transformations, the number of unlabeled markers (i.e. not including the pattern) can be as high as 15 due to artifacts of the capture.

There is less of an issue with duplicated gestures than with postures, but we still advise evaluating the dataset with a leave-one-user-out approach. Once again, a 'User' attribute is provided to accommodate this strategy.

4.2 File Format

Data is provided as a CSV file. Two header rows provide the name of each attribute. The first header row indicates the attributes for an entire sequence of frames that together constitute a single gesture. The beginning of a gesture is heralded by the word "Start" at the beginning of the first header. The second header indicates the attribute names for individual frames. An initial dummy sequence composed entirely of zeros is provided immediately after the headers as an example and should be ignored. Question marks are used to indicate missing values. Descriptions of each attribute are provided in the following list organized by attribute name.

- 'Class': Integer. The class ID of the given record. Ranges from 1 to 6 with

$$\begin{aligned}
1 &\mapsto \text{Click (or poke with pointer finger)},\\
2 &\mapsto \text{SwipeLeft (casual backhand as if swiping away)},\\
3 &\mapsto \text{SwipeRight (opposite motion of SwipeLeft)},\\
4 &\mapsto \text{TurnGrab (same as grab, but with left-handed}\\
&\quad\ \text{rotation about forearm axis)},\\
5 &\mapsto \text{Grab (hand closes with fingers outstretched)},\\
6 &\mapsto \text{Release (opposite motion of grab)}.
\end{aligned}$$

- 'User': Integer. The ID of the user that contributed the record. No meaning other than as an identifier.
- 'Origin-X': Float. The x-coordinate of the origin marker of the rigid pattern.
- 'Origin-Y': Float. The y-coordinate of the origin marker of the rigid pattern.
- 'Origin-Z': Float. The z-coordinate of the origin marker of the rigid pattern.
- 'XMarker-X': Float. The x-coordinate of the X-axis marker of the rigid pattern.
- 'XMarker-Y': Float. The y-coordinate of the X-axis marker of the rigid pattern.
- 'XMarker-Z': Float. The z-coordinate of the X-axis marker of the rigid pattern.
- 'YMarker-X': Float. The x-coordinate of the Y-axis marker of the rigid pattern.
- 'YMarker-Y': Float. The y-coordinate of the Y-axis marker of the rigid pattern.
- 'YMarker-Z': Float. The z-coordinate of the Y-axis marker of the rigid pattern.
- 'Extra-X': Float. The x-coordinate of the extra marker of the rigid pattern.
- 'Extra-Y': Float. The y-coordinate of the extra marker of the rigid pattern.
- 'Extra-Z': Float. The z-coordinate of the extra marker of the rigid pattern.

- 'Xi': Float. The x-coordinate of the *i*-th unlabeled marker position. 'i' ranges from 0 to 15.
- 'Yi': Float. The y-coordinate of the *i*-th unlabeled marker position. 'i' ranges from 0 to 15.
- 'Zi': Float. The z-coordinate of the *i*-th unlabeled marker position. 'i' ranges from 0 to 15.

Each record is a set in a sequence of sets. The *i*-th marker of a given record does not necessarily correspond to the *i*-th marker of a different record. One may randomly permute the visible (i.e. not missing) markers of a given record without changing the set that the record represents. For the sake of convenience, all visible markers of a given record are given a lower index than any missing marker. A class is not guaranteed to have even a single record with all markers visible.

References

[1] Vicon datastream sdk developer's manual, Jan. 2013.

[2] S. Alexanderson, C. O'Sullivan, and J. Beskow. Robust online motion capture labeling of finger markers. In *Proceedings of the 9th International Conference on Motion in Games*, MIG '16, pages 7–13, New York, NY, USA, 2016. ACM.

[3] L. Dipietro, A. M. Sabatini, and P. Dario. A survey of glove-based systems and their applications. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(4):461–482, July 2008.

[4] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, Nov. 2009.