This section provides precise details of how features are extracted for pairs of objects in Unity3D scenes. The code for the feature extraction process is provided in the feature extraction folder of the data collection software provided in the data archive[21] and updated versions are provided in the github repository.[49]

shortest_distance and *contact* are calculated by considering the distance of vertices of the figure, given by its mesh, to an approximation of the ground.[50] The shortest distance between F and G is taken to be the shortest distance from any vertex on the mesh of F to the approximation of G; as seen in Figure A.1.

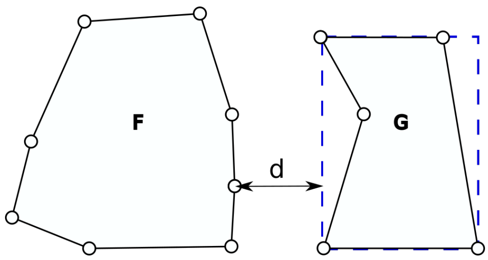


Figure A.1: Shortest distance.

The degree of contact between F and G is the number of vertices of F which are under a threshold distance (the default offset in Unity3D) to the approximation of G, divided by the total number of vertices of F. In Figure A.2 red vertices represent vertices of F which are in contact with G, *contact* is therefore $3 \div 8 = 0.375$.

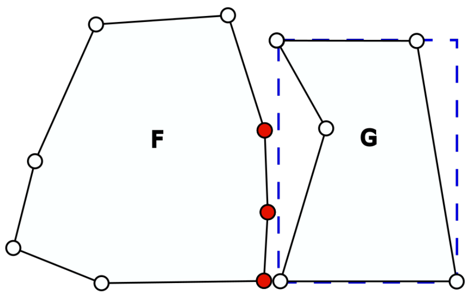


Figure A.2: Contact.

above_proportion and *below_proportion* are calculated by counting the number of

[49]https://github.com/alrichardbollans/spatial-preposition-annotation-tool-unity3d/tree/master/Unity3D%20Annotation%20Environment/Assets

[50]A collider is used to represent the ground which for simple objects is given by the object mesh, but for complex objects is approximated by a box, sphere or collection of boxes.

vertices of F which are above/below the highest/lowest point of G. In Figure A.3, the vertices above G are given in red and the vertices below G are given in blue. *above_proportion* is therefore $2 \div 8 = 0.25$ and *below_proportion* is $3 \div 8 = 0.375$.

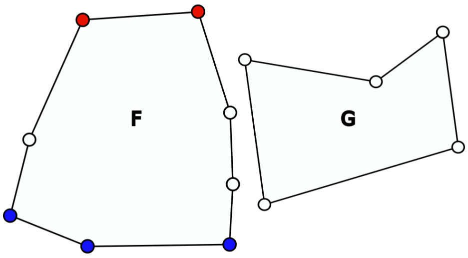


Figure A.3: Above/Below.

containment is calculated as the proportion of the axis-aligned bounding box of the figure which overlaps with the axis-aligned bounding box of the ground. In Figure A.4, *containment* is equal the volume of the purple shaded area divided by the total area of the bounding box of F.

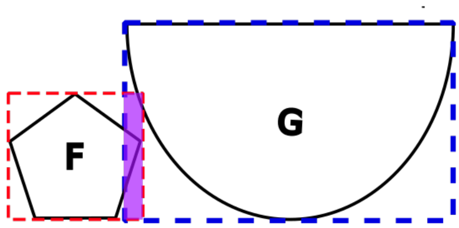


Figure A.4: Containment.

horizontal_distance calculates the horizontal distance between the centres of mass of the figure and ground, as given (in 2D) in Figure A.5.

f_covers_g aims to represent the degree to which the figure covers the ground. This is calculated by considering the degree to which the horizontal areas of the objects overlap. The greater the height separation of the figure from the ground, the larger the figure must be in order to provide effective covering. Therefore the effective area of the ground to be covered is extended (given by the blue dashed line in Figure A.6) taking into account the height separation, h: the area of G is extended on each side by $h \times \tan(5°)$. The effective overlap (given by the purple line in Figure A.6) is divided by the extended area of G in order to give the value of *f_covers_g*.

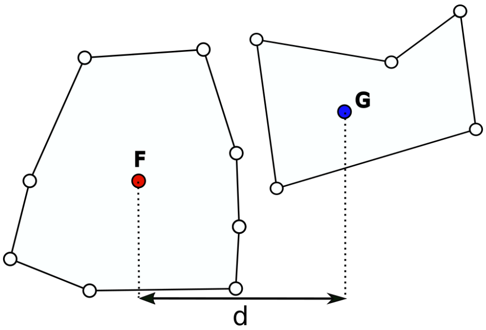


Figure A.5: Horizontal distance.

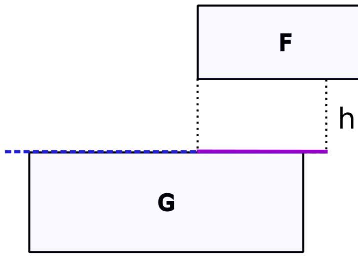
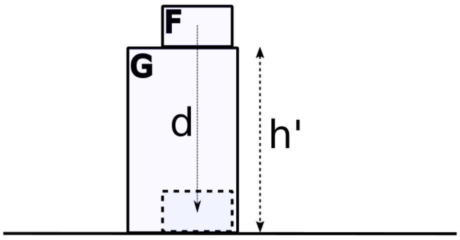


Figure A.6: F covers G.

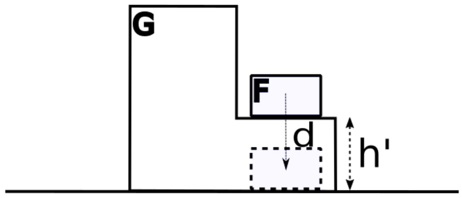
To calculate *support*, the distance fallen by the centre of mass of the figure is assessed when the ground is removed from the scene, given by d in the following diagrams. The distance in then divided by a normalising height h'.

In the canonical case shown in Figure A.7(a), h' is simply equal to the height of the ground and *support* $sim 1$. In the case shown in Figure A.7(b) which commonly occurs where an object is attached to the side of another, h' is equal to the height difference from the bottom of the figure to the bottom of the ground and *support* $sim 1$. In the case shown in Figure A.7(c) which commonly occurs where an object is leaning on another, h' is equal to the height difference from the centre of mass of the figure to the bottom of the ground and *support* is often less than 1. In all other cases not accounted for here, h' is set as the height of the ground.

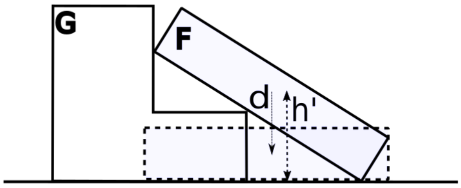
To calculate *location_control* a force is applied to the ground in a particular direction and the distance moved by the centre of mass of the figure is divided by the distance moved by the centre of mass of the ground (in the given direction). An example is given in Figure A.8, here location control is being assessed in the positive x direction and the



(a) Canonical Support: defined by the bottom of the figure being above the top of the ground.



(b) Support by attachment: defined by not being a canonical case and the bottom of the figure being above the bottom of the ground.



(c) Leaning support: defined by not being a canonical or attachment case and the centre of mass of the figure being above the bottom of the ground.

Figure A.7: Calculations of h' for support calculation.

given value is $L_{x+} = d \div n$. This is repeated for each of the other cardinal directions and the final value of *location_control* is the average, given by $\frac{L_{x+}+L_{x-}+L_{y+}+L_{y-}}{4}$.

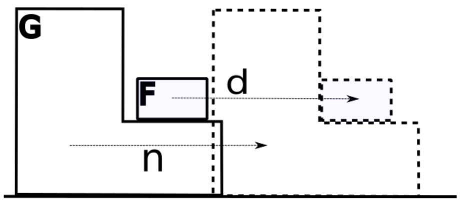


Figure A.8: Location Control.