

DETECTING MOVING SPHERES IN 3D POINT CLOUDS VIA THE 3D VELOCITY HOUGH TRANSFORM

Anas Abuzaina, Thamer Alathari, Mark S. Nixon, John N. Carter

School of Electronics and Computer Science, University of Southampton, UK
{aa6g08, tsalgl1, msn, jnc}@ecs.soton.ac.uk

ABSTRACT

We present a new approach to extracting moving spheres from a sequence of 3D point clouds. The new 3D velocity Hough Transform (3DVHT) incorporates motion parameters in addition to structural parameters in an evidence gathering process to accurately detect moving spheres at any given point cloud from the sequence. We demonstrate its capability to detect spheres which are obscured within the sequence of point clouds, which conventional approaches cannot achieve. We apply our algorithm on real and synthetic data and demonstrate the ability of detecting fully occluded spheres by exploiting inter-frame correlation within the 3D point cloud sequence.

1. INTRODUCTION

3D parametric object detection and tracking in 3D point clouds is fundamental to 3D computer vision and there have been a few techniques aimed to solve such problems [1-3]. Most available techniques consider information present in a single point cloud and work on a frame-to-frame basis by performing spatial alignment between consecutive frames based on locating feature descriptors or other attributes, rather than performing global analysis over the whole data sequence. In this work we follow a holistic view of the sequence and extract moving objects rather than track them.

We extend the velocity Hough Transform (VHT) [4], from optimum shape and motion parameters of conic sections moving in parametrically described fashion in a given sequence of images, to 3D point clouds in order to detect spheres moving with linear velocity. Considering a sequence of point clouds enables any evidence gathering technique to concurrently extract optimal structural and motion parameters exhibited by objects in the evidence gathering process. This gives improved performance than frame-by-frame tracking methods especially in the case of missing or occluded structural information, as any missing structural data in one point cloud is compensated from other point clouds. The significant advantage of the VHT which our 3DVHT inherits is that the correspondence problem is avoided; there is no need for feature detection and correspondence analysis between consecutive point clouds, which many 3D tracking methods rely on – and is implicit

in this new technique. Moreover, due to the global scope of the 3DVHT, there is no need to initialize the evidence gathering algorithm to a specific volume.

2. 3D HOUGH TRANSFORM FOR DETECTING SPHERES

The Hough Transform HT [5] is a well-known technique originally developed to extract straight lines, since then it has been extended to extract parametric (such as circles) and non-parametric shapes in 2D images. The Hough transform is recognized as a powerful tool in shape analysis that gives good results even in the presence of noise and occlusions.

Recently there have been approaches to extend the Hough transform for circle detection to three dimensions for detecting spheres [6] and cylinders [3] in 3D datasets. Equations (1), (2) and (3) define surface points on a sphere centered at (c_x, c_y, c_z) with radius r . For a fixed radius, each surface point defines a center of a sphere in the 3D accumulator space. After gathering evidence for all surface points, the maximum peak in the accumulator space corresponds to parameters of the original sphere.

$$x = c_x + r \cos \theta \sin \phi \quad (1)$$

$$y = c_y + r \sin \theta \sin \phi \quad (2)$$

$$z = c_z + r \cos \phi \quad (3)$$

Although it seems straightforward, the extension of the Hough accumulator space for 3D data can lead to excessive storage requirements which can make its application impractical. There have been approaches just to design the accumulator space to avoid such problems [6,1]. Accumulator space design is for later and we shall here demonstrate basic capability of moving 3D shape extraction.

3. VELOCITY HOUGH TRANSFORM

The HT conventionally concerns the extraction of structural parameters of stationary features, within single frames. The Velocity Hough Transform [4] identifies moving objects in 2D image sequences by extending the standard HT to include linear motion in the voting process; extending the HT from the spatial to the temporal domain. The VHT

combines feature extraction and motion analysis and estimates both structure and motion of a feature through global evidence gathering; evidence from the whole sequence is gathered in a single accumulator space and optimal structural and motion parameters are extracted concurrently. As a result of including motion in the evidence gathering process, the VHT is more efficient in extracting shapes than frame-to-frame tracking applications, especially if the tracked feature is occluded or contaminated by noise.

With the VHT, there is no need to match correspondences; since all correspondences are examined implicitly in the accumulation phase, the closest correspondences produce the highest peaks.

The major disadvantage of the VHT is that any extension of the motion or structure model will increase the dimensionality of the accumulator space, increasing the computational cost.

4. 3D VELOCITY HOUGH TRANSFORM

To detect spherical structures with constant linear motion in a sequence of point clouds, we introduce a generic evidence gathering technique which combines concepts from the VHT and the 3D Hough Transform, we call it the 3D Velocity Hough Transform (3DVHT).

The 3DVHT extends the polar sphere equations (1), (2) and (3) by adding velocity terms representing constant linear motion:

$$c_x = x - v_x t - r \cos \theta \sin \varphi \quad (4)$$

$$c_y = y - v_y t - r \sin \theta \sin \varphi \quad (5)$$

$$c_z = z - v_z t - r \cos \varphi \quad (6)$$

where v_x , v_y , v_z are the velocity of the sphere along the x , y , z axes respectively, measured by cm per time unit t , where t is the time reference value relative to the initial point cloud, i.e. for cloud 1 $t = 0$ and, cloud 2, $t = 1$ and so on. Where θ and φ vary in the range ($0 \leq \theta \leq 2\pi$ and $0 \leq \varphi \leq \pi$).

Algorithm 1 describes the voting process of the 3DVHT employing a multi-dimensional accumulator space to estimate the sphere's structural (c_x , c_y , c_z , r) and also motion parameters (v_x , v_y , v_z).

For each point cloud in the sequence, at each 3D point in the cloud, equations (4), (5) and (6) are employed to obtain the possible center of the sphere in a given range of radii and, x , y and z velocities, provided the time reference t of the point cloud. Points corresponding to each parameter combination are incremented in the accumulator space.

This process is performed for point clouds voting in the same accumulator. The peak of the accumulator is the best estimate of the sphere's structural and motion parameters; (c_x , c_y , c_z) represent the sphere center at $t = 0$ and (v_x , v_y , v_z) represent the corresponding velocities.

For (No. of point clouds)

For every point (x,y,z)

For (r = r_{min}; r < r_{max}; r++)

For (V_x = V_{xmin}; V_x ≤ V_{xmax}; V_x++)

For (V_y = V_{ymin}; V_y ≤ V_{ymax}; V_y++)

For (V_z = V_{zmin}; V_z ≤ V_{zmax}; V_z++)

For (θ = 0; θ ≤ 2π; θ++)

For (φ = 0; φ ≤ π; φ++)

$c_x = x - v_x t - r \cos \theta \sin \varphi$

$c_y = y - v_y t - r \sin \theta \sin \varphi$

$c_z = z - v_z t - r \cos \varphi$

Accumulator[c_x][c_y][c_z][v_x][v_y][v_z][r]++

Search Accumulator for peak.

Algorithm 1. 3DVHT.

To detect the sphere at any point cloud we simply apply equation (7); same equation applies for y and z coordinates. To reduce the complexity of the voting process, the number of dimensions of the accumulator space can be reduced by predefining one or more parameters. All the experiments discussed in the next section have predefined sphere radius.

$$c_x|_t = c_x|_{t=0} + v_x t \quad (7)$$

5. ANALYSIS

To test our method, we first apply it on a synthetic sphere sequence moving with linear x , y , z velocities, then we test the 3DVHT on a real sequence of ball acquired by the Microsoft Kinect sensor, then we introduce occlusions to the real sequence and demonstrate how the 3DVHT can locate spheres in fully occluded point clouds.

5.1. Synthetic sphere sequences

Figure 1 shows a sequence of 10 point clouds (merged) of a synthetic scene including a sphere of specified radius $r = 90$ cm moving in three dimensions with constant 3D velocity ($v_x = 110\text{cm/t}$, $v_y = 140\text{cm/t}$ and $v_z = 180\text{cm/t}$), starting at 3D center location of $c_x = 190\text{cm}$, $c_y = 100\text{cm}$ and $c_z = 120\text{cm}$.

The voting process for this sequence employs the algorithm 1 on a six-dimensional accumulator to find the sphere's center and velocities. A search algorithm is implemented on the 6D accumulator matrix and a peak is detected at the exact parameters values as mentioned above. We used a range of 125 values of velocities in all three dimensions with step size of 3cm. Experimental analysis showed that the step size of velocities has an inverse relationship with sphere's resolution; low resolution spheres would require larger step size for the 3DVHT to detect spheres correctly.

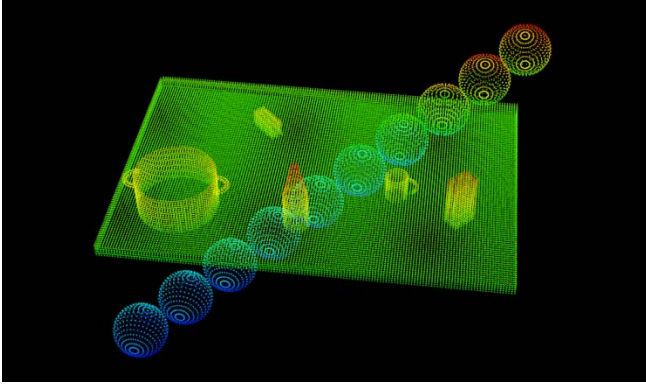


Fig. 1. Synthetic sphere sequence moving with constant 3D velocity.



Fig. 2. Real sequence: football moving with constant 3D velocity



Fig. 3. Optimized global cloud resulting from merging and cropping the sequence in fig 2.

5.2. Real sphere sequences

Figure 2 shows a sequence of five point clouds of a football moving in a constant velocity of 30cm per point cloud in the x direction. These point clouds are generated by the Kinect sensor. The Kinect generates point clouds of approximately 300K points which can be very expensive to handle. To demonstrate basic capability, we down-sample all point clouds and merge them together in a single cloud, keeping a record of which point cloud each point in the global cloud belongs to, further more we “crop” the global point cloud to reduce the volume for the evidence gathering algorithm. Figure 3 shows the resulting optimized point cloud sequence. For further optimization, we restrict the football to move only in one dimension (x). We employ the same approach as we did for synthetic sequence, and this time accumulating only on a four-dimensional accumulator space (as $v_y = v_z = 0$). Using equation 7, the 3DVHT was able to detect the football at any frame (point cloud) with a significant accuracy with root-squared mean error of less than 1cm, the error is due to the internal random error of the Kinect [7]. Figure 4 shows detected spheres at each frame superimposed on the real point cloud sequence. We also derive the results for applying the 3D HT for static spheres on each cloud, separately. To highlight the advantage of the 3DVHT, we perform it on the same point cloud sequence,

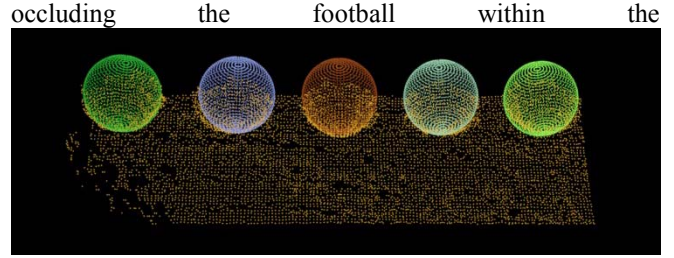


Fig. 4. Real sequence with spheres generated by 3DVHT.

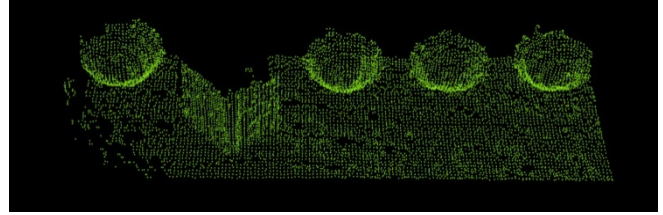


Fig. 5. Sphere occluded in one frame.

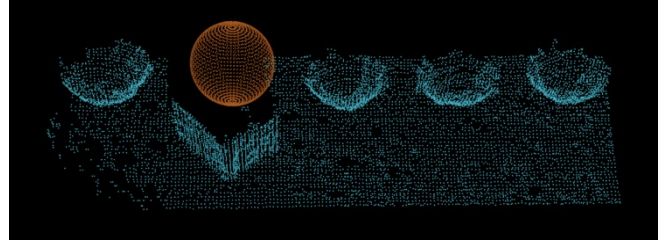


Fig. 6. Detected occluded sphere.

second cloud, figure 5. Of course the static HT will fail completely on the cloud with full occlusion when applied in a cloud-by-cloud basis. When applying the 3DVHT on the sequence with occlusion, we were able to detect the sphere with only 1cm root-squared error more than the un-occluded sequence. Figure 6 shows the detected sphere by the 3DVHT super imposed on the occluded sequence.

6. EVALUATION

By including motion in the 3DVHT, evidence is gathered throughout the 3D sequence in single accumulator space, providing a mean of inter-cloud mapping. This allows exploiting data lost in one point cloud though present in another of the same sequence, such as in the case of occlusion, which gives the 3DVHT its superior performance advantage over 3D tracking techniques which consider point clouds one by one.

For evaluation, we compared the performance of the 3DVHT with the 3D Hough transform for detecting spheres over a real point cloud sequence including an occluded sphere at one point cloud with different amounts of occlusion. The 3DVHT gives a robust and accurate detection performance even if a sphere is fully occluded in any point cloud, presented in figure 6 above. However the tracking performance standard 3D HT applied on cloud-to-cloud bases decreases as the amount of occlusion increases. Figure 7 shows the Euclidean distance error for the

parameters employed to quantify the performance of 3DVHT compared with the standard 3D HT at increasing occlusion ratios. The ratios are quantified by dividing the number of points on the occluded sphere by the number of points on the original non-occluded sphere. The root mean square error is measured by equation (8):

$$error = \sqrt{(x_G - x)^2 + (y_G - y)^2 + (z_G - z)^2} \quad (8)$$

Where x_G , y_G and z_G represent the ground truth coordinates of the sphere's center.

For optimization purposes, we will consider in our future work more sophisticated accumulator design approaches that do not require high dimensionality and provide better peak confidence measures.

7. CONCLUSION

We presented a new global Hough Transform based for detecting spheres undergoing linear motion in a sequence of point clouds, the 3-Dimensional Velocity Hough Transform (3DVHT). Analysis has been performed on synthetic and real point cloud sequences. We demonstrated the significant advantage of the 3DVHT which is the ability to detect fully occluded spheres at any frame of the sequence. The superior performance is attributed to the concept of incorporating motion in the voting process, accumulating evidence in a single accumulator space for the whole sequence, resulting for the spherical structural information in each cloud to be integrated over the complete sequence. This enables motion and structural parameters to be accurately determined, even in the case of a full occlusion.

8. REFERENCES

- [1] D. Borrmann, J. Elseberg, K. Lingemann, and A. Nüchter, "The 3D Hough Transform for Plane Detection in Point Clouds - A Review and A new Accumulator Design," *3D research*, Springer, **2**(2), pp 1-13, 2011.
- [2] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz, "Towards 3D Point Cloud Based Object Maps for Household Environments," *Robotics and Autonomous Systems*, **56**(11), pp. 927–941, November 2008.
- [3] T. Rabbani and F. van den Heuvel, "Efficient Hough Transform for Automatic Detection of Cylinders in Point Clouds," in *ISPRS WG III/3, III/4, V/3 Workshop Laser Scanning*, pp. 60–65, 2005.

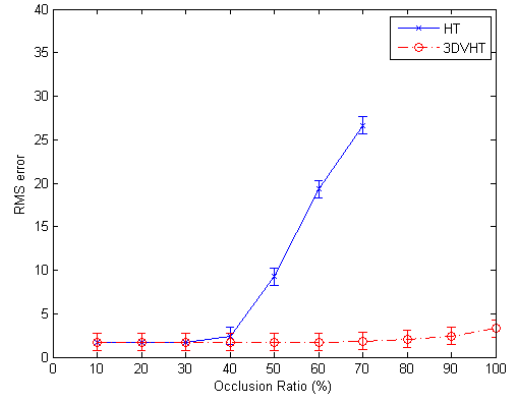


Fig. 7. RMS error metric (in centimeters) of the standard HT and the 3DVHT at increasing occlusion ratio. The HT fails completely to detect a sphere for occlusion ratios above 70%. The 3DVHT was performed with velocity range of 0 to 100 cm/cloud with 3cm steps. This analysis has been performed on a sequence of five point clouds of a football with radius of 11 cm.

- [4] J. M. Nash, J. N. Carter, and M. S. Nixon, "Dynamic Feature Extraction via The Velocity Hough Transform," *Pattern Recognition Letters*, **18**(10), pp. 1035-1047, October 1997.
- [5] P. V. C Hough, "Methods and Means for Recognizing Complex Patterns," US Patent 3069654, 1962.
- [6] O. O. Ogundana, C. R. Coggrave, R. L. Burguete, and J. M. Huntley, "Fast Hough Transform for Automated Detection of Spheres in Three-Dimensional Point Clouds," *Optical Engineering*, **46**(5), pp. 1-11, 2007.
- [7] K. Khoshelham and S. O. Elberink, "Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications," *Sensors*, **12**(2), pp.1437-1454, 2012.