

# A Comparison of Relative Position Descriptors for 3D Objects

Andrew R. Buck\*, Derek T. Anderson\*, James M. Keller\*, Robert H. Luke III<sup>†</sup>, and Grant Scott\*

\*Electrical Engineering and Computer Science (EECS) Department, University of Missouri, Columbia, MO, USA

<sup>†</sup>US Army DEVCOM C5ISR Center, Fort Belvoir, VA, USA

Email: {buckar, andersondt, kellerj, scottgs}@missouri.edu, robert.h.luke2.civ@army.mil

**Abstract**—The spatial configuration of objects in a scene is important to many applications. In particular, 3D environments constructed from point cloud observations are often used for navigation and planning with real-time requirements. In these settings, the ability to recognize and distinguish one set of objects from another may depend largely on how they are positioned with respect to each other. In this article, we explore two different approaches for describing the relative spatial relationship between two objects represented as 3D points: the histogram of forces, and a method using bounding boxes and fuzzy numbers. We use 2D axis-aligned projections of the objects to facilitate the computation of force histograms, and compare this approach to the bounding box and fuzzy number method. Our experiments are performed on the NPM3D dataset, consisting of hand-labeled point cloud objects in an outdoor street-level environment. The results highlight the strengths and weaknesses of each approach and we discuss the most appropriate applications for both.

## I. INTRODUCTION

How can the relative spatial arrangement of two objects be described? This question lies at the heart of many spatial reasoning applications and guides the ways in which we design algorithms that can understand the physical world. Real-world environments can be captured as 3D point clouds using a LiDAR scanner or depth camera, representing the locations of objects in the scene and resulting in a large volume of spatial data. As we seek to understand this data and construct spatial models, it becomes important to represent distinct objects and the ways in which they are positioned relative to each other. In this article, we investigate ways in which the spatial relationship between two 3D point clouds can be described.

Throughout this work, we consider the concept of a static 3D object that occupies some position in space and has some finite, non-zero volume. We represent such an object as a set of 3D points, typically captured from some physical sensor or a simulated depth image at a single moment in time. The points represent a portion of the object’s visible surface and may only cover a limited part of the entire underlying object. In these situations, the full extent of the object may be unknown and the spatial relationships that exist between objects need to be capable of handling such uncertainty.

There are several established methods for computing spatial relationships between 2D objects, some of which have been extended to 3D [1]. The histogram of forces (HoF) [2], [3], [4] is one such approach that has been broadly adopted by the community and has several desirable properties. It is a robust and informative descriptor that has been applied in a variety

of spatial reasoning applications [5], [6], [7], [8]. However, although 3D versions exist [9], the HoF has not been used widely for 3D applications or with point cloud objects. This may be due to the added complexity of computing forces in 3D, or the inefficiencies that arise from being unable to treat objects as solid volumes.

Large outdoor scenes can be captured dynamically by mobile sensing platforms such as unmanned aerial vehicles (UAVs) or autonomous vehicles. Often in these applications, processing power is limited due to the use of embedded hardware and real-time requirements. In these situations, it can be advantageous to represent scene objects using simplified representations such as axis-aligned bounding boxes. In [10], the authors propose constructing a spatial relationship graph (SRG) of a labeled point cloud dataset using bounding boxes and fuzzy numbers. This allows a user to run a spatial query on the scene, such as “Find all objects that are close to me and to the north.”

Although bounding box representations can satisfy the needs of many applications, their properties as relative position descriptors have not been fully compared to more established methods, such as the histogram of forces. Following a brief discussion of related work in Section II, Section III discusses how the commonly used 2D HoF methods can be applied to 3D point cloud objects by considering orthogonal 2D projections, and how they can be used to compare the spatial relationships between two pairs of 3D point cloud objects. We then introduce a descriptor based on bounding boxes in Section IV and define several different ways to compute spatial similarity. In Section V, we evaluate and compare these methods on simple 2D examples, and then examine how they perform on real-world data in Section VI. Section VII concludes with a discussion on the differences between methods and how they can best be used in practice.

## II. RELATED WORK

In this work, we are mainly concerned with spatial relationships involving *distance* and *direction*. Another commonly used concept, particularly in the area of spatial databases and GIS, is that of topological relationships [11]. In these systems, the representation of imprecise boundaries is important and queries often ask the degree to which such objects intersect. See [12] for a recent survey of fuzzy approaches in this area.

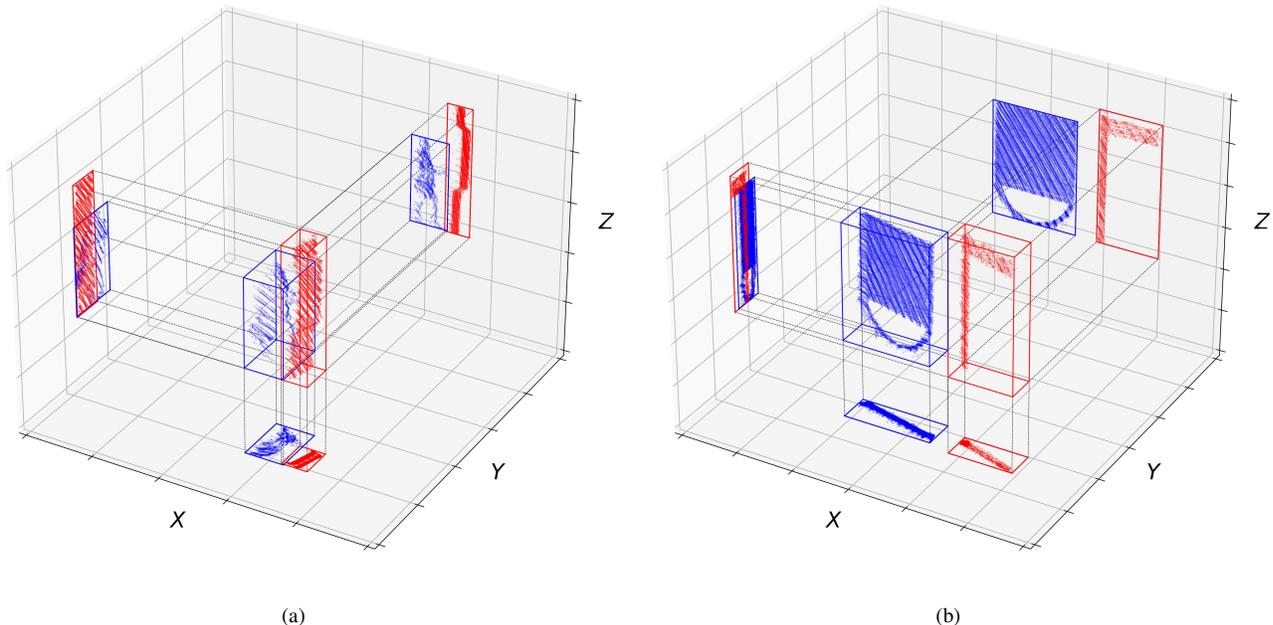


Fig. 1. Two examples of object pairs represented as 3D point clouds. (a) shows a person standing near an information sign and (b) shows two other signs in a similar spatial configuration. For each pair, we consider the red object as the referent and the blue object as the argument. Each object is projected into 2D along the three principle axes. Bounding boxes are shown for each object.

Fuzzy sets have also long been used to model spatial relationships in image processing applications [13]. Usually restricted to 2D domains, these approaches can be used to evaluate relationships such as above, beside, near, surrounds, between, etc. The histogram of forces is one such method for describing relative position that we explore in this work.

### III. HISTOGRAMS OF FORCES

#### A. Force Histograms

The histogram of forces is a well-established means of capturing the spatial relationship between two objects. To compute a force histogram, each angle  $\theta$  is considered and the strength of the proposition “Object  $A$  is in direction  $\theta$  from object  $B$ ” is evaluated. Various types of forces can be computed depending on how the distance between objects is used. Two commonly used histograms are those of constant and gravitational forces. The histogram of constant forces ( $F_0$ ) is independent of the distance between objects and the histogram of gravitational forces ( $F_2$ ) is biased toward nearby areas.

In cases where the objects overlap, the distance  $d$  between objects is reduced to zero. This can cause issues for the histogram of gravitational forces, which uses a  $1/d^2$  term. To accommodate this, a histogram of hybrid forces was proposed [2], [14] that acts as a blending of both the  $F_0$  and  $F_2$  histograms, being equivalent to  $F_2$  at far distances and resorting to  $F_0$  when the objects overlap. We refer to this hybrid histogram as  $F_{02}$  and use it in place of  $F_2$  throughout this work. Although the  $F_{02}$  histogram incorporates elements

of both the  $F_0$  and  $F_2$  histograms, we continue to use both  $F_0$  and  $F_{02}$  as they capture different properties for non-intersecting objects.

Most of the existing work done with the HoF focuses on 2D objects. While some methods have been explored to adapt the HoF to 3D [9], the 2D implementations are much more mature. Furthermore, we use the raster-based implementations of these methods as they are easier to generalize than the vector-based approaches. The next section explains how we prepare a 3D point cloud object as a set of raster image projections along the principle axes.

#### B. Axis-Aligned Projections

We use a point cloud representation for 3D objects as a matter of practicality. As an example, we consider the NPM3D dataset [15], which consists of a large 3D point cloud with labeled objects in an outdoor urban environment. Let  $\mathcal{O}$  be an object consisting of a set of points  $(x, y, z)$  in  $\mathbb{R}^3$ . The points can be projected along any of the three axes to give projections in the XY, XZ, and YZ planes. See Fig. 1 for two examples of point cloud object pairs from the NPM3D dataset and their 2D projections. The projected points can be discretized into binary raster images with a fixed pixel size as shown in Fig. 2. Here, we use the red and blue channels to show the two object images. By choosing an appropriate pixel size for the given point density (we use 5 cm for our experiments), the projected 2D images are made to capture the objects’ shapes as seen from looking down each axis.

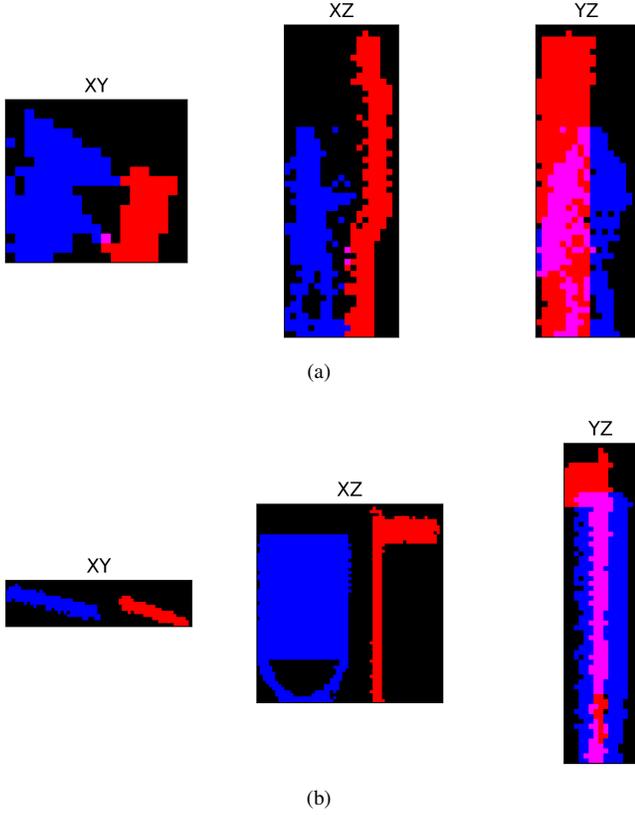


Fig. 2. The projected points from the examples in Fig. 1a (a) and Fig. 1b (b) are discretized into three raster images representing the XY, XZ, and YZ planes.

Since we are interested in the spatial relationships between objects, projections are done in pairs. To build the projected images for a given pair of objects, the image bounds are set to cover the extents of both objects. This can lead to a compact image if the objects are nearby, but may result in large (mostly empty) images if the objects are far apart. This is done to facilitate the computation of the force histograms, which require common image sizes for both objects.

For each of the three 2D projections, we compute both the histogram of constant forces ( $F_0$ ) and the histogram of hybrid forces ( $F_{02}$ ) with an angular resolution of  $2^\circ$ . This gives a total of six histograms, each being a 180-dimensional vector. We refer to this set of histograms as the HoF relative position descriptor (HoF-RPD),

$$\text{HoF-RPD}(A, B) = \begin{pmatrix} F_0^{AB_{XY}} & F_0^{AB_{XZ}} & F_0^{AB_{YZ}} \\ F_{02}^{AB_{XY}} & F_{02}^{AB_{XZ}} & F_{02}^{AB_{YZ}} \end{pmatrix} \quad (1)$$

where  $F_0^{AB_k}$  is the histogram of constant forces computed from projection  $k$  and likewise  $F_{02}^{AB_k}$  is the histogram of hybrid forces.

The force histograms computed between each pair of objects from each projection in Fig. 2 are shown in Fig. 3. The top row of the figure shows the histograms of constant forces ( $F_0$ ) for each of the three projections, and the bottom row shows the histograms of hybrid forces ( $F_{02}$ ). Note that because we

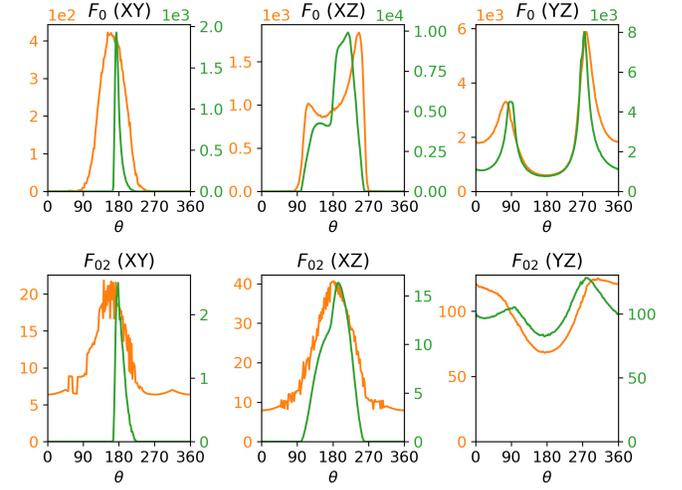


Fig. 3. The HoF relative position descriptors (HoF-RPDs) for the object pairs in Fig.1. The  $F_0$  and  $F_{02}$  histograms are computed for each of the three projected images in Fig. 2. The histograms from pair (a) are shown in orange and those from pair (b) are shown in green. The horizontal axis shows the angle  $\theta$  in degrees, and the vertical scales of each histogram are normalized to aid in visual comparison.

often encounter overlapping objects in the 2D projections, we choose to use the hybrid histogram over the histogram of gravitational forces, which can lead to unpredictable results when objects overlap. The histograms from both pairs are plotted on the same axes and normalized to aid in visual comparison.

### C. Force Histogram Similarity

We now consider how to evaluate the similarity of two HoF-RPDs. The HoF provides a rich feature descriptor that captures much of the nuance between different object shapes and positions. In general, however, we would like to compute a similarity measure  $S_{\text{HoF}}(A, B, A', B') \in [0, 1]$  that represents how similar the object pairs  $(A, B)$  and  $(A', B')$  are to each other. There are many ways this could be defined, all with variations on what aspects of the spatial relationship are most important.

In this work, we consider the following three histogram similarity measures used in [4],

$$\mu_T(h_1, h_2) = \frac{\sum_{\theta} \min(h_1(\theta), h_2(\theta))}{\sum_{\theta} \max(h_1(\theta), h_2(\theta))}, \quad (2)$$

$$\mu_P(h_1, h_2) = 1 - \frac{\sum_{\theta} |h_1(\theta) - h_2(\theta)|}{\sum_{\theta} |h_1(\theta) + h_2(\theta)|}, \quad (3)$$

$$\mu_C(h_1, h_2) = \frac{\sum_{\theta} h_1(\theta)h_2(\theta)}{\sqrt{\sum_{\theta} h_1^2(\theta)}\sqrt{\sum_{\theta} h_2^2(\theta)}}, \quad (4)$$

where  $h_1$  and  $h_2$  are force histograms and  $h_i(\theta)$  is the value in direction  $\theta$ .  $\mu_T$  is a Tversky index (Tanimoto coefficient),  $\mu_P$  is a Pappis' measure [16], and  $\mu_C$  is the normalized cross-correlation. All three methods produce values in the range  $[0, 1]$  with 0 representing no similarity and 1 representing

TABLE I  
FORCE HISTOGRAM SIMILARITIES

	$F_0$			$F_{02}$		
	XY	XZ	YZ	XY	XZ	YZ
$\mu_T$	0.236	0.201	0.722	0.016	0.207	0.868
$\mu_P$	0.381	0.334	0.839	0.031	0.343	0.929
$\mu_C$	0.548	0.859	0.948	0.482	0.899	0.989

TABLE II  
OBJECT PAIR SIMILARITIES USING THE HoF-RPD

	$s^{XY}$	$s^{XZ}$	$s^{YZ}$	$S_{\min}$	$S_{\text{mean}}$
$\mu_T$	0.126	0.204	0.795	0.126	0.375
$\mu_P$	0.206	0.338	0.884	0.206	0.476
$\mu_C$	0.515	0.879	0.968	0.515	0.787

equivalence. The  $\mu_C$  measure also has the property of being independent to the scale of each histogram.

To compute an overall similarity score for two object pairs  $(A, B)$  and  $(A', B')$ , the histogram similarities for each force type and projection need to be aggregated. Because the histograms of constant and gravitational (hybrid) forces are considered to be of equal importance, their similarity scores are averaged together. For a given histogram similarity measure  $\mu$  and projection  $k$ , the similarity is computed as

$$s_{\mu}^k(A, B, A', B') = \frac{\mu(F_0^{AB_k}, F_0^{A'B'_k})}{2} + \frac{\mu(F_{02}^{AB_k}, F_{02}^{A'B'_k})}{2}. \quad (5)$$

Across the three projections, we then consider both the minimum and mean as two ways to aggregate the combined scores. The minimum provides a pessimistic view of the overall similarity and the mean gives a more balanced approach. Other methods such as the maximum or an OWA operator could be used but are not considered here. Formally, we have

$$S_{\text{HoF}, \min, \mu}(A, B, A', B') = \min(s_{\mu}^{XY}, s_{\mu}^{XZ}, s_{\mu}^{YZ}), \quad (6)$$

and

$$S_{\text{HoF}, \text{mean}, \mu}(A, B, A', B') = \frac{1}{3}(s_{\mu}^{XY} + s_{\mu}^{XZ} + s_{\mu}^{YZ}). \quad (7)$$

Taken with the three different histogram similarity measures, a total of six different measures are defined using the HoF to compare the spatial similarity of two object pairs. Table I shows the histogram similarities computed for the HoF-RPDs in Fig. 3. Each row of the table reports a different histogram similarity measure computed for each of the six histogram pairs. Table II shows the averaged similarities for each projection and the two overall similarity measures. We note that for this single example, the range of similarity scores varies from 0.126 using  $S_{\text{HoF}, \min, \mu_T}$  to 0.787 using  $S_{\text{HoF}, \text{mean}, \mu_C}$ .

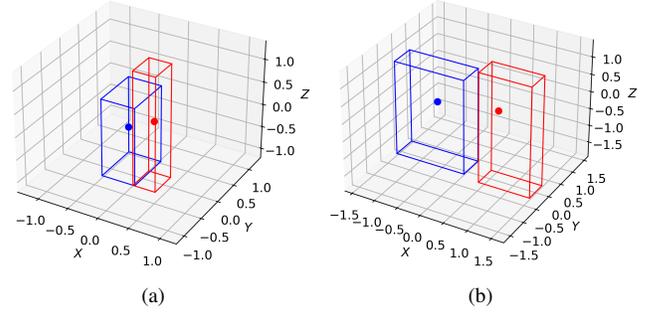


Fig. 4. 3D object pairs from Fig. 1a (a) and Fig. 1b (b) shown as AABBs and their centroids. The pairs have each been translated to the origin to simplify the presentation. (This does not change the relative position between the objects in each pair.)

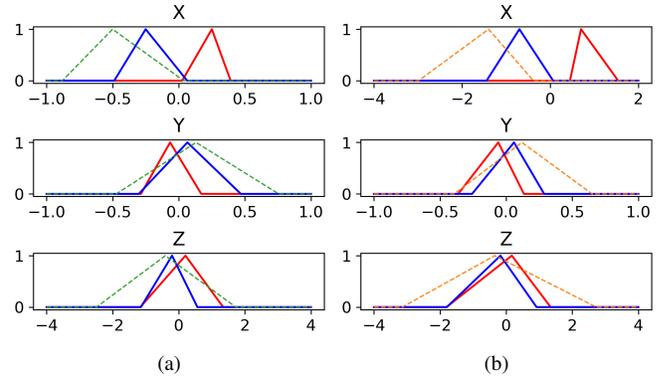


Fig. 5. Triangular Fuzzy Numbers (TFNs) for the object pairs shown in Fig. 4a (a) and Fig. 4b (b). Along each axis, the difference TFN is shown as a green dashed line for (a) and an orange dashed line for (b). These difference TFNs comprise the TFN-RPD and are computed by subtracting the position of the argument object (blue) from the referent (red).

## IV. BOUNDING BOX RELATIONS

### A. Triangular Fuzzy Number Descriptor

In [10], Buck et al. proposed a framework for constructing a spatial relationship graph (SRG) over a labeled point cloud dataset using bounding boxes and triangular fuzzy numbers (TFNs). For each object in a scene, the set of 3D points comprising that object are reduced to an axis-aligned bounding box (AABB) and a centroid. The relative position between objects can then be computed by considering the extent of the object in each dimension as a TFN. Distance and directional queries can be performed using this simplified representation, using the uncertainty associated with each TFN to guide the selection.

Consider two objects  $A$  and  $B$ , each comprised of a set of 3D points. We denote the AABB of  $A$  with two new points,  $A_{\min}$  and  $A_{\max}$ . Likewise, the AABB of  $B$  is given by  $B_{\min}$  and  $B_{\max}$ . The centroids of both objects are given by  $A_{\text{mean}}$  and  $B_{\text{mean}}$ . As an example, see Fig. 4, which shows the AABBs and centroids for the example object pairs in Fig. 1.

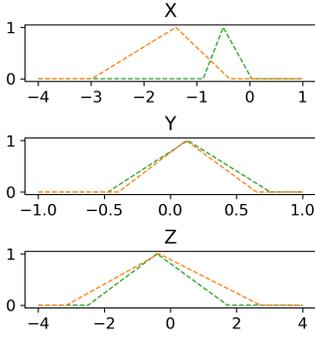


Fig. 6. TFN-RPDs for the object pairs in Fig. 4 plotted on the same axes to aid in comparison. Similar membership functions indicate similar spatial configurations.

The location of  $A$  and  $B$  along each of the three axes ( $x$ ,  $y$ , and  $z$ ) can be expressed as the TFNs  $A_x$ ,  $A_y$ ,  $A_z$  and  $B_x$ ,  $B_y$ ,  $B_z$ . Each TFN spans the range of the object (min to max) with the peak at the object’s centroid (see Fig. 5). The difference between  $A$  and  $B$  along any of the three axes can be computed by subtracting the TFN in that dimension of  $B$  from that of  $A$ . The result is a new set of three TFNs that define the “vector”  $D$  representing the minimum, maximum, and average displacements of a point from within the bounds of  $A$  such that it would end inside  $B$ . Here, the average displacement is the vector between the object centroids, i.e. from  $A_{\text{mean}}$  to  $B_{\text{mean}}$ . Formally, we have

$$D_x = B_x - A_x \quad (8)$$

$$D_y = B_y - A_y \quad (9)$$

$$D_z = B_z - A_z. \quad (10)$$

We call this set of three difference TFNs the TFN relative position descriptor (TFN-RPD) for the objects  $A$  and  $B$ ,

$$\text{TFN-RPD}(A, B) = (D_x, D_y, D_z). \quad (11)$$

Fig. 5 shows the TFNs that describe the object pairs in Fig. 4. For each pair, and along each axis, the TFN for object  $A$  (the referent) is shown in red and the TFN for object  $B$  (the argument) is shown in blue. The TFN of the difference vector  $D$  is shown as a dashed line and indicates the TFN that results from computing  $B - A$ . The difference vectors (TFN-RPD) from (a) are shown in green and the ones from (b) are shown in orange.

### B. Bounding Box Relation Similarity

To compare two object pairs using the TFN-RPD, we need to define the similarity  $S_{\text{TFN}}(D, D') \in [0, 1]$ , where  $D$  and  $D'$  are the difference TFNs as computed above. In this section, we discuss several different ways to define this measure used in the literature. First, we consider the similarity between the TFNs directly for each axis. Then we present a method using bounding box similarity measures.

TABLE III  
OBJECT PAIR SIMILARITIES USING THE TFN-RPD  
SINGLE-AXIS METHODS

	$s^X$	$s^Y$	$s^Z$	$S_{\min}$	$S_{\text{mean}}$
$\mu_{\max}$	0.358	0.990	0.988	0.358	0.778
$\mu_{\text{IoU}}$	0.053	0.846	0.717	0.053	0.539
$\mu_{\text{PD}}$	0.507	0.920	0.855	0.507	0.761

1) *Single-Axis Methods*: The TFN-RPDs for the example objects in Fig. 4 are shown again in Fig. 6, this time plotted on the same axes. The ranges of the TFNs indicate the relative position of the argument object from the referent along each axis. If the two TFN-RPDs overlap, there is some agreement that they share a similar spatial configuration. We can measure this by defining a similarity measure between two TFNs. We consider the following three methods in this work.

- **Max**: In this method, the similarity  $\mu_{\max}$  is defined as the maximum of the minimum of the two membership values taken over all real numbers as in [10]. If the TFN memberships are defined as  $m_D(x)$  and  $m_{D'}(x)$ , then

$$\mu_{\max}(D, D') = \max_{x \in \mathbb{R}} \{ \min(m_D(x), m_{D'}(x)) \}. \quad (12)$$

- **IoU**: This method defines the similarity  $\mu_{\text{IoU}}$  as the intersection over union of the areas of the two TFNs as though they were (potentially) overlapping triangles. Compared to the Max method, the IoU considers the full width of the membership function and not just a single point of intersection. Given two TFNs  $D$  and  $D'$ , the similarity is defined as

$$\mu_{\text{IoU}}(D, D') = \frac{|D \cap D'|}{|D \cup D'|}. \quad (13)$$

- **PD**: Here, the similarity  $\mu_{\text{PD}}$  is defined as the percent difference between the endpoints and centers of the two TFNs. If there is no overlap between the two membership functions, the previous two methods will give a similarity of zero, whereas this approach measures how much one TFN would need to be scaled as a fraction of the total support width in order to match the other TFN. The distance is then inverted to become a similarity measure in the range  $(0, 1]$ . If the two TFNs are defined as  $D = (d_1, d_2, d_3)$  and  $D' = (d'_1, d'_2, d'_3)$ , then

$$\text{PD}(D, D') = \sum_{i=1}^3 \frac{|d_i - d'_i|}{d_3 - d_1 + d'_3 - d'_1}, \quad (14)$$

and

$$\mu_{\text{PD}}(D, D') = \frac{1}{1 + \text{PD}(D, D')}. \quad (15)$$

In the single-axis approach, the overall similarity between two TFN-RPDs  $D$  and  $D'$  is computed by aggregating the TFN similarities for each axis, similar to the HoF-RPD. Using one of the above TFN similarity methods  $\mu$ , we consider both the minimum and the mean similarities to give the following overall measures,

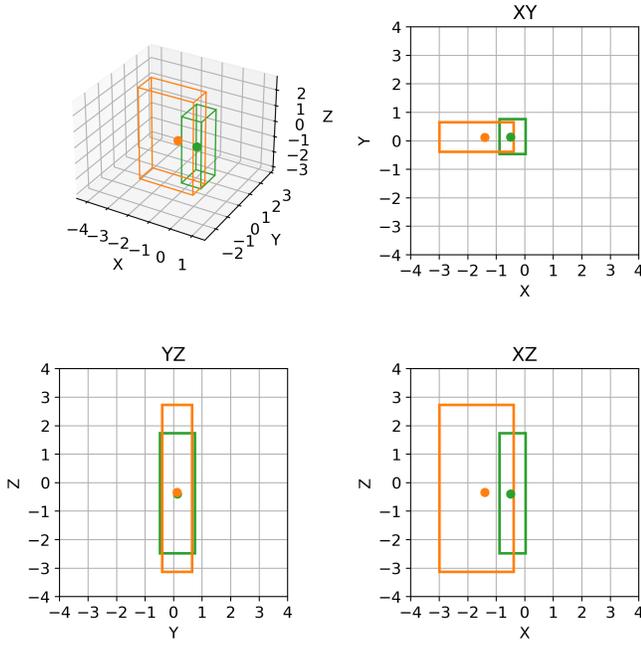


Fig. 7. Plots of the TFN-RPDs from Fig. 6 in 3D space as bounding boxes and centroids. The 2D views along each axis are also shown.

$$S_{\text{TFN-SA,min},\mu}(D, D') = \min(s_{\mu}^X, s_{\mu}^Y, s_{\mu}^Z), \quad (16)$$

and

$$S_{\text{TFN-SA,mean},\mu}(D, D') = \frac{1}{3}(s_{\mu}^X + s_{\mu}^Y + s_{\mu}^Z), \quad (17)$$

where  $s_{\mu}^i$  is the similarity between the TFNs on axis  $i$  using method  $\mu$ . The similarity scores for the example in Fig. 4 using the single axis methods are shown in Table III.

2) *Bounding Box Methods*: Another way to interpret the TFN-RPD is as its own bounding box and centroid in 3D space. Fig. 7 shows the TFN-RPDs from Fig. 6 plotted as bounding boxes, as well as projections along each axis. One limitation of the single-axis methods is that they do not directly account for the interactions between axes. By considering 2D bounding box projections or the full 3D bounding boxes, these correlations can be captured more completely. We consider the following three bounding box comparison methods, which can be applied in both 2D and 3D.

- **IoU**: This method computes the intersection over union (IoU) metric of the two bounding boxes as  $\mu_{\text{IoU}}$ . Given two TFN bounding boxes  $D$  and  $D'$ , the similarity is defined as the same as before with Eq. 13, except now using only the membership function endpoints as bounding box coordinates rather than the triangular membership function itself.
- **GIoU**: This method uses the generalized IoU to compute the similarity measure  $\mu_{\text{GIoU}}$ , as described in [17]. Since the IoU gives a similarity of zero if the two boxes do not overlap, the GIoU considers also the convex hull of

TABLE IV  
OBJECT PAIR SIMILARITIES USING THE TFN-RPD  
BOUNDING BOX METHODS

	$s^{XY}$	$s^{XZ}$	$s^{YZ}$	$S_{\min}$	$S_{\text{mean}}$	$S_{3D}$
$\mu_{\text{IoU}}$	0.157	0.123	0.637	0.123	0.306	0.119
$\mu_{\text{GIoU}}$	0.526	0.542	0.797	0.526	0.622	0.484
$\mu_{\text{PD}}$	0.486	0.467	0.796	0.467	0.583	0.449

the two boxes  $C$  and subtracts the percentage of the hull covered from the IoU score. The original GIoU is in the range  $[-1, 1]$ , so we modify it to be in the range  $[0, 1]$  as follows,

$$\mu_{\text{GIoU}}(D, D') = \frac{1}{2} \left( \frac{|D \cap D'|}{|D \cup D'|} - \frac{|C \setminus (D \cup D')|}{|C|} + 1 \right). \quad (18)$$

- **PD**: The final method is the same  $\mu_{\text{PD}}$  similarity as before, only now applied in two or three dimensions. Unlike the IoU and GIoU methods, the PD method considers the centroid location as well as the bounding box edges. For each axis  $j$ , consider the TFNs along axis  $j$  as  $D_j$  and  $D'_j$ . Then the PD similarity is computed using Eq. 14 as

$$\mu_{\text{PD}}(D, D') = \frac{1}{1 + \sum_{j \in \{x, y, z\}} \text{PD}(D_j, D'_j)}. \quad (19)$$

For the bounding box methods, the overall similarity between two TFN-RPDs  $D$  and  $D'$  can be computed as before, by aggregating the similarities from each 2D projection, or by considering the full 3D bounding boxes. The 2D methods more closely align with the way in which the HoF similarity is computed, whereas the 3D approach is more straightforward. We have the following overall bounding box similarity measures,

$$S_{\text{TFN-BB,min},\mu}(D, D') = \min(s_{\mu}^{XY}, s_{\mu}^{XZ}, s_{\mu}^{YZ}), \quad (20)$$

$$S_{\text{TFN-BB,mean},\mu}(D, D') = \frac{1}{3}(s_{\mu}^{XY} + s_{\mu}^{XZ} + s_{\mu}^{YZ}), \quad (21)$$

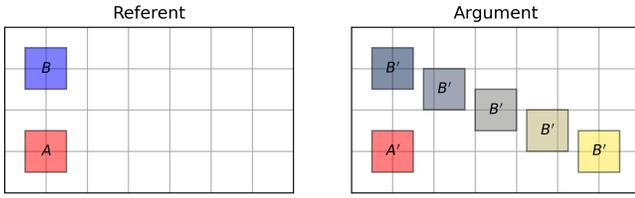
and

$$S_{\text{TFN-BB,3D},\mu}(D, D') = s_{\mu}^{XYZ}, \quad (22)$$

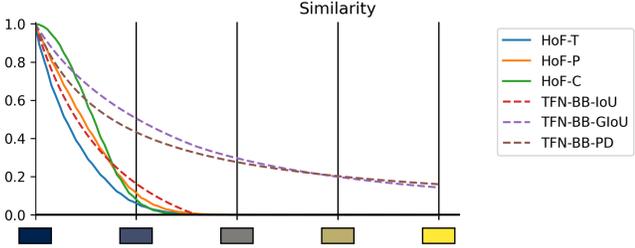
where  $s_{\mu}^k$  is the similarity between the bounding boxes using the axes in  $k$  and method  $\mu$ . The similarity scores for the example in Fig. 4 using the bounding box methods are shown in Table IV.

## V. COMPARISON IN 2D

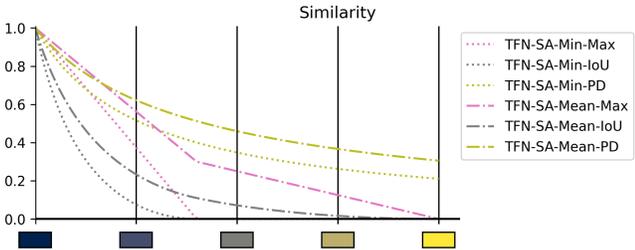
Having defined several measures for evaluating the similarity of two spatial relationships using both force histograms and bounding box methods, we now consider ways to compare the different approaches. We start by examining how the measures behave on simple 2D objects. Fig. 8 and Fig. 9 show two example cases using flat square objects of uniform weight.



(a)



(b)

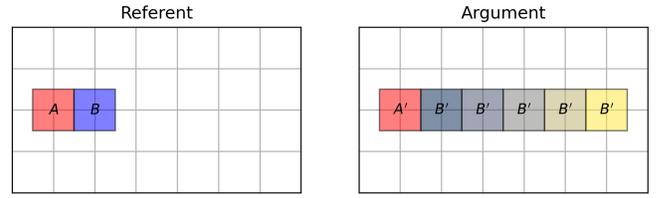


(c)

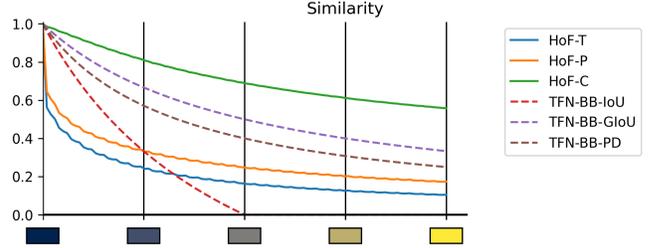
Fig. 8. (a) A referent and argument pair of 2D objects to compare. The left plot shows a static pair of objects  $A$  and  $B$  to be used as the referent spatial relationship. The right plot shows  $A'$  and several variations of  $B'$ . (b), (c) Similarity scores between the referent and argument spatial relationships as  $B'$  shifts down and to the right (changes from dark blue to yellow).

In both examples, a pair of objects is defined as the referent relationship and we evaluate the similarity of this object pair to an argument relationship where one of the objects moves.

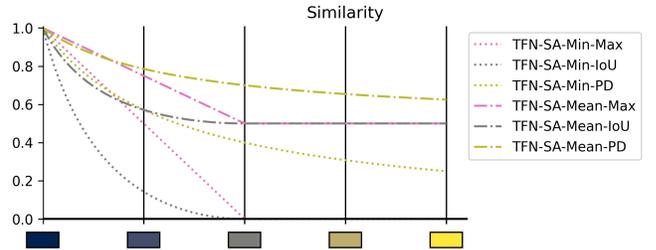
Consider first the example in Fig. 8a. Here the referent relationship has two objects,  $A$  and  $B$ , with  $B$  positioned above  $A$ . The argument relationship has objects  $A'$  and  $B'$ , with  $A'$  being static and  $B'$  moving from the top side of  $A'$  down and to the right. As  $B'$  moves, it changes color in the figure from dark blue to yellow. Fig. 8b and Fig. 8c show the similarity  $S(A, B, A', B')$  for several different measures as  $B'$  shifts in position. The scores all begin at 1 for the perfect match between the two pairs where  $B'$  is directly above  $A'$  and decrease as  $B'$  moves to the lower right (becomes more yellow). In Fig. 8b, three HoF similarity measures and three TFN bounding box measures are shown. All HoF measures and the TFN-BB-IoU method have decreased to a similarity of 0 by the middle position of  $B'$ , whereas the other two TFN-BB methods continue to have non-zero similarity even in this quite different configuration. Fig. 8c shows the TFN single-axis similarity measures for this example. All methods except PD have been reduced to zero by the final position of  $B'$ . Of note is the sharp inflection point of the TFN-SA-Mean-Max



(a)



(b)



(c)

Fig. 9. (a) A referent and argument pair of 2D objects to compare. The left plot shows a static pair of objects  $A$  and  $B$  to be used as the referent spatial relationship. The right plot shows  $A'$  and several variations of  $B'$ . (b), (c) Similarity scores between the referent and argument spatial relationships as  $B'$  shifts farther to the right (changes from dark blue to yellow).

method, which highlights the point at which the x-axis TFN no longer contributes to the similarity.

The second example in Fig. 9a shows a similar analysis, this time with the objects moving only along the x-axis. Here, the two objects  $A$  and  $B$  are adjacent in the referent pose, and  $B'$  moves away from  $A'$  in the argument. Fig. 9b and Fig. 9c show the similarity scores generally decreasing as  $B'$  moves farther away. Of the HoF methods, the HoF-C method maintains a significantly higher degree of similarity than the others, likely due to the implicit normalization of the histograms, which reduces sensitivity to distance. The TFN-BB-IoU, TFN-SA-Min-Max, and TFN-SA-Min-IoU methods have all reached zero similarity by the halfway point, demonstrating the restrictiveness of these approaches.

## VI. 3D EXPERIMENTS

The HoF-RPD and TFN-RPD features introduced in the previous sections are ways to capture the spatial relationships between two 3D point cloud objects. We have defined several similarity measures that can be applied to the descriptors to assess how similar one pair of objects is to another spatially. In total, 21 different methods have been described, each

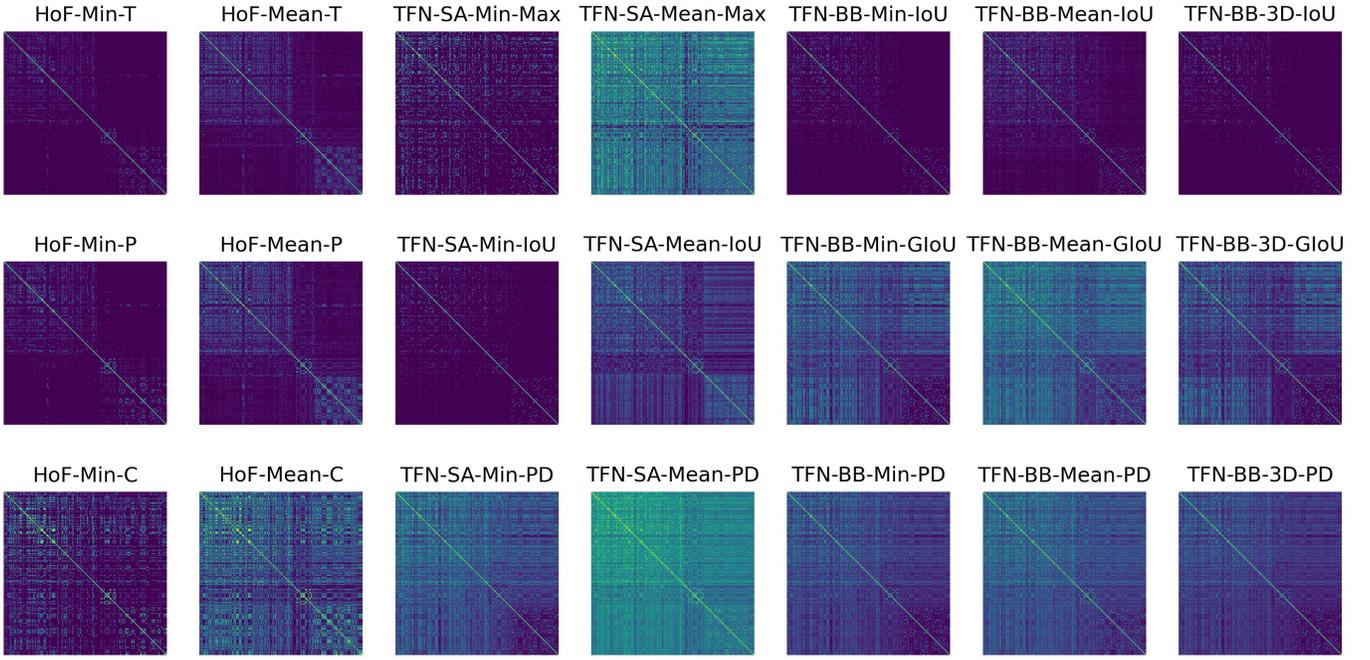


Fig. 10. Similarity matrices computed for 202 object pairs from the NPM3D dataset. Each matrix shown uses a different method to compute the similarity between two pairs of 3D point cloud objects. All images are scaled to the range  $[0, 1]$ .

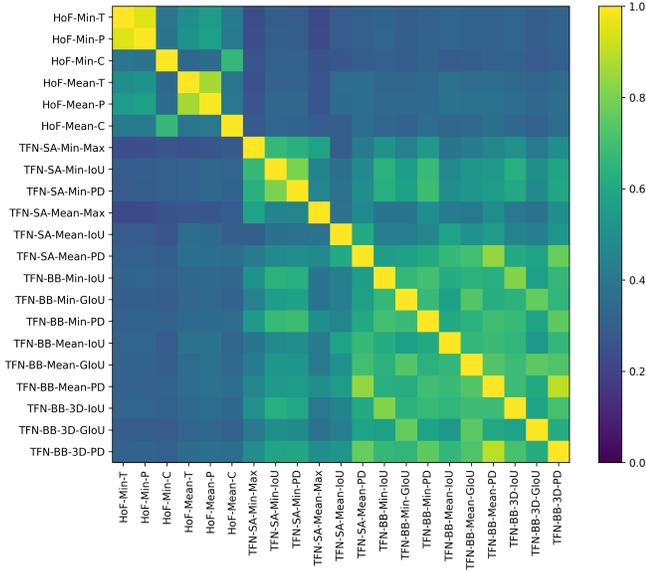


Fig. 11. Similarity matrix between all 21 methods of computing spatial similarities between object pairs.

producing a similarity  $S$ . In this section, we compare these methods on the NPM3D hand-labeled point cloud dataset [15].

We begin by computing a spatial relationship graph (SRG) over a portion of the data using the approach described in [10]. We then extract a subgraph containing only edges where the maximum distance between objects is less than 5 meters. This ensures that projected images maintain a reasonable size

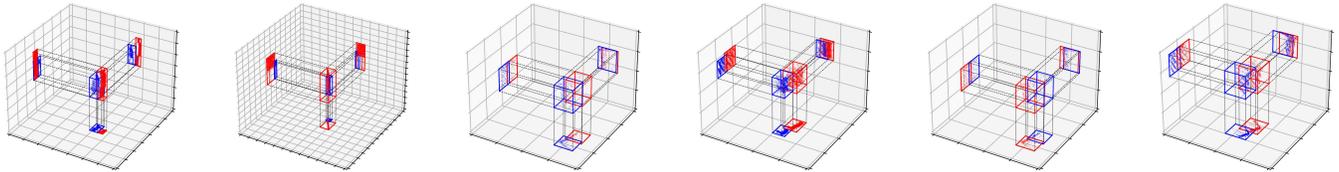
during rasterization. The resulting subgraph contains 202 object pairs, each consisting of a referent and argument 3D point cloud such as the ones in Fig. 1.

For every pair of graph edges (two object pairs), we compute the similarity score using each of the methods described above and produce a similarity matrix (Fig. 10). The matrices retain the original ordering of the pairs, so patterns can be observed between them. While it is difficult to draw any meaningful conclusions from these matrices alone, we can get a sense of how restrictive each method is at assigning similarity. For example, the TFN-SA-Mean-PD method generally assigns high similarity scores, whereas the HoF-Min-T method is much more likely to give low scores.

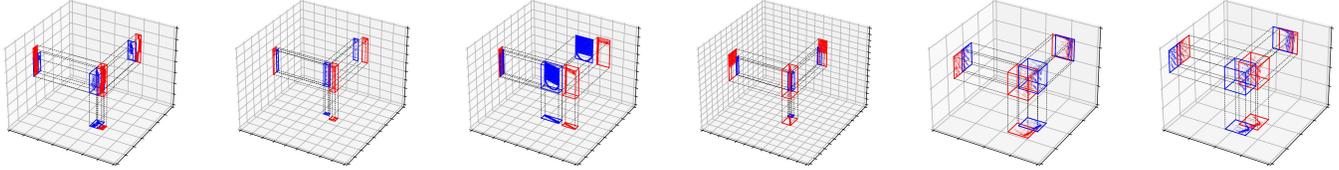
We next look at how consistent the methods are at identifying the same set of similar object pairs. For any pair of methods  $(i, j)$  and object pair index  $k$ , we identify the top 5 matches (other than  $k$ ) using each method. Let  $M_{ik}$  be the set of 5 object pairs that most closely match object pair  $k$  according to method  $i$ , and likewise  $M_{jk}$  the set of 5 pairs according to method  $j$ . We define the method similarity as

$$s_m(i, j) = \frac{1}{|K|} \sum_{k \in K} \frac{|M_{ik} \cap M_{jk}|}{|M_{ik} \cup M_{jk}|}. \quad (23)$$

$s_m(i, j)$  represents the average Jaccard index between the top 5 matches of methods  $i$  and  $j$  over all object pairs. The similarity matrix for all methods is shown in Fig. 11. From this we see that there is general agreement within the HoF and TFN methods, but a clear difference between these two broad categories of approaches.

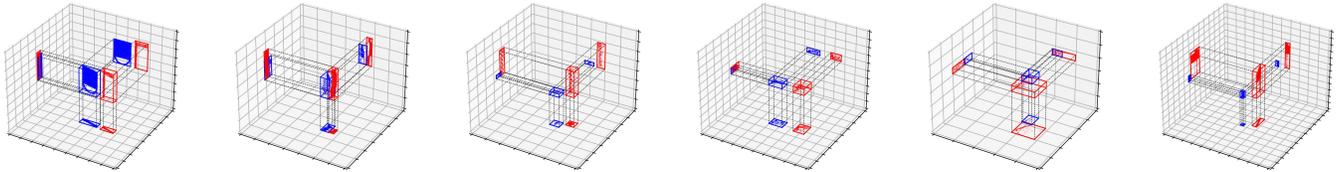


(a)

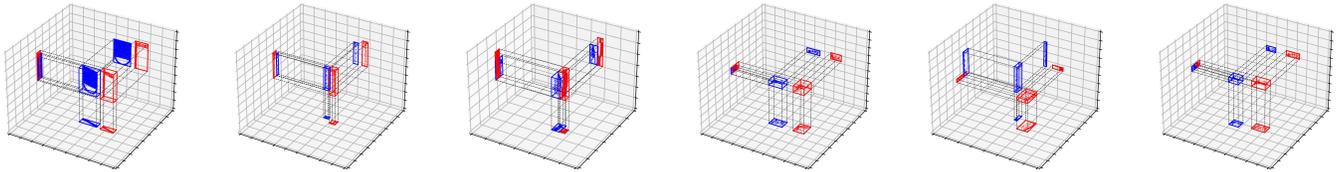


(b)

Fig. 12. Top matches (decreasing left to right) for the referent object pair from Fig. 1a using the HoF-Min-T (a) and TFN-BB-3D-PD (b) methods.



(a)



(b)

Fig. 13. Top matches (decreasing left to right) for the referent object pair from Fig. 1b using the HoF-Min-T (a) and TFN-BB-3D-PD (b) methods.

Lastly, we look at a few examples in detail to gain a better understanding of how two of the methods compare in practice. Fig. 12 shows the top 6 matches for the object pair in Fig. 1a using the HoF-Min-T and TFN-BB-3D-PD methods. (The top match is always an identical match.) The HoF-Min-T method was chosen as a representative of the HoF approaches for its sensitivity to the distance between objects and the TFN-BB-3D-PD method was chosen to demonstrate the full capabilities of the 3D bounding box TFN-RPD. Not counting the top identical match, 2 of the 5 object pairs are common to both methods. We note that the reference object pair in Fig. 13 appears as the second best match using the TFN-BB-3D-PD method, but does not appear in the top 5 matches of the HoF-Min-T method.

Fig. 13 shows another example using the object pair in Fig. 1b and the same two similarity methods. Again, 2 of the 5 top matching object pairs are common to both methods. In this case, the reference objects from Fig. 12 appear as top matches for both methods.

## VII. CONCLUSION

The relative position between two 3D point cloud objects can be captured in a variety of ways, each with its own set of strengths and weaknesses. The HoF-RPD and TFN-RPD features described above and the various methods to compare them all share the fundamental quality of being able to describe the spatial relationship between two objects and evaluate the similarity of that relationship to another pair of objects. This is a critical component of many spatial reasoning applications, and depending on the problem specifics, some approaches may be better suited than others.

The HoF-based methods as used here have several shortcomings in the context of evaluating spatial relationships between 3D point clouds. Chiefly among these is the current requirement to rasterize the pair of objects into a common image space, limiting the distance at which the force histograms can be computed. Objects that are very far apart may not actually have any computed angles  $\theta$  that intersect

both objects, resulting in an empty histogram. This issue could be mitigated by using a vector-based HoF method or by increasing the histogram resolution. Another strategy might be to utilize a 3D-based HoF method directly, or one specifically adapted for point clouds.

The difficulty in applying the HoF methods at long distances suggests that they may not be well-suited for applications where this is the main type of spatial relationship. As objects move farther apart, the relationship between them can be more closely modeled as a single crisp vector, and the more complex spatial features become less relevant. In these contexts, the TFN-based methods seem more appropriate as they make direct use of this directional vector. In particular, methods that do not rely on overlapping sets, such as the GIoU and PD methods, can be more robust to small variances in the object positions.

Although the TFN methods are not as descriptive as the HoF methods, they can still be quite useful for identifying and characterizing general relationships between objects. By utilizing the centroids of the point clouds, the PD method is most capable of the TFN methods at capturing the underlying spatial distribution of the objects. Furthermore, the full 3D bounding box methods seem to be most appropriate to use in most cases, unless there is a specific need to consider axes independently.

It should be noted that we have not explored the notion of rotation and scale invariance in the present work. Such issues are often critical aspects of spatial reasoning systems and depend on the ability to normalize and compare descriptors in a common frame of reference. Several strategies have been employed for the histogram of forces [18], however the need for normalization remains application dependent.

Simplifying an object to its AABB introduces empty space that is not part of the original object. The use of the point cloud centroid is an attempt to mitigate this effect, but other approaches such as oriented bounding boxes or a hierarchical method could also be used. AABBs are inherently dependent on the frame of reference, and the impact of choosing an appropriate frame will be explored in future work.

In conclusion, while the HoF methods are best suited for cases where objects are relatively close together and consist of intricate shape details that need to be captured, the TFN methods can provide an alternative approach that may be easier to implement and more appropriate for use cases where shape discrimination is less important than relative position. Both approaches have qualities that make them desirable to use in certain situations, and many variations on how they can be applied. Ultimately, the utility of a relative position descriptor is dependent on the problem domain and the application requirements, and subjective analysis can make any one of these strategies a good solution.

## REFERENCES

- [1] M. Naem and P. Matsakis, "Relative position descriptors: A review," in *Proceedings of the International Conference on Pattern Recognition Applications and Methods*. Lisbon, Portugal: SCITEPRESS - Science and Technology Publications, 2015, pp. 286–295.
- [2] P. Matsakis, "Relations spatiales structurelles et interprétation d'images," Ph.D. dissertation, IRIT, Université Paul Sabatier, Toulouse, France, 1998.
- [3] P. Matsakis and L. Wendling, "A new way to represent the relative position between areal objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 7, pp. 634–643, Jul. 1999.
- [4] P. Matsakis, J. Keller, O. Sjahputera, and J. Marjamaa, "The use of force histograms for affine-invariant relative position description," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 1–18, Jan. 2004.
- [5] P. Matsakis, J. Keller, L. Wendling, J. Marjamaa, and O. Sjahputera, "Linguistic description of relative positions in images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 31, no. 4, pp. 573–588, Aug. 2001.
- [6] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D. Brock, "Spatial language for human-robot dialogs," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 2, pp. 154–167, May 2004.
- [7] O. Sjahputera and J. M. Keller, "Scene matching using F-histogram-based features with possibilistic C-means optimization," *Fuzzy Sets and Systems*, vol. 158, no. 3, pp. 253–269, Feb. 2007.
- [8] A. R. Buck, J. M. Keller, and M. Skubic, "A memetic algorithm for matching spatial configurations with the histograms of forces," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 4, pp. 588–604, Aug. 2013.
- [9] J. Ni, P. Matsakis, and L. Wawrzyniak, "Quantitative representation of the relative position between 3D Objects," in *Proceedings of the Fourth IASTED International Conference on Visualization, Imaging, and Image Processing*, 2004.
- [10] A. R. Buck, D. T. Anderson, J. M. Keller, R. H. Luke, and G. Scott, "A fuzzy spatial relationship graph for point clouds using bounding boxes," in *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Jul. 2021, pp. 1–8.
- [11] X. Tang, "Spatial object modeling in fuzzy topological spaces: With applications to land cover change," Ph.D. dissertation, ITC, Enschede, The Netherlands, 2004.
- [12] A. C. Carniel and M. Schneider, "A Survey of Fuzzy Approaches in Spatial Data Science," in *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. Luxembourg, Luxembourg: IEEE, Jul. 2021, pp. 1–6.
- [13] I. Bloch, "Fuzzy spatial relationships for image processing and interpretation: A review," *Image and Vision Computing*, vol. 23, no. 2, pp. 89–110, Feb. 2005.
- [14] O. Sjahputera, "Object registration in scene matching based on spatial relationships," Ph.D. dissertation, University of Missouri-Columbia, 2004.
- [15] X. Roynard, J.-E. Deschaud, and F. Goulette, "Paris-Lille-3D: A large and high-quality ground truth urban point cloud dataset for automatic segmentation and classification," *arXiv:1712.00032 [cs, stat]*, Apr. 2018.
- [16] C. P. Pappis and N. I. Karacapilidis, "A comparative assessment of measures of similarity of fuzzy values," *Fuzzy Sets and Systems*, vol. 56, no. 2, pp. 171–174, Jun. 1993.
- [17] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE, Jun. 2019, pp. 658–666.
- [18] M. Jazouli, J. Wadsworth, and P. Matsakis, "Normalization of the histogram of forces," in *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*. Prague, Czech Republic: SCITEPRESS - Science and Technology Publications, 2019, pp. 630–639.