Playing The Strengths of Synthetic Data

Carson Trego

November 2024

1 Defining Pose, Origin, Object Space

Let the projection of 3 dimensional information onto a 2 dimensional plane (as in the case of a digital photograph with a depth map) be represented by a n x m x 4 matrix, where n and m represent the pixel resolution of the image and depth map, and each pixel has normalized red, blue, green, and depth information, where 1 is the maximum value that can be detected by the camera or render, and 0 is the minimum value that can be detected by the camera or render.

Assume there exists 3 dimensional object that is perfectly rigid, opaque, and has an arbitrarily-defined point that is defined by its position relative to every unit of information in the 3 dimensional object. For example, if a point is arbitrarily defined to be in the centroid of a sphere, then the point will always be located in the position where all points on the surface of the sphere are equally far from the arbitrarily-defined point. In another example, if the object is a cone, and the point is arbitrarily defined to be the centroid of the circular face of the cone, than the arbitrarily-defined point can be located by finding the location in which the distance from the centroid of the circular face and the arbitrarily-defined point is zero, and the distance from the arbitrarily-defined point and the tip of the cone is the height of the cone. Note that while only a single point on the object, and its subsequent spherical angles and distance from the arbitrarily defined point are required to find the arbitrarily-defined point, the arbitrarily-defined point is defined relative to all of the points on the 3 dimensional object. Also assume that all transformations to the points of the object are applied to the arbitrarily-defined point, so if the object is a sphere with the arbitrarily-defined point being its centroid, and the sphere is translated 10 units up on the z-axis, then the arbitrarily-defined point will also ascend 10 points on the z-axis. With this setup, both the point and the object exist in their own 3 dimensional space that can be transformed independently of the outer environment they both exist in. For example, if the sphere is in an outer environment, e.g. a city, and the sphere is translated 10 units vertically on the z axis, then the 3 dimensional space that the object and point exist in ascend 10 points on the z axis, but the buildings and streets within the city are unchanged. This 3 dimensional space extends infinitely in 3 dimensions, and thus the object and the point can be large, finite distances from each other while still being theoretically able to deduce the same information about the point from the object.

Another way of thinking about the arbitrarily-defined point as the origin of the 3 dimensional space that the object, and only the object, occupy. We will use this definition, and refer to the arbitrarily-defined point as the "origin" and the space that the object, and only the object, occupy as the "object space". Do note: the object can exist in wider contexts while also occupying the object space, so the object and origin can exist within an outer context, such as a city or a forest, while simultaneously occupying the object space, of which none of the other objects in the city or forest exist in. The object space is a 3 dimensional space that can be overlayed on "top" of another 3 dimensional space, and transformed relative to it.

In addition to the object space having an origin, there are 3 axis, both extending in 6 directions from the origin. As with the origin of the object space, these directions are arbitrarily-defined for an object, and may be unrelated to any intuitive description of the object itself. For example, if the object is a cardboard box with a "THIS SIDE UP" arrow-symbol on it, it is intuitive to think of the origins being in the centroid of box and the z-axis increasing in the direction of the "THIS SIDE UP" arrow, however the object space could be defined differently, with the origin being millions of units away from the centroid, and the "THIS SIDE UP" arrow pointing towards the direction that is 45 degrees between the positive x-axis and the positive y-axis.

2 Limitations With RGBD Projections

In the context of pose estimation, the goal of pose estimation is to find deduce all of the information about the object space (the object information, the origin, and axis directions), relative to the camera that forms the RGBD projection, from an object alone. The full information of the object in its environment is also not fully available, and the information of the object space from the object must be determined from 4 x 2 dimensional projections of information about the object, given by the RGBD matrix.

How easily it is to determine the pose (i.e. locations of the origin of the object space and all congruence classes of both angle parameters in spherical coordinates for one of the axis of the object space), depends on the bijection of the projection and the position and orientation of the object space. All orientations and positions of the object space will produce an RGBD projection, and, assuming all projections are taken with the object space present, all projections will have an accompanied orientation and position of the object space. With this information, if an object and object space is present, then the object space and RGBD projection will necessarily have a subjective relationship.

If the relationship between the object space and the projection is also injective, and thus bijective, then any projection can be mapped to a unique position and orientation of the object space. This would mean that all RGBD projections are unique, and the position and orientation of the object space could be, theoretically, deduced from any RGBD projection

An RGBD projection of an object may not be unique. If we define an object as having a unique RGBD projection when all possible poses result in a sequence of RGBD matrices in which no matrix has the same values for all elements in the RGBD matrix, then we can easily disprove. Obviously if the object is not within view of the camera which the 3D object is projected onto, then all RGBD projections with the object not in scene will be identical (assuming the camera and background is stationary).

Even if we remove this case and only consider projections where the object is present and in "full view", a projection may not be unique. Here, we will define "full view" as when an object and camera are positioned such that the set of all 2D slices parallel to the camera, extending from the direction of the camera, and limited by the viewing angle of the camera, contain all of the 3D information of the object, and there are no external objects in the environment that obstruct the object from being seen by the camera, meaning that all points on the 3D object should have an a path to the location on the projection without crossing any other visible objects.

To show that not all objects in full view of the camera produce unique projections, we will perform a disproof by counterexample. Assume the object is a triangular prism with length L, with the origin being at the centroid of the prism, the x-axis and y-axis being parallel to the faces of the triangles, and the y-axis intersecting and extending toward the top ridge of the triangular prism, and assume the z-axis is perpendicular to the faces of the triangle, extending towards the face of the triangle, and the camera positioned i units away from the face of the object, and $i+L/2$ units from the origin. Assume the object is in full view. In this case, assuming infinite resolution and gray, opaque surface color of the prism, the projection of the object will be a set of points asymptotically forming a triangle, where centroid of the triangles is the center of each layer of the RGBD projection. If the object space, and thus the object, is rotated on the z-axis 180 degrees, then the new projection will also form a set of points that is asymptotically a triangle, with the centroid of the triangles being the center of each RGBD projection layer. In this case, both RGBD projections will have the exact same values, meaning RGBD projections are not unique, and the relationship between the location-orientation of the object space and the RGBD projection is not bijective.

In practical terms, this means that perfect pose estimation from a single RGBD projection is not necessarily possible.

3 Resolutions

While pose of an object from a single image is not necessarily possible, it is possible if there exist one equivalence class of RGBD projections that has a length of one. As stated before, it is possible for multiple poses to produce the same projection. All poses that produce the same projection can be grouped in the same "equivalence class". In the case of the triangular prism, all projections that produce the same set of 3 triangles and the same numeric values for each matrix element. For example, given a set distance from the centroid, d, 6 poses will all produce the same triangle shape in each matrix with the same values, and thus all poses exist in the same equivalence class. In the case of a perfect sphere with a the origin at the centroid, all orientation angles of the sphere have the same equivalence class.

However, if there exists at least one equivalence class with only one pose, the relationship between the projection and the pose is one to one, and the pose can be deduced from the RGBD projection.

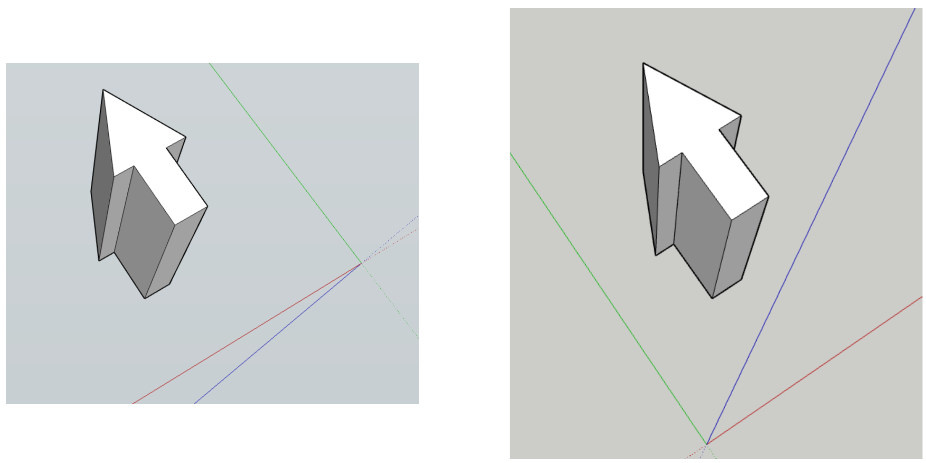

Figure 1: Two arrow prisms at different poses with nearly same appearance. Assuming a perfect angle and distance, these would be in the same equivalence classes.

In the case of objects with no unique equivalence classes, as in the case of a perfect arrow prism, useful information may still be extracted if the pose orientation modular divisor is changed. For example, it may not be important if the front face or back face of the arrow prism is showing. In this case, we can define the orientation of one of the rotation axis as being from 0 to pi instead of 0 to 2pi, as the RGBD projection of the arrow prism resets after a 180 degree rotation. In the case of a sphere, the equivalence class modular divisor for both orientation angles would be infinitely small such that all orientation angles would share the same equivalence class.

For simplicity, we will fist focus on the case where at least one equivalence class with only one pose exists.

(Note, at this point I would like to rigorously show/test if an object has this trait, but I have yet to develop a truly specific test for it, as there are infinitely many orientations in an object. My current solution is just rendering a large number of images with a clear background that uses a color least present on the object's surfaces, maximizing the difference, and then doing some quantitative comparison such as k nearest neighbors.)