

Open Sound Controller interface

The format of the Open Sound Controller (OSC) packets is very similar to the custom packet format described above:

```
gesture_event, siiiisffff, tstamp, id, t, r, posture, x, y, s, a
```

Note that the OSC identifier is `gesture_event` and that the cryptic `siiiisffff` encodes the type information for all arguments: the following four arguments are integers, posture is a string argument, and the last four arguments are float numbers.

4.2.9 The vision conductor configuration file

For the HandVu application programmer and user who desires more control over the interface operation, the vision modules' main settings are stored in and read from a configuration file. This can be conveniently modified to fit the specific needs. Due to the orchestrating nature of the settings, we termed it a “vision conductor” file. We will briefly describe its format and refer to the respective places in the dissertation that cover the details. The following is a typical example of a conductor configuration file.

HandVu VisionConductor file, version 1.5

```
camera calibration: -
#camera calibration: config/FireFly4mm_calib.txt

camera exposure: software
#camera exposure: camera

detection params: coverage 0.3, duration 0, radius 10.0

tracking params: num_f 30, min_f 10, win_w 7, win_h 7, \
  min_dist 3.0, max_err 400
tracking style: OPTICAL_FLOW_COLORFLOCK
#tracking style: CAMSHIFT_HSV
#tracking style: CAMSHIFT_LEARNED

recognition params: max_scan_width 0.4, max_scan_height 0.6

1 detection cascades
```

```

config/closed_30x20.cascade
area: left 0.6, top .2, right 0.94, bottom .84
params scaling: start 1.0, stop 8.0, inc_factor 1.2
params misc: translation_inc_x 2, translation_inc_y 3, \
  post_process 1

0 tracking cascades

1 recognition cascades
config/all_hands_combined.cascade
area: left 0.47, top .2, right 0.94, bottom .84
params scaling: start 1.0, stop 8.0, inc_factor 1.2
params misc: translation_inc_x 2, translation_inc_y 3, \
  post_process 0

7 masks
config/Lpalm.mask
config/Lback.mask
config/sidepoint.mask
config/closed.mask
config/open.mask
config/victory.mask
config/closed_30x20.mask

```

The backslash `\` at line endings in the above printout indicates that there must not be a line break in the actual configuration file. All configuration settings must be present in the order shown above. Blank lines and comment lines, prefixed with a pound `#`, are ignored.

- **camera calibration** specifies whether a correction for lens distortion is to be performed and what file holds the calibration information. See Section 4.2.4 for details. A dash `-` indicates that no calibration is desired.
- **camera exposure** can be either **camera** or **software** and specifies whether the camera's automatic exposure control is to be used or the software-based, area-selective exposure control as introduced in Section 4.2.2.
- **detection params** are three general settings pertaining to hand detection: the **coverage** specifies the relative amount of masked hand area that has to have skin color as determined with the fixed color histogram method from Section 5.7. The **duration** gives an amount in milliseconds that a hand must

be detected in every successive frame for it to be considered a match and a valid system initialization. A value of 0 prompts acceptance with only one frame. The `radius` parameter is only used for durations greater than 0 and delimits the radius in pixels in which subsequent hand detections must lie from the first one to be considered a match. The discussion in Section 5.10 explains when these settings might be helpful.

- **tracking params** are used exclusively for the Flock of Features tracking style and specify: `num_f` the target number of features that is maintained, `min_f` the minimum number of features that has to be successfully tracked from one frame to the next or tracking is considered lost, `win_w` the width of the search window for KLT features, `win_h` the window height, `min_dist` the minimum-distance flocking constraint, and `max_err` the maximum area mismatch before a KLT feature is considered lost. All units but the last are in pixels. More details about the meaning of these parameters can be found in Chapter 6.
- **tracking style** determines the method to be used for tracking a once-detected hand:
OPTICAL_FLOW_COLORFLOCK causes tracking with a Flock of Features,
CAMSHIFT_HSV with CamShift based on a fixed HSV skin color distribution,
CAMSHIFT_LEARNED with CamShift based on a color distribution learned at detection time.
Again, please see Chapter 6 for more.
- **recognition params** limit the maximum size of the area that is scanned for hand postures during tracking to the width and height specified through `max_scan_width` and `max_scan_height`, relative to the video size.
- **n detection cascades** is a list of length `n` of detector cascades and their detection parameters. The first line of a list entry points to a file that describes a detector cascade. In addition to all weak classifiers, each cascade file contains a textual identifier (a fanned detector contains multiple identifiers). This name is used for associating the correct masks (probability maps) and giving detected appearances a name, for example, for reporting detected postures (see Section 4.2.8). Specifics about the detection method and cascades can be found in Chapter 5. The remaining three lines in a list entry are described in the following.
- **area** defines a rectangular region that is to be scanned with the respective cascade, in relative coordinates.

- **params scaling** specifies the scales at which the respective cascade is to be scanned across the area. For example, a **start** scale of 1.0 is the minimum template resolution, a **stop** scale of 8.0 says to increase the scale incrementally while it is smaller than eight times the template resolution, and an **inc_factor** of 1.2 asks for scale increase steps of 20% over the previous size.
- **params misc** specify the translation of the cascade during scanning in pixel-sized increments, both in the horizontal and the vertical dimension. The increments are for the smallest scale and scaled with the cascade size thereafter. **post_process** can be 0 or 1, where 1 means that all intersecting matches found in a single frame are to be combined into a single rectangular area as suggested by Viola and Jones in [180], and 0 causes all individual matches to be reported. See Section 2.3.8 for more details on detector scanning.
- **n tracking cascades** is currently not used and **n** must be 0.
- **n recognition cascades** are the cascades used for recognizing different postures as described in Chapter 7. The list of cascades has the same format as for the detection cascades, but the **area** line is ignored, only **params scaling** and **params misc** are used.
- **n masks** are the names of **n** files that contain the hand pixel probability maps as described in Section 5.8. Each of these files contains a textual posture identifier that is used to match a map to its cascade and a template-sized matrix of probabilities for the respective pixel to belong to the hand area.

A file that follows these specifications can be read with the **LoadConductor** API call. Upon successful parsing, the changes are assumed immediately.

4.3 Vision system performance

The quality and usability of any vision-based interface is determined by four main aspects of the computer vision methods: speed, accuracy, precision, and robustness. In addition, usability of the application interface is of course an important factor, but this shall not be considered here. While the main results of user studies and runtime data are reported in the following chapters, this section summarizes the performance as it pertains to the entire vision system.