3
Hand gestures for manipulation

The previous chapter demonstrated how we can parametrize individual parts of the human hand, using biomechanical structures with various total numbers of degrees of freedom. Given this data (the trajectory of hand and its parts in space and time), we can recognize gestures made by the hand.

3.1
One-handed and gestures in general

As per Figure 3.1 adopted from [9], any hand movement can be classified as 1) gesture or 2) unintentional movement [10], [11], [9].

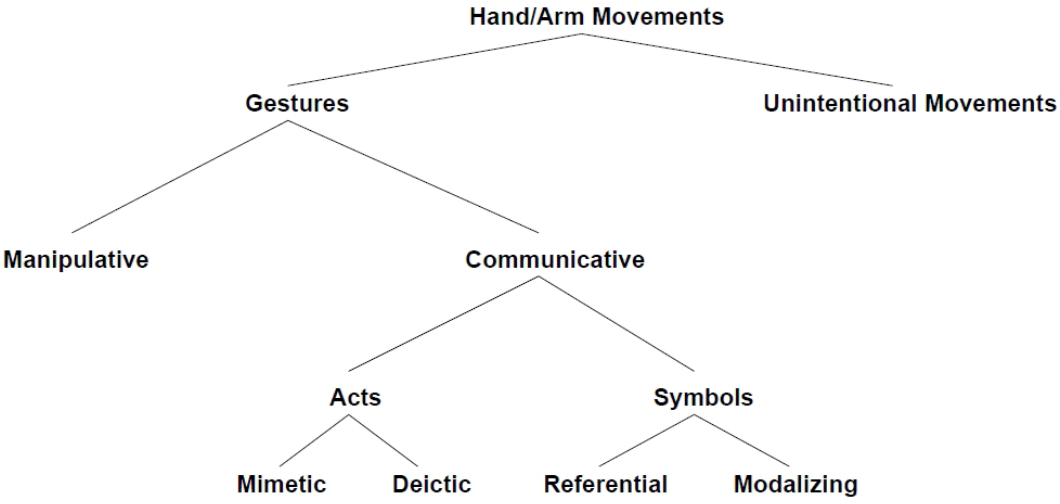

Figure 3.1: Taxonomy of gestures. In this work mostly the manipulative gestures (see extreme left of the figure) will be considered

There are two major classes of gestures:

1. **Manipulative gestures** — in this work, we are mainly interested in these gestures. Manipulative gestures act directly on objects in the real or virtual environment, like grabbing, moving, touching, rotating and stretching an object. For example [11], a pianist‘s hand movements are meant to touch the piano keys. Differently from an orchestral conductor’s hand motions, the pianist’s hand and finger motions are not meant to

communicate with anyone, but to simply touch (*manipulate*) the piano keys in order to produce music.

Some manipulative gestures are amenable to be processed by computer vision techniques, in order to manipulate 3D virtual objects. For example, a closed fist can be interpreted as a grabbing of a virtual object.

2. **Communicative gestures** — meant for visual interpretation, not for acting on objects. These gestures try to convey a message, an information. For example, an orchestral conductor's hand motions are intended to communicate temporal, affective and interpretative information to the orchestra, NOT to act on an object.

   Communicative gestures are further subdivided into (see Figure 3.1):

   - **Acts** — the movements performed relate directly to the intended interpretation.
     - **Mimetic gestures** — imitate some actions and can be considered pantomimes. These are characterized by their "iconicity". For example, a smoker going through the motion of "lighting-up" with a cigarette in his mouth indicates that he needs a light. Such gestures are usually generated on-the-fly without predetermined convention. The more novel the pantomime, the more exaggerated the motion would be to convey its intent.
     - **Deictic gestures** — point at something, and as such are very interesting for HCI. Subdivided into: **specific** — to select a particular object or location, **generic** — elicit the identity of a class of object by picking one of its members, and **metonymic** — when pointing at an object in order to signify some entity related to it.
   - **Symbols** — a type of motion short-hand.
     - **Referential gestures** — designated objects or concepts. For example, circular motion of index finger may be a referent for a wheel. Or, rubbing the index finger and the thumb in a circular fashion, is referential to money.
     - **Modalizing gestures** — serve in conjuction with some other means of communication (for example speech) to indicate the opinion of the communicator. For example, at a party, one might say to another, "Have you seen your husband?" (holding her hands apart to indicate that he is overweight). The resulting chuckle would not be understandable if one listened only to an audio transcript of the exchange.

3.2 Two-handed gestures

The seminal work for two-handed gestures is Guiard's work [12]. Guiard classifies manual tasks into the following three categories:

1. **Unimanual tasks** — to perform these tasks only one hand is necessary. Examples include combing one's hair, or dropping an object.

2. **Bimanual tasks** — two hands are necessary to perform these tasks. Can be divided into:

   - **Bimanual symmetric tasks** — both hands have an equal role in performing an activity/task which can be either 1) in the same phase (for example, rowing) or 2) out of phase (for example, climbing a mountain, or boxing).
   - **Bimanual asymmetric tasks** — for these tasks, in right-handed people, motion produced by the right hand tends to be articulated with motion produced by the left. A complex coordination between both hands is required, as in for example playing a guitar, or writing something on a paper with a pen.

Section 6.4 on page 61 shows how:

- manipulation operations `OP_SELECT`, `OP_DESELECT`and `OP_TRANSLATE`were implemented in this work, in order to perform *unimanual* tasks of selection, deselection, and translation of virtual 3D objects.
- manipulation operation `OP_SCALE`was implemented in this work in order to perform *bimanual-symmetric* task of scaling virtual 3D objects.
- manipulation operation `OP_ROTATE`was implemented in this work in order to perform *bimanual-asymmetric* task of rotating virtual 3D objects.

3.3 Modeling of hand gestures

In this section, the theoretical background on hand gesture modeling (spatial and temporal) will be given, based on exposition in [9]. Hand gesture modeling can be:

- **spatial**, and
- **temporal**.

3.3.1 Spatial modeling of gestures

Hand and arm movements trace trajectories in 3D space, therefore an useful gesture model must include a formal parametrization of the paths that all bones of the hand (finger bones, palm bones ...) and the arm (humerus, ulma, radius ...) trace in the 3D space.

Definition 1 *A complete gesture model for HCI is one whose parameters belong to the parameter space S for one-handed gestures defined as:*

$$S = \{\vec{x} : \textit{position of all hand and arm joints, and fingertips, in 3D space}\}$$

Note that the dimensionality of S is relatively high (at least $23 \times 3 = 69$), because we have 23 hand and arm joints, and fingertips, for one hand.

Definition 2 *Let $\vec{h}(t) \in S$ be a vector that describes the hand pose within a 3D Euclidean space at time t in the parameter space S. A* **hand gesture** *is then represented by a trajectory in the parameter space S over a suitably defined time interval $I = (a, b) \subset R$.*

The time interval I is also called "gesture interval".

3.3.2 Temporal modeling of gestures

Human gestures happen in time. In order to differentiate gestures from unintentional hand and arm movements, we have to determine the gesture interval I. Kendon [13] calls this gesture interval I a "gesture phrase", consisting of the following three phases:

1. **Preparation** — a preparatory movement that sets the hand in motion from some resting position.

2. **Nucleus (Peak, Stroke)** — has some definite form and enhanced dynamic qualities.

3. **Retraction** — here hand either 1) returns to the resting position or 2) repositions for the new gesture phase.

The only exception to this "preparation-nucleus-retraction" rule are the so-called "beats" that are related to the rhythmic structure of the speech.

Definition 3 *The following set of rules determines the temporal segmentation of gestures:*

1. *Gesture interval consists of three phases: preparation, stroke and retraction.*

2. *Hand pose during the stroke follows a classifiable path in the parameter space S.*

3. *Gestures are confined to a specified spatial volume (workspace).*

4. *Repetitive hand movements are gestures.*

5. *Manipulative hand gestures have longer gesture interval lengths than communicative gestures.*

3.4 Hand gesture recognition

Gesture recognition attempts to classify the trajectory of arm and hand in the parameter space S as a member of some meaningful subset of the parameter subset S. See Figure 3.2 adopted from [9].

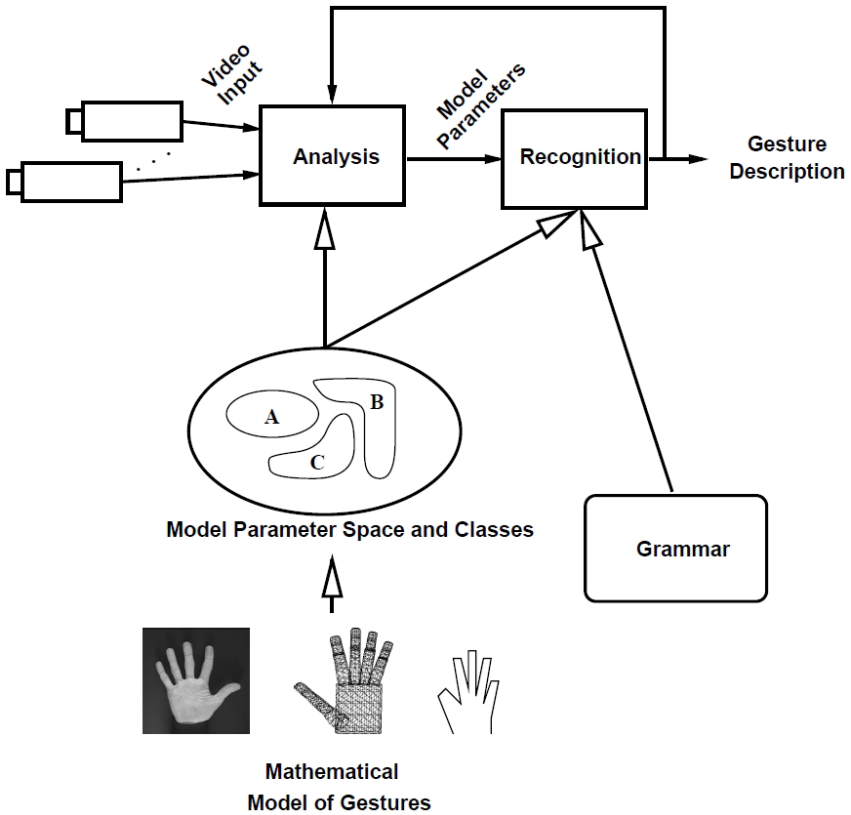

Figure 3.2: Gesture interpretation. The Recognition phase has the hand pose (Model Parameters), the database of all defined gestures (classes of trajectories) and a Grammar (serving to influence the gesture recognition depending on the current working context) as input parameters

Relevant classification methods include:

– **Hidden Markov models** — here a gesture being modelled is assumed to be a Markov process with unknown parameters, and the goal then is to determine these unknown parameters from the estimated hand pose.

– **K-means algorithm** — this algorithm classifies input gestures into clusters, where each cluster is defined by its object attributes.

– **Neural networks** — here a network of artificial neurons, which has previously been trained, tries to classify a hand trajectory in S as a gesture.