Object Classification Using Tripod Operators

David Bonanno, Frank Pipitone, G. Charmaine Gilbreath,
Kristen Nock, Carlos A. Font, and Chadwick T. Hawley
US Naval Research Laboratory, 4555 Overlook Ave. SW, Washington DC 20375

David.bonanno@nrl.navy.mil; 202-404-1947

ABSTRACT

Over the last few decades, we have seen an increase in both quality and quantity of 3D data sets. These data sets primarily come in the form of discrete points that are projected onto the surface of the object (point clouds) and are often derived from either LIDAR data (in which case, the surface points are actively sensed) or stereoscopic pairs (in which case, the surface points are derived using two dimensional (2D) feature matching algorithms).

As these data sets become larger and denser, they also become harder to sift through which demands methods for automatic object classification through computer vision processes. In this paper we revisit a method of recognizing objects from their surface features known as Tripod Operators.*[1] More specifically, we explore how matching multiple features from an unknown object to a known shape allows us to determine the extent to which the objects are similar using the resultant Digital Elevation Model (DEM) or Surface Elevation Model (SEM) that results from manipulation of point clouds.. We apply this method to determine how to separate objects of various classes.*

Keywords: Tripod Operator, Object Classification, 3D, Computer Vision

I. INTRODUCTION

A variety of techniques have been developed over the years to classify objects in images. A popular method is to find feature points (sometimes called key points) in an image and create a feature vector. Techniques such as the Scale Invariant Feature Transform (SIFT) are able to not only describe features in an image, but are also able to describe these feature points in a way that is invariant to translations, rotations, and scale [2]. While certainly a robust and studied technique, the feature matching is ultimately limited in matching objects with dramatically different poses (i.e. a picture of a car from the front, and a separate from the side). While these pictures relate to the same object they would likely have only a few 2D features in common.

For this reason we turn our attention from the realm of 2D feature descriptors to 3D feature descriptors. In the world of 3D images (created often from LIDAR systems and/or stereoscopic imagery), features can exploit invariances in shape in three dimensions, allowing for the possibility of feature descriptors which are invariant to position, rotation, and also pose of the object.

This paper explores a technique to differentiate three dimensional (3D) objects into classes using 3D features. Our 3D feature extractors are known as *Tripod Operators (T.O.s)* [1,3] which are described in Section I-A. Section I-B introduces a method for creating a library of these surface features for a known reference object. Finally, a matching metric is presented whereby a single feature vector can be said to 'match' or 'not match' the library within a tolerance.

The percentage of matches (PoM) between random surface features extracted from the reference object and its own library is explored in Section II-A. In Section II-B dissimilar objects are introduced and their surface features are compared to the reference library generating separable PoM leading to classification. Finally, we show the tolerance can

Active and Passive Signatures V, edited by G. Charmaine Gilbreath,
Chadwick Todd Hawley, Proc. of SPIE Vol. 9082, 90820C · © 2014
SPIE · CCC code: 0277-786X/14/$18 · doi: 10.1117/12.2069529

be optimized to create an optimal classifying system. All of these computations are conducted by sampling DEMs or SEMs using T.O.'s.

A. Tripod Operators

The tripod operator is a class of feature extraction operators for range images which facilitate the recognition and localization of objects. It consists of three points in 3-spacefixed at the vertices of an equilateral triangle and a procedure for making several scalar measurements in the coordinate frame of the triangle [1]. A major advantage of the T.O. is that it can describe a 3D surface as a 1D feature vector making it computationally efficient to store and sort the features of a target object [3][4]. Essentially, an equilateral triangle "opens" one triangular face at a time. The face opens until its vertex intersects the target object's surface. The angle between the adjacent triangular faces is the angular feature. A nine-dimensional tripod operator is illustrated in Figure 1.

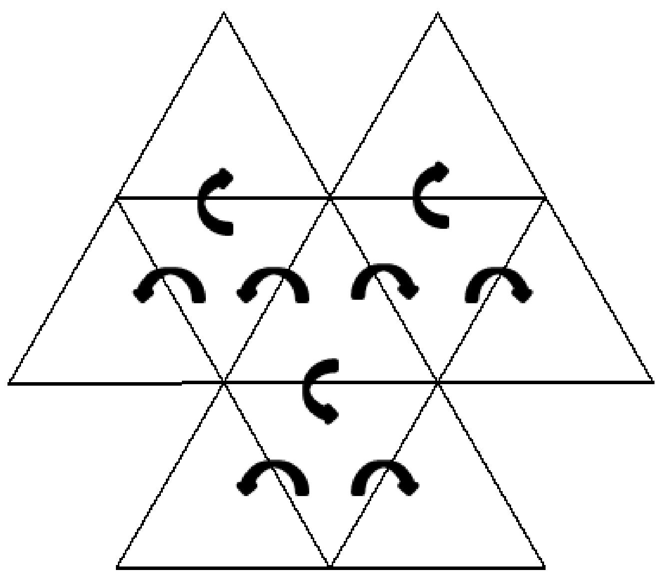

Figure 1: A nine dimensional Tripod Operator. Each triangular face opens up onto the surface of an object until the corner intersects the surface. The angle between the adjacent triangular faces is the angular feature.

Previously, we demonstrated the T.O. can be made to work on data that is generated from point clouds by processing the discrete points into a reconstructed surface [5]. These reconstructed surfaces can take on many forms including polygonal meshes, rigid voxels, and projected digital elevation models (DEMs). We choose to represent our data as DEMs which record the height of a surface at a pixel location [u, v] corresponding to real world coordinates (x, y). To create a T.O., an equilateral triangle is dropped onto the surface of an object. The sides of the equilateral triangle are scaled to the size of the object. A fourth point is generated by rotating around an edge of the base triangle at a fixed distance until the object intersects the DEM surface (as shown in Figure 2) resulting in a second equilateral triangle. This can be rapidly done using a binary search algorithm. The total angular displacement between the base triangle and the new triangle is the first feature in the feature array. It is independent to three degrees of rotation and three degrees of translation.

More triangular faces can be added in a similar manner creating N angular features with N+3 points on the surface of the object. Again, these features are independent of the position and orientation of the object. An example of a T.O. on a DEM surface, as well as the resultant feature array can be seen in Figure 3.

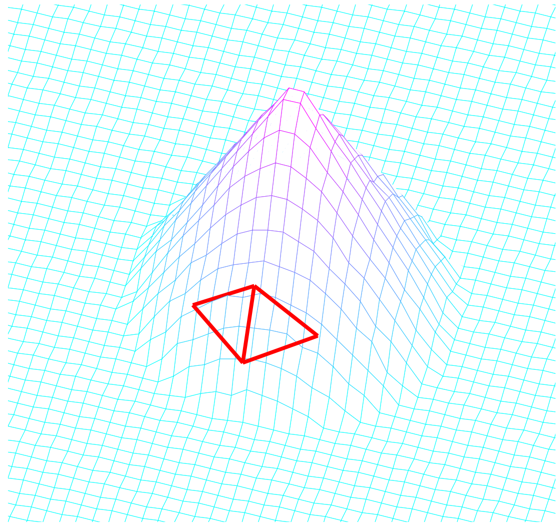

Figure 2: The tripod operator begins to open onto the surface. The two equilateral triangles are hinged about their shared side. Only the vertexes of the triangles necessarily touch the surface of the DEM.

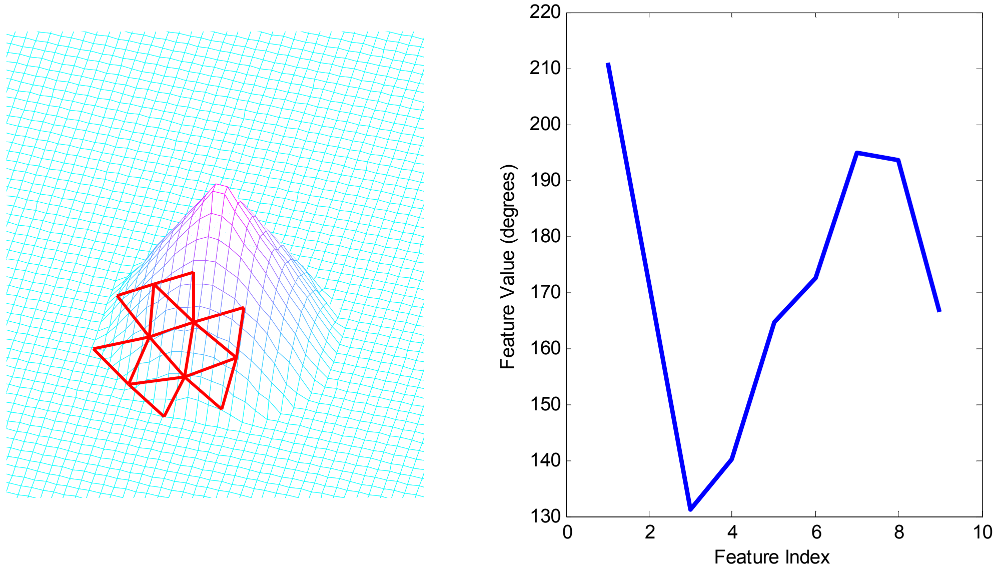

Figure 3: A complete nine dimensional tripod operator has unfolded onto the surface of a DEM (a). The resultant tripod feature showing the angular opening for each of the triangular faces (b).

B. Feature Matching

A library of *M* features can be generated for a reference object by extracting multiple *N*-dimensional surface features from the surface at random. It should be noted that in addition to the feature vector, each dropped T.O. also has position and orientation components. These can be useful for many applications including pose estimation [5] and relative feature positional invariance. We intend to explore this positional invariance in future papers.

The minimum amount of error (ε_o) between a feature vector and the library is calculated using a matching metric. We chose the sum of absolute differences for its computational simplicity.

$$\varepsilon(m) = \frac{1}{N}\sum_{n=0}^{N-1}|library(m,n) - target(n)| \quad (1)$$

$$\varepsilon_o = \min\big(\varepsilon(m)\big)$$

A binary decision can be made to determine if the minimum error is sufficiently close. The threshold decision is:

$$d = \begin{cases} 1, & \varepsilon_o < \tau \\ 0, & \varepsilon_o \geq \tau \end{cases}$$

Here a constant predetermined tolerance, τ, determines how close the feature vector needs to be to be considered a match. In the following section, the distribution of the PoM as well as the effects of changing the tolerance value τ is explored.

II. MULTIPLE FEATURE MATCHES

The previous section showed how a 3D feature vector can be created using Tripod Operators. These features can be extracted creating an M x N library. A new feature can be compared to this library using a matching metric resulting in a minimum error. Finally, a binary decision can be made where '1' represents a feature match and '0' represents no matches.

A. Expected Percentage of Matches

A single match between a feature and a library is not sufficient for classifying the entire object. Similarly, for a finite length library, not all new surface features extracted from an object will match its own library. For these reasons, multiple features are extracted from a single object and compared; the resulting PoM becoming the quantity necessary for classifying the object.

Additional *P* surface features are generated by randomly sampling the surface of the reference object. These new surface features are compared to the library with a fixed tolerance, τ. The PoM (represented as variable μ) between the new surface features and the library is recorded.

The expected value of the percentage of matches $E\{\mu\}$ is not dependent on the total number of surfaces features examined. As with many feature matching problems, the distribution of the percentage of matches can be modelled as a Gaussian random process [6]. Figure 4 shows the results of a Monte Carlo simulation creating the distribution of the percentage of matches between a target object and its library. In the first case, 100 features from the surface are extracted at a time. In the second test, 400 surface features are compared to the library. Notice that the expected percentage of

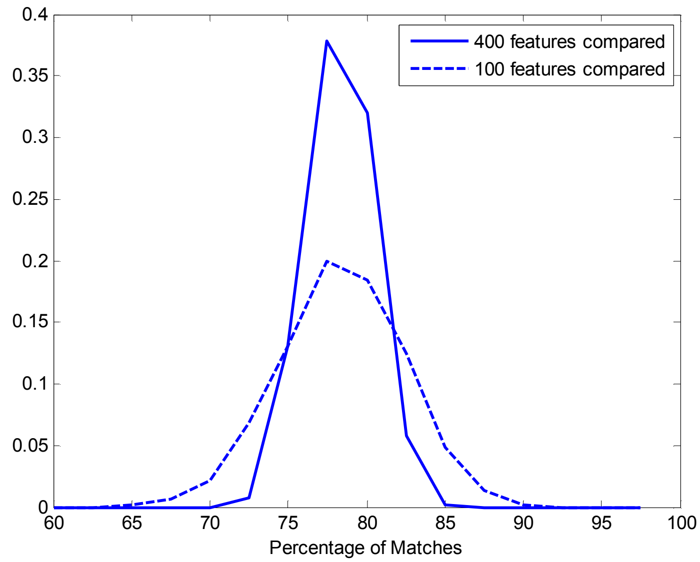

Figure 4: The results of a Monte Carlo simulation. No matter how many features are dropped at a time, the expected percentage of feature matches (within a tolerance) between an object and a library are the same. However, the standard deviation varies as a factor of $1/\sqrt{N}$.

matches remains the same; however, the standard deviation for the 400 drops is significantly lower (specifically half, as is expected with a Gaussian random process).

The expected PoM is also dependent on the tolerance. If the tolerance is low, then few of the features will match the library; however, if the tolerance is high enough every feature will match. The function $E\{\mu(\tau)\}$, while monotonic, is not necessarily smooth.

B. Object Classification

When surface features from dissimilar objects are compared to the reference library, they will also have a percentage of matches. Generally, these percentages of matches are lower than the PoM between the reference object and its library (Figure 5). Because of this, a classifier can be introduced. Should a μ collected from an object be greater than a predetermined PoM threshold (λ) the object will be classified as being similar to the reference object.

How separable the PoM between an object and its library are to the PoM of other objects to the same library is also dependent on the matching tolerance, τ. An optimal tolerance value can be determined by generated Receiver Operating Characteristic (R.O.C.) curves similar to the ones shown in Figure 6.

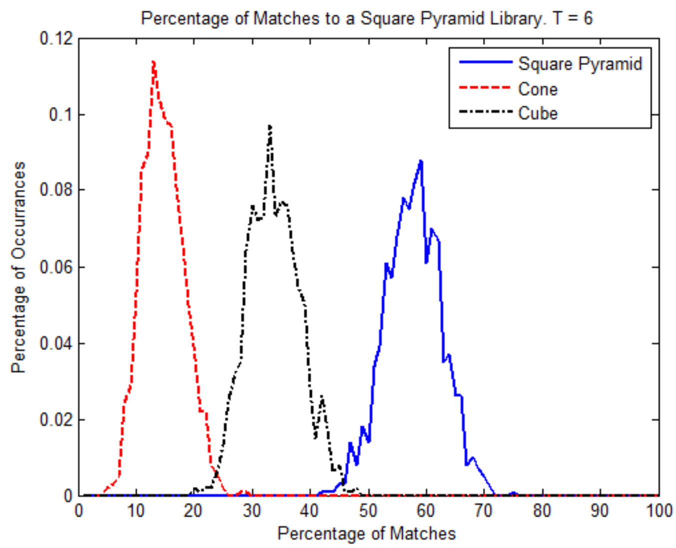

Figure 5: The percentage of matches between three shapes and a square pyramid library. Note that the square pyramid has the highest percentage of matches and is highly separable.

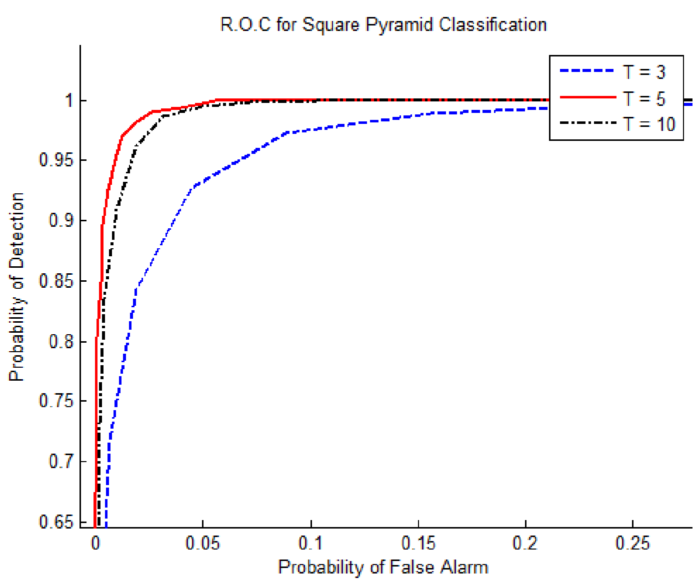

Figure 6: R.O.C curves for square pyramids. Should the tolerance be too low or too high the system will perform sub optimally.

III. CONCLUSIONS

A technique for classifying 3D objects from surface features generated with Tripod Operators has been presented. While the method developed focuses on classifying objects as matching or not matching a single library, it can easily be extended to include multiple libraries. We have shown how such a system can be optimized to achieve the best performance.

The technique presented lends itself to both supervised and unsupervised learning techniques. In future works we intend to extend the statistical methods developed in this paper to work into such systems allowing for automated processing of 3D data.

IV. REFERENCES

[1] Pipitone, F.; Adams, W., "Tripod operators for recognizing objects in range images: rapid rejection of library objects," *Robotics and Automation, 1992. Proceedings., 1992 IEEE International Conference on* , vol., no., pp.1596,1601 vol.2, 12-14 May 1992

[2] Lowe, D.G., "Object recognition from local scale-invariant features," *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* , vol.2, no., pp.1150,1157 vol.2, 1999

[3] Pipitone, F.; Adams, W., "Tripod operators for recognizing objects in range images: rapid rejection of library objects," *Robotics and Automation, 1992. Proceedings., 1992 IEEE International Conference on* , vol., no., pp.1596,1601 vol.2, 12-14 May 1992

[4] Pipitone, F.; Gilbreath, C.; Bonanno, D., "Tripod Operators for Efficient Search of Point Cloud Data for Known Surface Shapes," Proc. SPIE 8382, Active and Passive Signatures III, 83820O (7 May 2012);

[5] Pipitone, F, "Technique for estimating the pose of surface shapes using tripod operators," U.S. Patent: 6,393,143 B1, issued date May 21, 2002.

[6] Christmas, W.J.; Kittler, J.; Petrou, M., "Structural matching in computer vision using probabilistic relaxation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.17, no.8, pp.749,764, Aug 1995