APPENDIX

This appendix provides additional context on the design decisions given in the main paper text. In particular, we motivate our choice of horizontal plane and user-centered cylinder shapes to be the most appropriate for the given 3D control task. We also describe specific details of the experimental setup that may help one to reproduce our results.

A. Alternative workspace shapes

In this work, in addition to cylinder and horizontal plane, we also considered *sphere* and *vertical plane* as candidate workspace shapes; however, we rejected these options as they do not pass the following important usability constraints:

1) the ability to control the robot's height independently from the robot's horizontal position;
2) the ability to control the robot's horizontal position independently from its height;
3) the legibility of the active workspace configuration, i.e. the ability for the operator to know the active workspace configuration by only knowing: a) the active workspace shape, and b) the current position of the robot with respect to the operator himself.

Below we describe these alternative shapes and motivate our decision in more detail.

a) Sphere S_{sphere}: centered at the user's head (p_o); the sphere radius is the free parameter. Operating in this workspace roughly corresponds to the user holding a rod in their hand, with the robot affixed at its extremity. This mode of control is particularly effective for user-centered tasks like taking drone selfies.

b) Vertical plane $S_{v\text{-}plane}$: parallel to z-axis and perpendicular to the line connecting the user's head (p_o) and the robot; this shape therefore has *two* free parameters, which are set when the operator switches to it: the rotation of the plane around the z-axis, and the distance of the plane from the operator. This option allows one to change the robot's height without limitations and move it on a straight line with respect to the ground. This modality loosely resembles image plane object manipulation [1].

In order to reach any point in 3D space, the user has to switch between at least two different surfaces. To minimize the complexity of the system, we limit the user to toggle between *exactly two* possible shapes; after observing the experience of first-time users, we believe more than two workspace shapes would unacceptably increase the operator's cognitive load.

Among the possible choices, $\langle S_{\text{cylinder}}, S_{\text{sphere}}\rangle$ and $\langle S_{\text{v-plane}}, S_{\text{sphere}}\rangle$ pairs are discarded as they prevent controlling the robot's horizontal position independently from height. For example, if the initial position is close to the user's feet, reaching a point a few meters away can only be executed through a convoluted operation: choosing the cylindrical or vertical plane shape, raising the robot to the arbitrary height, then switching to a spherical shape and lowering the robot again while it gains distance.

Similarly, $\langle S_{\text{h-plane}}, S_{\text{sphere}}\rangle$ does not allow independent control of the robot's vertical position, which is something operators expect. Reaching a point at a considerable height in this model requires one to fly the robot farther than necessary using the plane shape, then switch to sphere to gain height while reducing horizontal distance.

The $\langle S_{\text{v-plane}}, S_{\text{h-plane}}\rangle$ combination is a reasonable alternative; however, as compared to S_{cylinder}, $S_{\text{v-plane}}$ has an important drawback: the orientation of the plane is set at the moment the workspace is switched to, and after this moment, this orientation can not be inferred by the robot position alone. We believe that this makes the system less *legible*, as the user has to remember the vertical plane's rotation; unlike $S_{\text{h-plane}}$ shape that is always parallel to the ground and easy to visualize, $S_{\text{v-plane}}$ has no such environment reference. For this reasons we reject combinations involving $S_{\text{v-plane}}$ as overly complex and non intuitive. Note that using vertical planes parallel to one of the walls of a rectangular room (or any indoor environment with few clearly defined vertical planes) would be a good alternative, but requires some knowledge of the environment, which is an assumption we do not make in this work.

We therefore determined to allow the user to toggle only between $S_{\text{h-plane}}$ and S_{cylinder} workspace shapes.

B. Experimental environment

The experiments take place in the indoor netted volume of roughly $6\,\text{m} \times 6\,\text{m} \times 2\,\text{m}$. We use an Optitrack motion capture system to enable safe closed-loop drone control, and to gather the ground truth positions of the drone and the subjects. We also use it to acquire heights of the subjects to calibrate the pointing model at the beginning of each session; for this we scale the entire human body kinematics model of an average European according to the acquired height.

Inside the flying arena we placed three targets of different heights in the following (x, y, z) configuration:

$$\begin{aligned} T_1 &= (0.6\,\text{m}, 0\,\text{m}, 0.90\,\text{m}) \\ T_2 &= (-0.6\,\text{m}, 1.2\,\text{m}, 0.53\,\text{m}) \\ T_3 &= (-0.6\,\text{m}, -0.6\,\text{m}, 0.10\,\text{m}) \end{aligned}$$

The locations where the users should stand were marked with numbers on the floor and were placed at the following (x, y) positions:

$$\begin{aligned} P_1 &= (1.2\,\text{m}, 1.2\,\text{m}) \\ P_2 &= (0\,\text{m}, -1.8\,\text{m}) \\ P_2 &= (-1.8\,\text{m}, 0\,\text{m}) \end{aligned}$$

We arranged locations $P_{1,2,3}$ and the targets $T_{1,2,3}$ such a way that the users could not fly the drone without switching between workspace shapes. For example, if the user stands at equal distances from two targets they could easily use the cylinder workspace to move the drone between them as they would lie on the same radius.

While the users were asked to move between the positions in succession, the order in which the targets were given to them was randomized.

References

[1] J. S. Pierce, A. S. Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, and M. R. Mine, “Image plane interaction techniques in 3D immersive environments,” *Proceedings of the 1997 symposium on Interactive 3D graphics - SI3D ’97*, pp. 39–ff., 1997.