

Multi Scale Shape Index for 3D Object Recognition

Ujwal Bonde, Vijay Badrinarayanan, and Roberto Cipolla

Department of Engineering, University of Cambridge, Cambridge, UK

Abstract. We present Multi Scale Shape Index (MSSI), a novel feature for 3D object recognition. Inspired by the scale space filtering theory and Shape Index measure proposed by Koenderink & Van Doorn [6], this feature associates different forms of shape, such as umbilics, saddle regions, parabolic regions to a real valued index. This association is useful for representing an object based on its constituent shape forms. We derive closed form scale space equations which computes a characteristic scale at each 3D point in a point cloud without an explicit mesh structure. This characteristic scale is then used to estimate the Shape Index. We quantitatively evaluate the robustness and repeatability of the MSSI feature for varying object scales and changing point cloud density. We also quantify the performance of MSSI for object category recognition on a publicly available dataset.

1 Introduction

The availability of cheap IR sensors have considerably lowered the cost of real time 3D data acquisition [11] and has led to a renewed interest in 3D object recognition [13, 5]. This has encouraged research into the development of a number of shape inspired features for 3D, several of which are extensions to popular 2D features [23, 12, 5] and do not directly operate on point cloud data, while others [18, 17] are not robust enough to sensor noise. In this work, we propose a novel feature called the Multi Scale Shape Index (MSSI) which is jointly motivated by scale space filtering theory [21, 10] and the shape categorization work of Koenderink [6]. Shape Index (SI) maps points on surfaces to a linear scale $[-1 : 1]$ and thus classifies them into categories such as Umbilics, Parabolics and Saddle points. Fig. 1 shows a few canonical shapes and their corresponding shape index. The proposed MSSI feature operates directly on a point clouds and are robust to noise in the data.

The SI measure at a 3D point is a function of the principal curvatures at that point. This measure was originally proposed for the continuous domain. However, computing the principal curvatures at a point from noisy 2.5D or 3D data can be erroneous if the characteristic scale at that point is not known (see Fig. 2 for an example). In this work, we show how to compute the characteristic scale at a point in a discrete domain (point cloud) and then estimate the shape index at this scale. We then construct the MSSI feature at a point as a concatenation of its

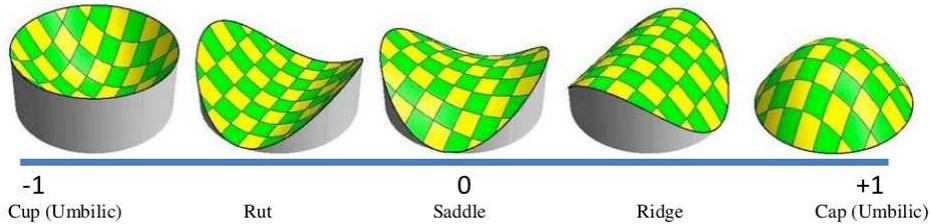


Fig. 1. Illustration of shape index measure mapping shapes to real number. From [6].

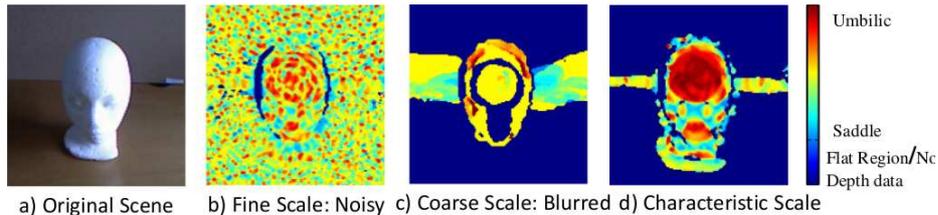


Fig. 2. Effects of computing Shape Index at an erroneous scale for a real world scene. Shape Index computed at the characteristic scale (d) is more stable as compared to one computed at a fine scale (b), which is sensitive to noise, or a coarse scale (c) which blurs high curvature regions. Best viewed in colour.

characteristic scale, shape index and a measure of curvedness [6]. An example of each of these features for a real world scene is shown in Fig. 3. The RGB image and the corresponding depth map from a Kinect sensor is shown in Fig. 3 (a,b). The dummy’s head has a large scale and is classified as an umbilic (doubly convex shape) in Fig. 3 (d). It also has low curvedness as seen in Fig. 3 (e). The tip of the nose has low scale, and is an umbilic with high curvedness. The map of the triplet of these three features is the MSSSI map shown in Fig. 3 (c). To show the efficacy of our proposed feature for category recognition we compare it with the work of Lai *et al.* [7] using their publically available dataset.

The remainder of our paper is organised as follows. In Sec. 2 we discuss relevant literature. The MSSSI feature computation is described in Sec. 3. Our experimental setting and results are elaborated in Sec. 4. We conclude in Sec. 5.

2 Literature Review

There exist many 3D features in literature that try to capture local shape. Recently, Zaharescu *et al.* [23] provided an extension of HOG features for meshes, by binning the directional derivatives of the mean curvature. However, computing the mean curvature in the discrete domain is not straight-forward as we show in section 3. The Heat kernel signature proposed in [18] is based on the fundamental solution to the heat diffusion equation. However this method is sensitive to noise and changes in the configuration of the original mesh. Furthermore, although approaches for triangulating and generating surfaces/meshes from a point cloud do exist, they are slow, noise sensitive or require dense point clouds as pointed out in the survey in [3].

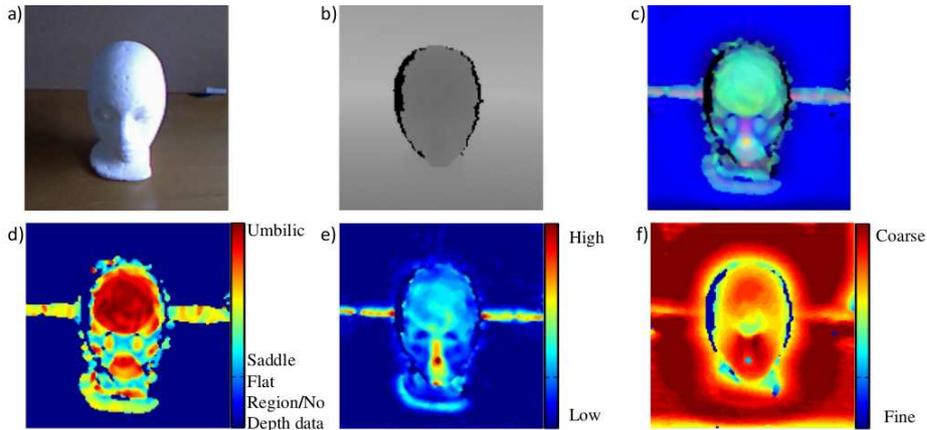


Fig. 3. RGB image of an example scene (a), its depth map obtained from kinect sensor(b). The shape index map (d) categorizes forehead and nose as umbilics, while the nose bridge is estimated as a saddle. Background is correctly assigned as a flat region. The characteristic scale map(e) assigns the nose tip to a fine scale, while the forehead has a relatively coarser scale. The Curvedness map (f) assigns the nose tip to a very high curvedness value while the background has very low curvedness. The three components together gives the MSSSI feature map (c). Best viewed in colour.

Features that operate directly on 3D point clouds do exist [19, 5] and are extensions of popular 2D features(SIFT, SURF). These methods however do not assign stable canonical frames which are needed for them to be rotationally invariant. To address this issue, the authors of [12] provide a method to compute a stable canonical frame. Unlike their previous approach [19] which worked directly on point clouds, this method requires a mesh structure.

Many of the advances made in 2D Object Recognition in the past decade have been adopted for 3D Object Recognition and have shown promising results. Knopp *et al.* [5] extends the Implicit Shape Model (ISM) model proposed by [9] to 3D Object Recognition. They use a Hessian-based interest point detector that encodes an extension of 2D SURF features to 3D [2]. These interest points are then clustered to form the ISM. Their method showed promising results on clean meshed data. However, they do not report any results on real world 3D/2.5D point cloud data. Lai *et al.* combine colour and depth information using pyramid Histogram Of Gradients(HOG) features [7, 4]. These features are used with a linear Support Vector Machine(SVM) to perform sliding window based object detection. They show competitive results using depth and colour features individually, and improve it further by combining both features on one of the largest publicly available 2.5D dataset. Both these methods concentrate on a fusion of depth and intensity/colour features for recognition in real world scenarios. However, there is no explicit attempt to capture shape for recognition. In this work, we propose a local shape based feature (MSSI) to exploit depth data and demonstrate that competitive results can be achieved with lesser training data.

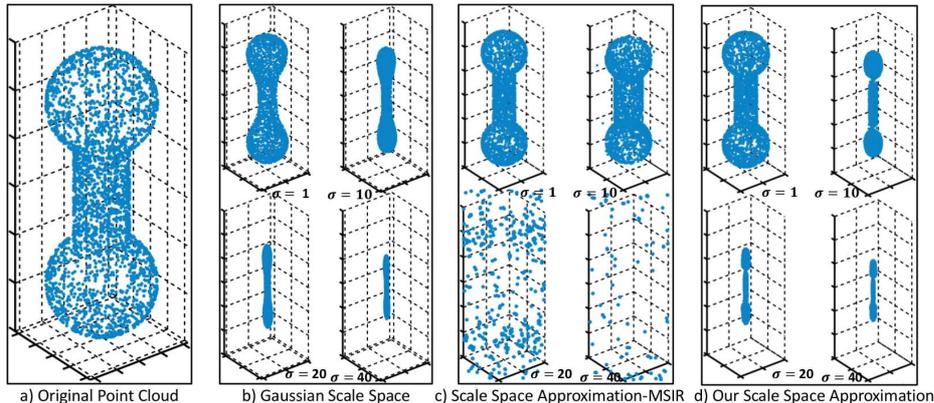


Fig. 4. Scale space approximation for a synthesized dumbbell shape. Our method approximates the original scale space while predictions from MSIR [20] are erroneous.

3 Proposed Multi-Scale Shape Index (MSSI) Feature

Shape index (SI) as proposed by Koenderink [6] is a function of the principal curvatures (κ_1, κ_2) ¹. Principal curvatures are primarily defined for a continuous parameterization of the 3D surface. While there exist methods to approximate them for discrete spaces (point clouds) [15] they require a *support region* to compute them. However, the size of the support region itself is dependent on the principal curvatures. This can be seen from Fig. 4; when the shape index is computed at a fine scale, it is sensitive to noise thus falsely classifying noisy low curvature (flat) regions as umbilics. On the other hand, at a coarser scale, regions of high curvature get blurred out (nose) while low curvature regions are classified correctly. To address this ambiguity in the size of the support region, we propose to obtain a *characteristic scale* automatically by relating the effect of blurring at different scales to the underlying local shape of the point clouds. Our approach is motivated by Multi-Scale Interest Region (MSIR) [20] approach to locate *interest regions*. However, the scale space model in their work is neither accurate for basic shapes nor is stable as shown in Fig. 4 and Fig. 5. In the remainder of this section we derive the relationship between characteristic scale and principal curvature and compare it with MSIR.

Curves in 3D space We start by considering a continuous arc-length parameterized curve $\alpha(s)$ in \mathbb{R}^3 , where $s \in (-B, B)$. Here, B denotes extent of the curve. We define $A : \mathbb{R}^3 \times \mathbb{R}^+ \rightarrow \mathbb{R}^3$ as the family of curves obtained by filtering the original curve at different scales. i.e.,

$$A(\alpha(s), \sigma) = \int_{-B}^B \phi(s-u, \sigma) \alpha(u) du, \quad (1)$$

where, ϕ is the Gaussian kernel. We consider the evolution of a point x as it is filtered. Without loss of generality, we take this point to be $s = 0$ and define

¹ $SI = \frac{2}{\pi} \arctan\left(\frac{\kappa_2 + \kappa_1}{\kappa_2 - \kappa_1}\right) \quad \kappa_1 \geq \kappa_2,$

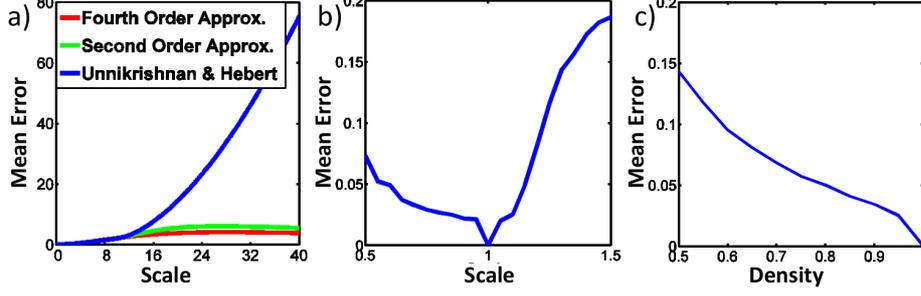


Fig. 5. a) Mean prediction error of the approximated scale space for the synthesized dumbbell shape in Fig 4. MSIR [20] is accurate only at lower scales but our fourth order model approximates quite accurately the Gaussian scale space at all scales. b) Mean estimation error of the computed shape index as we vary the scale for the head model, Fig. 3. Due to the relatively coarser scale of the head a larger error is seen as the scale is increased and a smaller error for lower scales. c) Mean estimation error of the computed shape index as we vary point cloud density is smooth for smaller changes in density. As the density decreases to low values very few points remain to correctly estimate both the characteristic scale as well as the shape index.

$x = \alpha(0)$. Performing a Taylor series expansion of α around x up to fourth order terms, we can approximate the integral as shown below:

$$A(x, \sigma) \approx \frac{1}{\sqrt{2\pi}\sigma} \int_{-B}^B e^{-\frac{u^2}{2\sigma^2}} \left(x + u\alpha'(0) + \frac{u^2}{2!}\alpha''(0) + \frac{u^3}{3!}\alpha'''(0) + \frac{u^4}{4!}\alpha''''(0) \right) du \quad (2)$$

For better readability we set $x' = \alpha'(0)$, $x'' = \alpha''(0)$ and so on. Observing that the second and fourth term in the equation go to zero and performing the integration over the remaining terms we get:

$$A(x, \sigma) = \Phi\left(\frac{B}{\sqrt{2}\sigma}\right) \left(x + x''\frac{\sigma^2}{2} + x''''\frac{\sigma^4}{8} \right) - \sqrt{\frac{2}{\pi}} B\sigma e^{-\frac{B^2}{2\sigma^2}} \left(\frac{x'}{2} + B^2\frac{x''''}{24} + 3\sigma^2\frac{x''''}{24} \right) \quad (3)$$

Using results from differential geometry (see the supplementary material) the above equation can be approximated as:

$$A(x, \sigma) \approx \Phi\left(\frac{B}{\sqrt{2}\sigma}\right) \left(x + \kappa\mathbf{N}\frac{\sigma^2}{2} + (3\kappa'\kappa\mathbf{T} - \kappa^3\mathbf{N})\frac{\sigma^4}{8} \right) - \sqrt{\frac{2}{\pi}} B\sigma e^{-\frac{B^2}{2\sigma^2}} \left(\frac{\kappa\mathbf{N}}{2} + B^2\frac{(3\kappa'\kappa\mathbf{T} - \kappa^3\mathbf{N})}{24} + 3\sigma^2\frac{(3\kappa'\kappa\mathbf{T} - \kappa^3\mathbf{N})}{24} \right), \quad (4)$$

Note that functions κ , \mathbf{T} , \mathbf{N} are evaluated at x which is suppressed for better readability. Here, κ , \mathbf{T} , \mathbf{N} are the curvature, tangent and normal to the curve α and Φ is the error function. If $\sigma \ll B$ the error function can be approximated to 1 and the second term $\rightarrow 0$. This resulting equation is similar to the MSIR model. However, for a bounded curve, as the scale of blurring increases ($\sigma \rightarrow B$) the contribution of the second term and the error function is significant and cannot be ignored (see Fig. 5). We next extend these equations to surfaces in 3D space.

Extension to 3D Surfaces: Let x be a point on a surface M . Further, let the normal at x be denoted as N and its tangent plane as T_x . There then exists

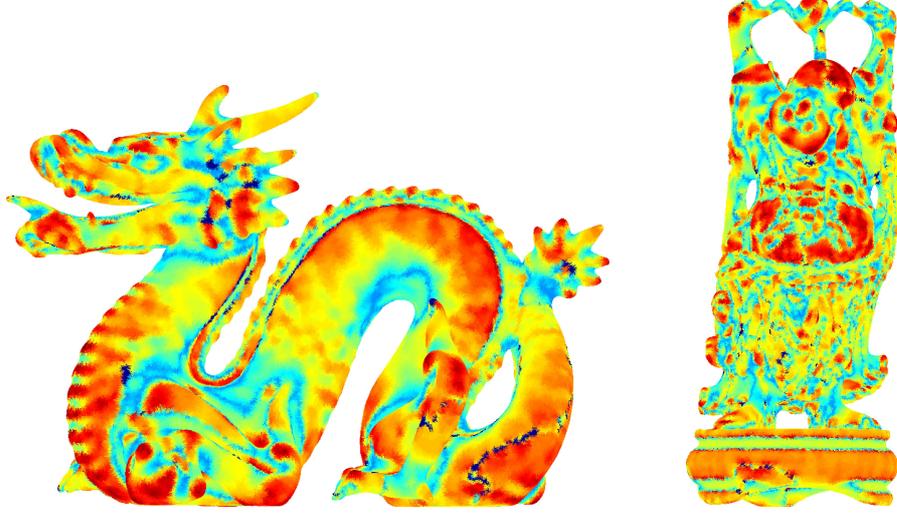


Fig. 6. Computed shape index map on stanford dragon and happy buddha models. Spikes on the back of dragon and the pointed tail are estimated as umbilics. Regions where the dragons body twists are estimated as saddle. The intricate structure on happy buddha produces more variations in the shape index. Best viewed in colour.

a family of planes Π_θ that contain the normal N . The normals of these planes lie in the tangent plane T_x . Let the angle subtended by these planes with the first principal direction be θ [6]. These planes then intersect the surface to give a family of curves $\alpha_\theta(s)$ which are called the *normal sections*. Now, the net *displacement* of the point $x = \alpha_\theta(0)$, after blurring, will be equal to the average displacement caused by each normal section. Using Eq.(1) we have:

$$A(x, \sigma) = \frac{1}{2\pi} \int_0^{2\pi} A(\alpha_\theta(0), \sigma) d\theta = \frac{1}{2\pi} \int_0^{2\pi} \left(\int_{-B_\theta}^{B_\theta} \phi(0 - u, \sigma) \alpha_\theta(u) du \right) d\theta. \quad (5)$$

Solving this equation is not trivial as both (B_θ) and (α_θ) are functions of θ . Moreover, we do not have any explicit form for B_θ which represents the extent of the 3D surface in all directions. Using empirical evidence, we propose setting B_θ to be a constant value B proportional to the average geodesic distance in all directions. In practice, for a given point in a discrete point cloud, this is equal to the average geodesic distance of that point to all other points. From Eqs.(4),(5) we get (see supplementary material for further details):

$$\tilde{A}(x, \sigma) \triangleq \frac{A(x, \sigma)}{\Phi\left(\frac{B}{\sqrt{2}\sigma}\right)} \approx \left(x + HN \frac{\sigma^2}{2} - \frac{H}{16}(5H^2 - 3G)\sigma^4 \mathbf{N} \right), \quad (6)$$

Again, note that functions H , G and \mathbf{N} are evaluated at x which is suppressed for better readability. Here, H , G and \mathbf{N} are the mean curvature, Gaussian curvature and normal to the surface M . $\tilde{A}(x, \sigma)$ is the normalized family of surface obtained from Gaussian blurring with scale σ . This can be further rearranged to give:

$$D(x, \sigma) \triangleq \|\tilde{A}(x, \sigma) - x\|_2 = \left(H \frac{\sigma^2}{2} - \frac{H}{16} (5H^2 - 3G)\sigma^4 \right) = \frac{H\sigma^2}{2} \left(1 - \frac{\sigma^2 H^2}{4} \left(1 + \frac{1.5}{S^2} \right) \right),$$

where, S relates to the shape index via $S = \tan(2\pi \times SI)$, and $D(x, \sigma)$ can be viewed as the approximate distance traveled by a point when blurred with a kernel of scale σ . Inspired by the principle of automatic scale selection, as defined by Lindeberg [10], we define the characteristic scale (σ_{max}) as the maxima in the normalized distance ($\frac{D(x, \sigma)}{\sigma}$) traveled by a point. This is given by $\partial \left(\frac{D(x, \sigma)}{\sigma} \right) / \partial \sigma = 0$, which gives:

$$\sigma_{max}^2 = \frac{4}{3H^2 \left(1 + \frac{1.5}{S^2} \right)}. \quad (7)$$

This derivation can also be carried out in the discrete domain by assuming a uniform point cloud sampling and approximating the integral by a sum.

Eq. (7) relates the shape index to the characteristic scale and thus motivates the term *Multi-Scale Shape Index*. The characteristic scale that we obtain is different from that proposed in MSIR due to two reasons: a) we explicitly consider the curve to be bounded and b) we model the effect of blurring until the fourth order of the Taylor series. Fig. 4 shows an example of a dumbbell shaped point cloud on which we demonstrate the effect of these changes. The actual scale space obtained by Gaussian blurring is shown, along with the prediction obtained using our model and that of MSIR. Fig. 5 shows a quantitative comparison of the mean error for the two models. MSIR is accurate only at lower scales but our fourth order model approximates the Gaussian scale space accurately.

Algorithm 1 gives a stepwise procedure to compute the characteristic scale from a point cloud. As input our algorithm requires the range of scales (σ_k) to search over and an initial smoothing parameter (σ_s) before computing the scale space. We set B to be proportional to the average geodesic distance and call this as the bounding factor (B_{fct}) which is also an input for our algorithm. The estimated characteristic scale is used as a *support region* to compute the shape index. We only calculate the magnitude of the shape index and not its sign. Fig. 6 shows the shape index map on some publically available 3D models.

We compare the robustness and repeatability of the computed shape index against variations in scale and point cloud density. We use a 3D point cloud of the head model used in Fig. 3 for these experiments. As we cannot establish the ground truth, we treat the shape index computed at the original scale and cloud density as the reference.

Change in Scale: We vary the scale from half the original scale to 1.5 times the original scale. The left panel in Fig. 5 plots the resulting deviation from the ground truth. All parameters (σ_k , σ_s , B_{fct}) are kept constant. Since the head is of a relatively coarser scale and the initial smoothing is kept constant, a higher rate of deviation from the ground truth is seen as we increase the size of the head model. On the other hand, decreasing the scale of the model while keeping the initial smoothing constant does not affect the coarser scale regions and thus a lower deviation from the ground truth is observed in this case.

Change in Density: We vary the density from the original density to half its density. The right panel in Fig. 5 plots the resulting deviation from the ground

```

input : Point Cloud:  $X = \{x_i\}$ , Range of Scales:  $\sigma_k$ , Bounding factor:  $B_{fct}$ ,
        initial smoothing:  $\sigma_s$ 
output: Characteristic Scale for each point  $\sigma_{max}(x_i)$ 
1. Compute a Disjoint Minimum Spanning Tree on  $X$  to form a graph  $\mathbb{G}$ .
2. Using Dijkstra's algorithm approximate the graph distance between points,
    $d_{\mathbb{G}}$ , as the geodesic distance.
3. Calculate the average geodesic distance for each point  $d_{avg}$ .
foreach  $x_i$  do
4.  $\hat{x}_i \leftarrow \frac{1}{n_{x_i}} \sum_j \exp\left(-\frac{d_{\mathbb{G}}^2(x_i, x_j)}{2\sigma_s^2}\right) x_j$ , where  $n_{x_i}$  is the normalizing factor.
   // Initial Smoothing
   foreach  $\sigma_k$  do
5.  $A(x_i, \sigma_k) \leftarrow \frac{1}{n_{x_i}} \sum_j \exp\left(-\frac{d_{\mathbb{G}}^2(x_i, x_j)}{2\sigma_k^2}\right) x_j$ 
6.  $\tilde{A}(x_i, \sigma_k) \leftarrow A(x, \sigma) \times \left(\Phi\left(\frac{B_{fct} * d_{avg}(x_i)}{\sqrt{2}\sigma_k}\right)\right)^{-1}$ 
7.  $D(x_i, \sigma_k) \leftarrow \|\tilde{A}(x, \sigma) - x\|_2$ 
   end
8.  $\sigma_{max}(x_i) \leftarrow \max_{\sigma_k} D(x_i, \sigma_k)$ 
end

```

Algorithm 1: Computation of the characteristic scale for point clouds

truth averaged over 10 different trials. Once again $\sigma_k, \sigma_s, B_{fct}$ are kept constant. A smooth deviation from the ground truth is observed as the density is reduced to 3/4 its original density. As the density decreases to low values very few points remain to correctly estimate both the characteristic scale as well as the shape index and thus a higher deviation from ground truth is observed at low densities.

3.1 Object Recognition with the MSSSI feature

The shape index alone does not capture all the information about the underlying shape [6]. Being a ratio of the principal curvatures, it does not provide any information about the magnitude of the curvatures. For example, a tennis ball and a football are both spherical, but have completely different size with the tennis ball having a higher magnitude of the principal curvatures compared to a football. This notion is captured by the curvedness measure proposed by Koendrink². The characteristic scale is used as another feature to capture the scale and thus we form a triplet of features, which we call the Multi-Scale Shape Index (MSSI) feature.

Detecting interest points, followed by a bag-of-visual-words approach is a common strategy in 2D object recognition [1, 8]. However in 3D, as reported in the survey by [22], corner detectors are relatively less robust to noise compared to region based methods. We therefore follow a region based approach to object shape encoding. We start by super-pixelizing the MSSSI feature map. We use the fast and efficient SLIC super-pixels [14]. Fig. 7 shows an example of super-pixels for different viewpoints of the head model. As seen from the images, these super-pixels are fairly stable across viewpoints. This empirical observation motivates us to use super-pixels for category recognition. To further capture the variations of

² $curvedness = \sqrt{\frac{\kappa_1^2 + \kappa_2^2}{2}}$.

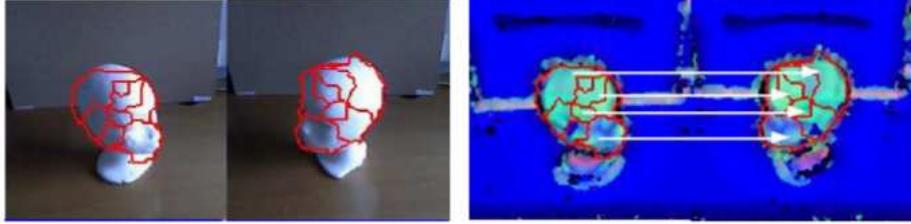


Fig. 7. An illustration of the stability of MSSSI feature based superpixels across view-points. Super-pixels with similar MSSSI features appear at approximately the same relative location in both view points. See supplementary material for an example on cluttered scenes. Best viewed in colour.

shape within a superpixel, we also include the angle between each pixel normal and its corresponding *superpixel normal*³. Although this is correlated to the variation in shape index, empirically we observed that it improves recognition rate by introducing redundancy.

We cluster the concatenation of MSSSI features and normals at each pixel into a preset number of clusters⁴. The super-pixel descriptors are obtained by binning the MSSSI+normal features for each of their pixels. As these super-pixels using MSSSI features, the resulting super-pixel descriptors are very sparse. Therefore to enrich the descriptor of a super-pixel, we compute a weighted average of descriptors of super-pixels that are at most two hops away from it (1 and 2 neighbourhood in a graph sense). The weights used are proportional to the depth difference between the super-pixels.

We train our super-pixel based recognition approach using an RBF kernel SVM. We use a 1-vs-all setup. The super-pixel in the test set are classified individually during testing. The resulting classification gives us an initial region of interest for possible object locations. Thresholding on the number of connected pixels within these region of interests gives the final object detection.

4 Experiments, Results and Discussion

Many 3D object recognition datasets have been introduced in the recent years [16, 13, 7]. Of these, one of the largest is the RGB-D dataset [7]. We compare our recognition algorithm using MSSSI features with the pyramid hog based depth features of Lai *et. al* [7]. We used the original authors code to obtain their results.

Dataset: The RGB-D dataset [7] contains challenges for both instance level as well as category level object recognition and detection. We perform our experiments for the category level object detection. For the category case, five items are ground truth-ed by the authors. Of these we choose four categories: cap, coffee mug, flashlight and soda can. We do not choose the bowl category since there is a large variation in the size of the bowl category. This results in a large variation in the characteristic scale (and thus in MSSSI feature) which is difficult to capture with limited training instances per category. Each category has 4 or

³ Mean of normals of all pixels within it.

⁴ We empirically fixed this to 300.

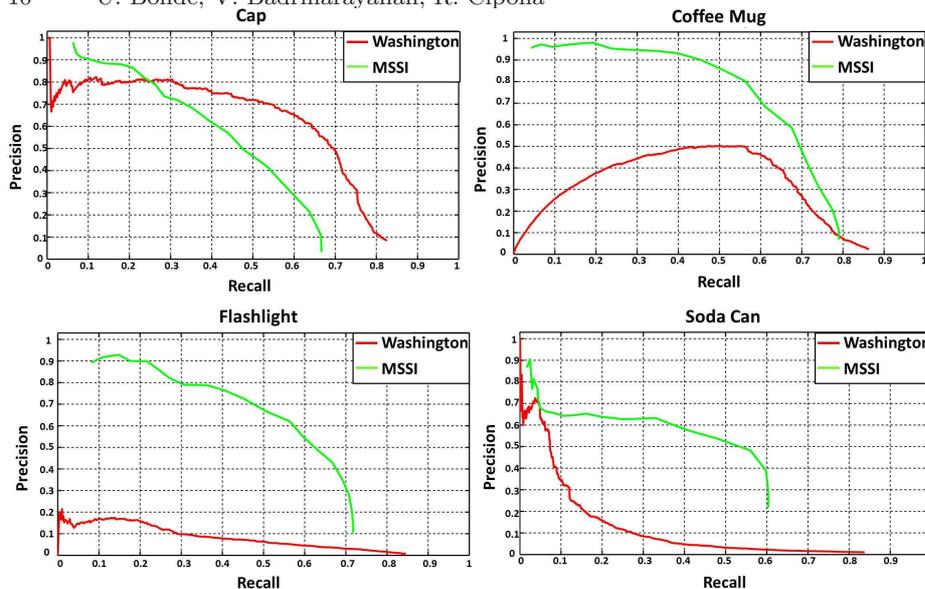


Fig. 8. Precision/Recall curves for our MSSSI feature based category recognition and its comparison with depth features based recognition of Lai *et. al* [7] (shown as Washington in plots). MSSSI features based recognition clearly outperforms the method of Lai *et. al* in the coffee mug and flashlight classes. For the cap class, we are only marginally worse off in performance, and for the soda can category, although our performance is relatively better than Lai *et. al*, both methods suffer due to small size of the object.

more instances and we train on only 2 instances. The training set contains about 600-1000 depth and RGB (which we do not use) images of 640x480 resolution each captured on a turntable at 3 different angles. As mentioned earlier since shape is fairly constant with small changes in viewing angles, we use only 1/3rd of the training data. The test set contains 8 video sequences with 98-230 frames per sequence. The number of objects in each sequence varies as does the clutter. We currently use 320x240 resolution images to process this large dataset and hence the results quoted in [7] are different from those computed here. At this resolution, we consider minimum object size to be at least 1000 pixels. Varying the threshold on the number of connected pixels we plot the resulting Precision-Recall (P-R) curve in Fig. 8. The qualitative recognition results of our system are shown in Fig.9.

We downsampled the dataset depth images by half to 320x240 resolution to speed up computation. This particularly affects performance on small object categories (soda can). We expect to perform better when our system is scaled up to a larger resolution of images. From Fig. 8 we see that our MSSSI features based recognition easily outperforms the method of [7] in two classes (coffee mug and flashlight). MSSSI features cope better with reduced training data and lower image resolution. For the soda can category, although our performance is



Fig. 9. Sample results of our object recognition and segmentation system. Our methods performs well in the presence of partial occlusion (cap), clutter and change in object pose. More results can be found in the supplementary material. Best viewed in colour.

relatively better than [7], both methods suffer due to the small size of the object. Only on the cap class, we are marginally worse off in performance.

We now discuss the effect of some of the influential parameters of our system:

Effect of Bounding Factor B_{fct} : Setting B_{fct} to a large value we obtain only coarse scale changes and mask the effect of smaller scale objects in the scenes. On the other hand by setting it to a small value we obtain small neighbourhood scale changes which mostly originate from the sensor noise present in the data. In general we found the average geodesic distance (or a fraction of it) to be a good approximation.

Weighting of super-pixel neighbourhoods To form the final descriptor for a super-pixel, we compute a weighted average of individual super-pixel descriptors in its 2-neighbourhood. We found the performance to vary based on the weights that were assigned to the neighbourhood super-pixels. In our experiments, we used a Gaussian weighting, based on depth difference between the super-pixels. We set the standard deviation for the 1-neighbours to 20 and 10 for the 2-neighbours in our experiments.

The current algorithm is computationally expensive, for example it takes about 15 minutes on the stanford dragon model on a single core CPU with our unoptimized code.

5 Conclusions

In this work we presented a novel shape based feature called multi scale shape index (MSSI). This feature is a triplet of shape index, curvedness and characteristic scale. The shape index component of this feature assigns a real valued index to shapes such as umbilics (double convex), parabolics (double concave) and saddle points (convex-concave). We developed a scale-space method to compute MSSI at each discrete point at its characteristic scale from noisy 2.5D data. We studied the robustness and repeatability of this feature and demonstrated its efficacy in category recognition. Our quantitative studies indicate that the MSSI feature based recognition outperforms the current state-of-the-art method and is better able to cope with lesser training data.

6 Acknowledgements

This research is supported by the Boeing Company.

References

1. S. Agarwal, A. Awan, and D. Roth. Learning to detect objects in images via a sparse, part-based representation. *TPAMI*, 26(11):1475–1490, 2004.
2. H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. In *ECCV*, 2006.
3. R. Fabio. From point cloud to surface: the modeling and visualization problem. In *International Workshop on Visualization and Animation of Reality-based 3D Models*, volume 34, page 5, 2003.
4. P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. In *CVPR*, 2008.
5. J. Knopp, M. Prasad, G. Willems, R. Timofte, and L. V. Gool. Hough transform and 3d surf for robust three dimensional classification. *ECCV*, 2010.
6. J.J. Koenderink and A.J. van Doorn. Surface shape and curvature scales. *Image and Vision Computing*, 10(8):557–564, 1992.
7. K. Lai, L. Bo, X. Ren, and D. Fox. A large-scale hierarchical multi-view rgb-d object dataset. In *ICRA*, 2011.
8. S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, 2006.
9. B. Leibe, A. Leonardis, and B. Schiele. Combined object categorization and segmentation with an implicit shape model. In *ECCV*, 2004.
10. T. Lindeberg. Feature detection with automatic scale selection. *IJCV*, 30(2):79–116, 1998.
11. R.A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A.J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *ISMAR*, 2011.
12. A. Petrelli and L. Di Stefano. On the repeatability of the local reference frame for partial shape matching. In *CVPR*, 2011.
13. M.T. Pham, O.J. Woodford, F. Perbet, A. Maki, B. Stenger, and R. Cipolla. A new distance for scale-invariant 3d shape recognition and registration. In *ICCV*, 2011.
14. R. Achanta and A. Shaji and K. Smith and A. Lucchi and P. Fua and S. Süsstrunk. Slic superpixels. *Technical Report 149300 EPFL*, (June), 2010.
15. R. B. Rusu and S. Cousins. 3d is here: Point cloud library (pcl). *ICRA*, 2011.
16. S. Savarese and L. Fei-Fei. 3d generic object categorization, localization and pose estimation. In *ICCV*, 2007.
17. T.B. Sebastian, P.N. Klein, and B.B. Kimia. Recognition of shapes by editing their shock graphs. *TPAMI*, 26(5):550–571, 2004.
18. J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Computer Graphics Forum*, 2009.
19. F. Tombari, S. Salti, and L. Di Stefano. Unique signatures of histograms for local surface description. *ECCV*, 2010.
20. R. Unnikrishnan and M. Hebert. Multi-scale interest regions from unorganized point clouds. In *CVPR Workshop on Search in 3D (S3D)*, 2008.
21. A.P. Witkin. Scale-space filtering. *Readings in computer vision: issues, problems, principles, and paradigms*, pages 329–332, 1987.
22. T.H. Yu, O.J. Woodford, and R. Cipolla. An evaluation of volumetric interest points. In *3DIMPVT*, 2011.
23. A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud. Surface feature detection and description with applications to mesh matching. In *CVPR*, 2009.