User Manual for 3D Semiautomatic Annotation Tool

The following steps are necessary to create a dataset with our proposed tool:

1. Record dataset using depth camera
2. Do the detection
3. If necessary, apply low-pass filter on 3D detection locations
4. Check if hand size fits the cropping window
5. Run reference frame selection
6. Annotate reference frames
7. Run annotation propagation and global optimization

For all these steps, see the included files:

- main_blender_semiauto.py – Demo for running or method
- main_labeling_detect.py – GUI demo for detection
- main_labeling_pose.py – GUI demo for pose annotation

on how to use our code.

There is a brief explanation of the provided GUI on the following pages.

Detection Tool

The GUI used for the detection tool is shown in Fig. 1. The left side shows the color-coded depth image (darker closer to camera, brighter further away), and the right side shows depth profiles along x- and y-axis. Using the right graphs you can check the depth of the annotation.

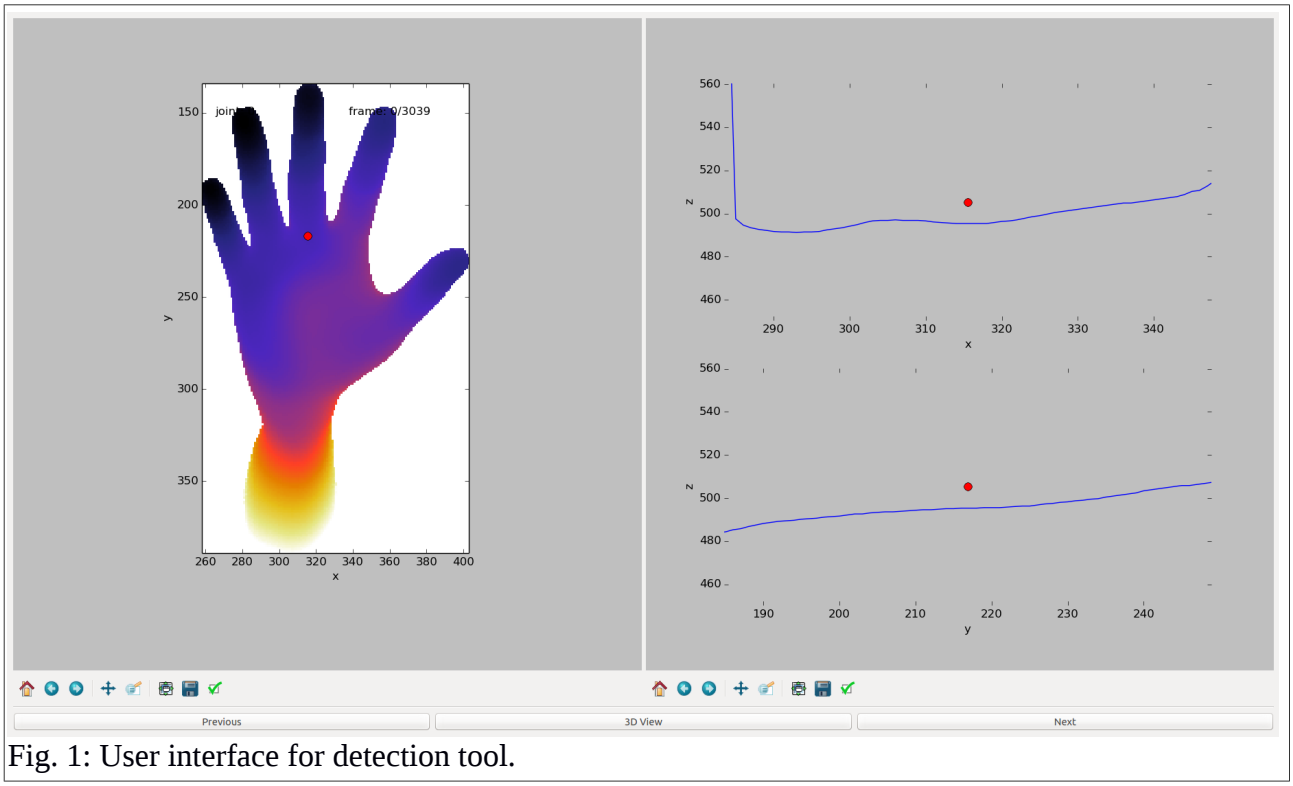

Fig. 1: User interface for detection tool.

The easiest joint for annotation is the MCP joint of the middle finger. Click on the desired location or press the left mouse button to drag the position marker in 2D. Use the keys “+” and “-” to move the joint in depth. The depth is automatically set behind the depth of the surface when dragged. Thus, if the joint is visible, there is no need to adjust the depth. With the “3D view” you can check the actual 3D location of the annotation. The 3D annotation should be located behind the surface, directly at the center of the joint.

Keyboard shortcuts:
- Press “+” to move joint further away from the camera
- Press “-” to move joint closer to the camera
- Press “3” to open 3D view, use mouse to rotate the hand
- Press “n” to go to next frame
- Press “p” to go to previous frame
- Press “q” to quit program
- Press “s” to save current state (done automatically with “n” or “p”)

Pose Tool

The GUI for annotating the pose, i.e. hand articulation, is shown in Fig. 2. Again, the 2D reprojections of the joints can be set by dragging the location markers. Each joint has to be set to its physically most plausible position. Further, visibility has to be defined for each joint, which can be toggled by pressing “v”. Joints that are not visible, i.e. occluded, must be specified and the marker changes to a triangle. The bottom row shows the interface for setting the depth order constraint (z-order). The depth order of several pairs of joints can be enabled, and set to a specific order. This resolves flipping ambiguities. Therefore, each joint has a number shown in the plot and also in the bottom row. The depth order constraint can be set for parent joints, which are indicated by the numbers. To enable a constraint check the check box and select the order of the joints. “X < Y” means that joint X is closer to the camera than joint Y, and “X > Y” indicates the opposite. The result of the inferred 3D joint locations can be shown by pressing the “Solve” button, which starts the optimization and shows a 3D view of the result. If the result does not look accurate, move the joint locations, check visibility, or add a depth order constraints at the bottom bar.
The left window can be adjusted by clicking the blue plus sign below. Hold and drag the left mouse button to move the image, and hold and drag the right mouse button to zoom. The right side of the window can be ignored.

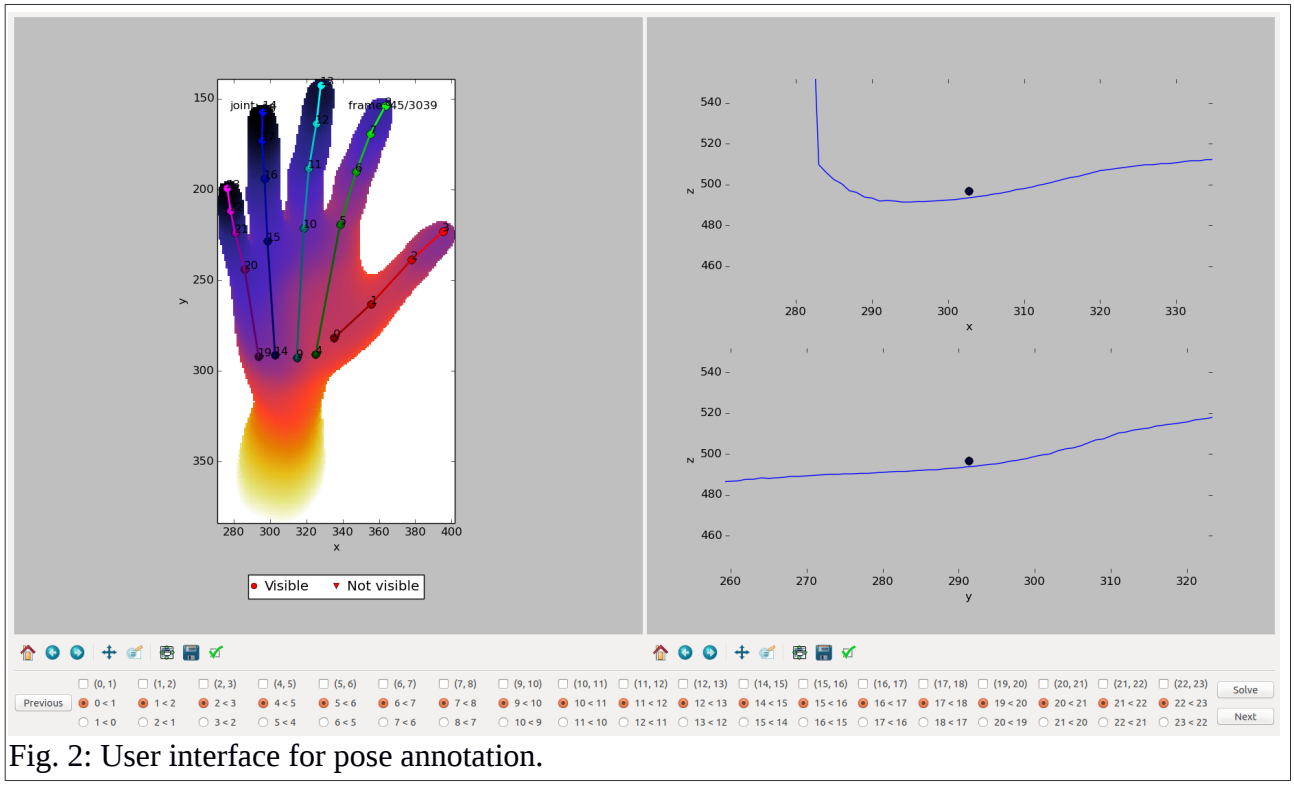

Fig. 2: User interface for pose annotation.

Keyboard shortcuts:
- Press “+” to move joint further away from the camera
- Press “-” to move joint closer to the camera
- Press “3” to open 3D view
- Press “v” to toggle visibility
- Press “d” to reset pose to default
- Press “a” to shift joints to current hand center
- Press “n” to go to next frame
- Press “p” to go to previous frame

- Press “q” to quit program
- Press “s” to save current state (done automatically with “n” or “p”)
- Press “m” to measure bone length of current hand, useful for setting up the constraints. Therefore annotate a flat hand and also adjust the depth of the joints (+/-), and then measure the bone length using “m”.

There are different choices of the model used for annotation, as shown in Fig. 3 and 4. Fig. 3 shows the used hand model for the Blender dataset. Fig. 4 shows a different configuration that was used for the MSRA dataset, which is easier to annotate in practice. We also use this annotation for our dataset. Note, that the colors indicate the different fingers and must be the same for all frames.

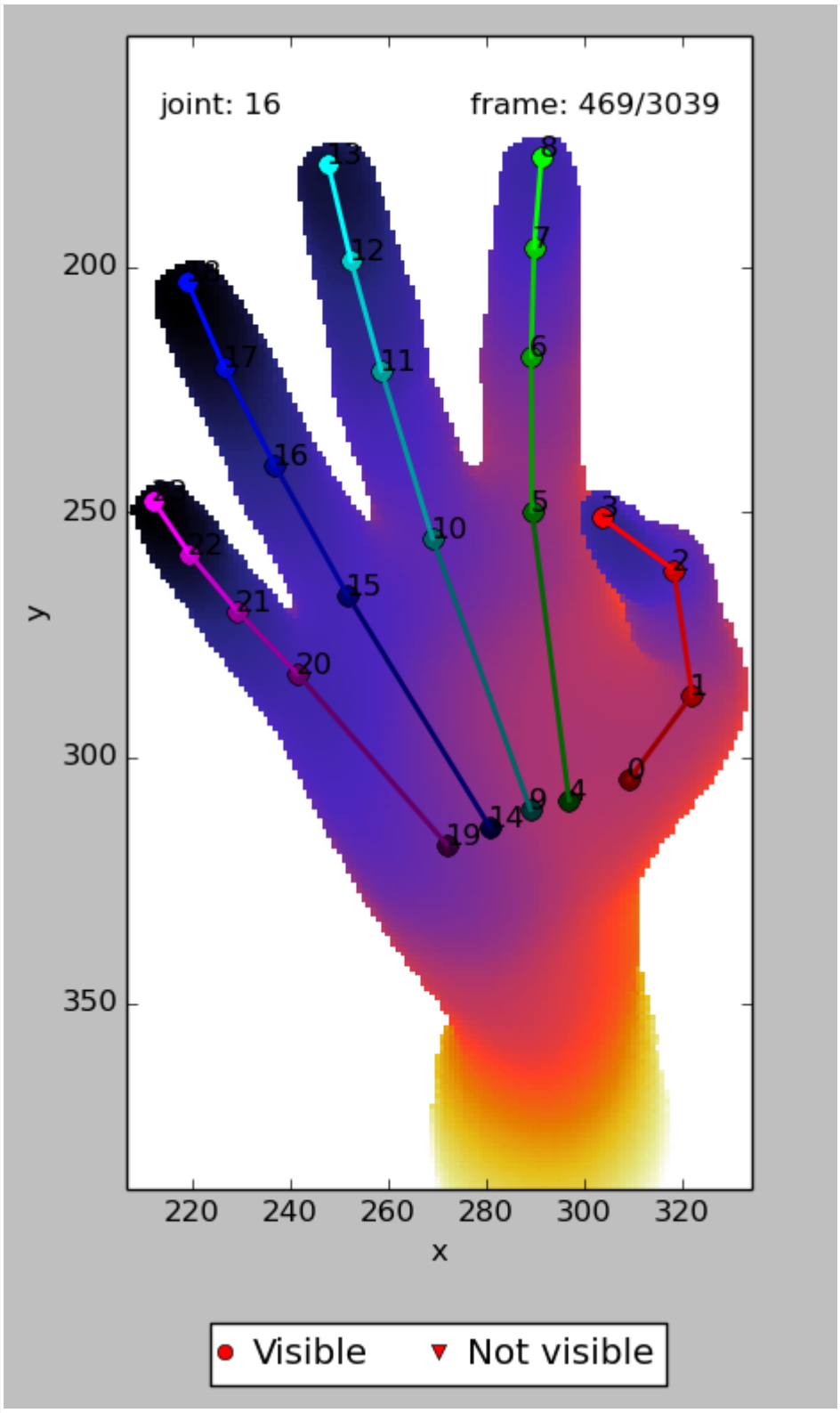

Fig. 3: Ground truth pose for Blender dataset.

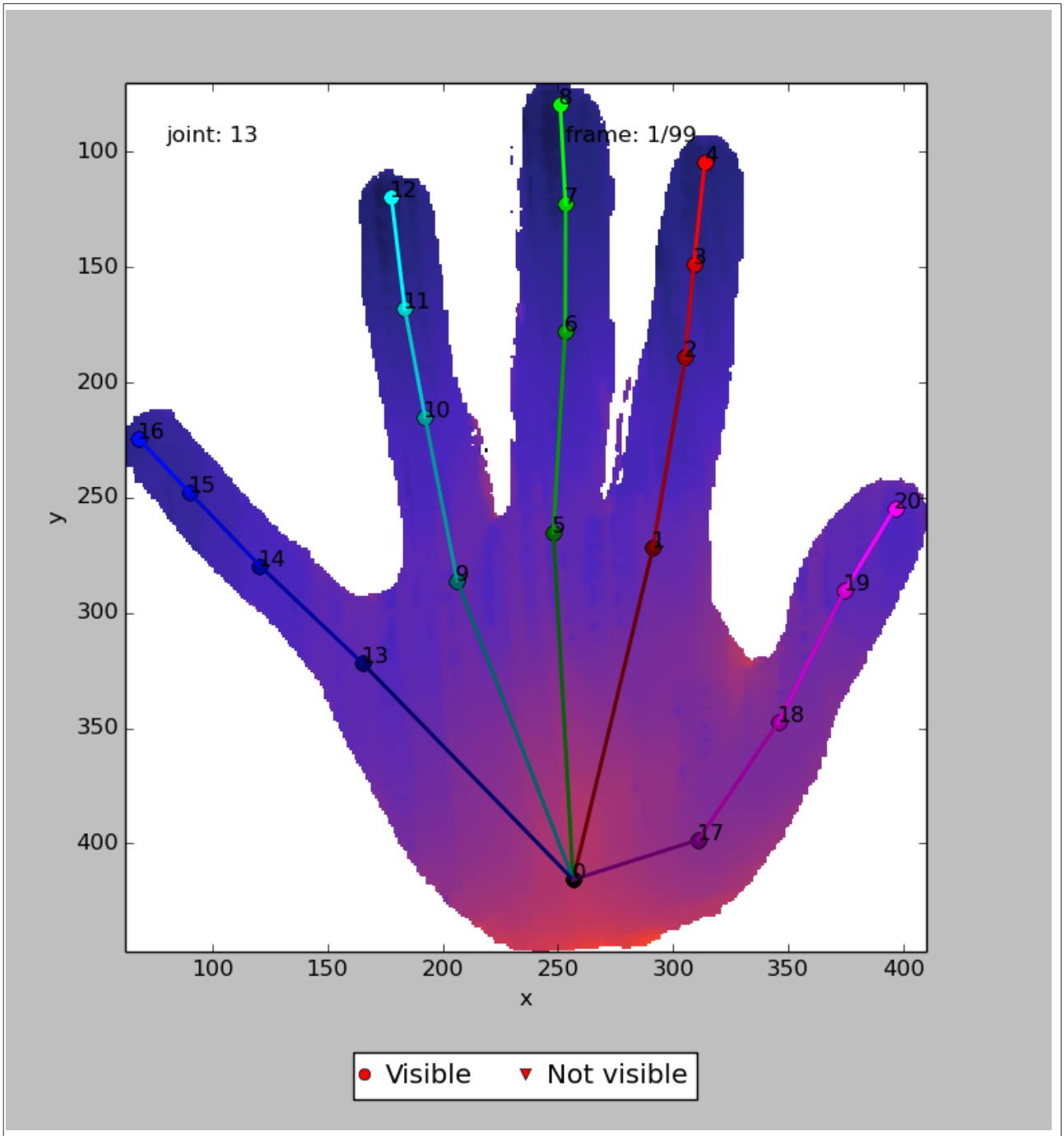

Fig. 4: Ground truth pose for our dataset.