

# A Performance Evaluation of Volumetric 3D Interest Point Detectors

Tsz-Ho Yu · Oliver J. Woodford · Roberto Cipolla

Received: 4 November 2011 / Accepted: 17 August 2012  
© Springer Science+Business Media, LLC 2012

**Abstract** This paper presents the first performance evaluation of interest points on scalar volumetric data. Such data encodes 3D shape, a fundamental property of objects. The use of another such property, texture (i.e. 2D surface colouration), or appearance, for object detection, recognition and registration has been well studied; 3D shape less so. However, the increasing prevalence of 3D shape acquisition techniques and the diminishing returns to be had from appearance alone have seen a surge in 3D shape-based methods. In this work, we investigate the performance of several state of the art interest points detectors in volumetric data, in terms of repeatability, number and nature of interest points. Such methods form the first step in many shape-based applications. Our detailed comparison, with both quantitative and qualitative measures on synthetic and real 3D data, both point-based and volumetric, aids readers in selecting a method suitable for their application.

**Keywords** 3D interest points · Volumetric interest points · Feature detection · Performance evaluation

## 1 Introduction

The applications of object detection, recognition and registration are of great importance in computer vision. Much

work has been done in solving these problems using appearance on 2D images, helped by the advent of image descriptors such as SIFT and learning-based classifiers such as SVM, and these methods are now reaching maturity. However, advancing geometry capture techniques, in the form of stereo, structured light, structure-from-motion and sensor technologies such as laser scanners, time-of-flight cameras, MRIs and CAT scans, pave the way for the use of shape in these tasks, either on its own or complementing appearance—whilst an object’s appearance is a function not only of its texture, but also its pose and lighting, an object’s 3D shape is invariant to all these factors, providing robustness as well as additional discriminative power.

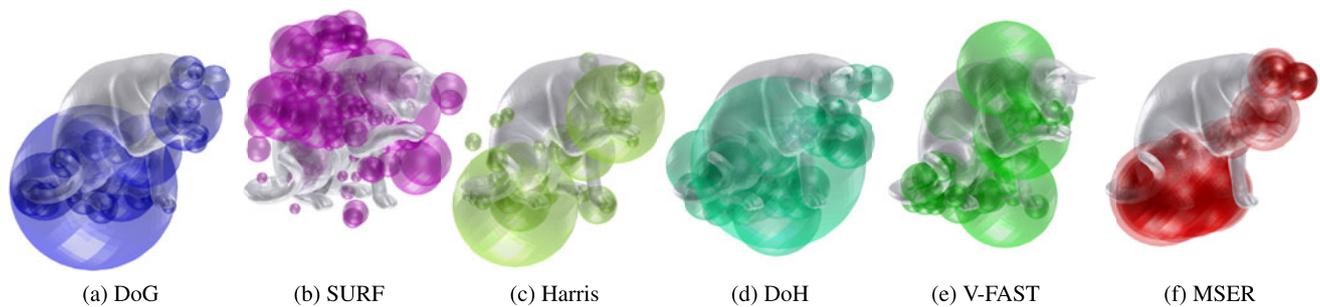
Detection and recognition of 3D objects is not new, e.g. (Fisher 1987), though such applications have seen a recent resurgence. Approaches range from the local to the global. At the global end are those which form a descriptor from an entire object. Such methods generally offer excellent discrimination plus robustness to shape variation, but, since the whole object is required and its extent known, they do not cope well with clutter or occlusion. Also, whilst suitable for recognition, the matching does not provide pose for registration applications. At the other end of the scale are highly local features, such as points. Being completely non-discriminant, such features are usually embedded in a framework that finds geometrical consistency of features across shapes, e.g. RANSAC (Brown and Lowe 2005; Papazov and Burschka 2011). The geometrical consistency framework makes matching, for detection and recognition, costly, but does provide pose for registration applications, and the local nature of features provides robustness to clutter and partial occlusion. Between these two extremes are those methods that describe local features of limited but sufficiently distinctive scope, thereby gaining the discriminability of global methods and the robustness of local methods.

---

T.-H. Yu (✉) · R. Cipolla  
University of Cambridge, Cambridge, UK  
e-mail: [thy23@cam.ac.uk](mailto:thy23@cam.ac.uk)

R. Cipolla  
e-mail: [cipolla@eng.cam.ac.uk](mailto:cipolla@eng.cam.ac.uk)

O.J. Woodford  
Toshiba Research Europe Ltd., Cambridge, UK  
e-mail: [oliver.woodford@crl.toshiba.co.uk](mailto:oliver.woodford@crl.toshiba.co.uk)



**Fig. 1** Different types of volumetric interest points detected on a test shape

The distribution of such features can be used for effective object detection, recognition and registration. The nature of these hybrid methods is reminiscent of the image descriptors of appearance-based methods, not least in the need for shape features to be chosen at points that are repeatably locatable in different datasets, and whose localities are distinctive. A common and crucial stage of such approaches is therefore the detection of *interest points* to be described.

This paper aims to conduct a performance evaluation of interest point detectors on scalar volumetric data, as shown in Fig. 1. Different from other data-specific 3D interest point detectors for meshes (Sipiran and Bustos 2011; Gomb 2009; Zaharescu et al. 2009) or point-clouds (Aanæs et al. 2010; Unnikrishnan and Hebert 2008), feature detection from scalar volumetric data is more versatile. Such data not only comes directly from volumetric sensors, e.g. MRIs, but can also be generated or converted from other three dimensional data such as point clouds, meshes or depth maps, making the evaluation result widely applicable. In addition, visual saliency of volumetric interest points is defined in a scalar volume but not on a local surface patch. This full three dimensional representation implies that interest points can be located off an object's surface, e.g. inside a cavity. Furthermore, the nature of the data—voxels, the 3D equivalent of pixels—makes repurposing the many 2D interest point detectors for 3D straightforward. The primary quantitative evaluation criterion used here is a novel measure combining both repeatability, based on the number of corresponding points found across two volumes, and the spatial accuracy of correspondences. Detected interest points have a (sub-voxel) location and a scale, and distances are computed in this space. This paper also presents a generic evaluation of qualitative characteristics of the interest points.

The following section reviews previous work relevant to interest point detectors and our evaluation framework. Section 3 introduces the interest point detectors used in the evaluation experiments, while Sect. 4 explains the proposed evaluation methodology. Section 5 presents the evaluation experiments and their corresponding quantitative and qualitative analyses, before we conclude in Sect. 6.

## 2 Previous Work

### 2.1 Interest Point Detectors

The process of interest point detection is the first stage of many computer vision applications, including object detection, recognition and reconstruction. An interest point detector localizes salient points from input visual data for further processing. The detected interest points are typically used to match corresponding points across two or more similar sets of data.

The majority of earlier studies focus on detecting features on 2D images. The paradigm of 2D interest point detection is now well studied; we refer the reader to a recent survey (Tuytelaars and Mikolajczyk 2008) for further details. Recent advancements in data acquisition techniques have greatly improved the availability of 3D shape data. Large scale synthetic and real 3D repositories, such as Google Warehouse (Lai and Fox 2010) and the B3DO dataset using the Kinect sensor (Janoch et al. 2011), have attracted much interest in 3D shape-based computing vision applications. Consequently, various 3D interest point detection techniques have been proposed alongside with this emerging field of computer vision research.

Existing techniques for 3D interest point detection can be categorized as volume-based or geometry-based detectors, according to the representation of input data. Volume-based detectors operate directly on the pixel/voxel values of volumetric scalar data. This kind of data includes CT scan volumes (Flitton et al. 2010), binary volumes generated from range data (Vikstén et al. 2008) or 3D meshes (Knopp et al. 2010), and space-time video data (Koelstra and Patras 2009; Laptev 2005; Willems et al. 2008; Yu et al. 2010). Geometry-based interest point detectors extract geometric information (e.g. contours, normals or surface patches) and find interest points based on these features. Their input data are usually synthetic meshes (Gomb 2009; Sipiran and Bustos 2011; Zaharescu et al. 2009) or point clouds (Unnikrishnan and Hebert 2008; Aanæs et al. 2010). For geometry-based interest point detectors, recent evaluations have been reported in Salti et al. (2011) and Dutagaci

et al. (2011). Nevertheless, unlike 2D interest points, performance evaluation of 3D interest points remains a largely unexplored topic. This paper aims at the quantitative and qualitative evaluation of volumetric interest point detectors, whose high versatility w.r.t. data representation enables a much wider coverage of potential applications.

The remainder of this section will review the interest point detectors used in our evaluation, which we divide into two classes based on their definitions of local features.

### 2.1.1 Corner Detection

The first class of interest point detectors aim to find corners (i.e. areas of high change in gradient in orthogonal directions) in the input data. The Harris interest point detector (Harris and Stephens 1988) is a classic example of corner detection in 2D, which is still widely used today. It detects interest points by analyzing the eigenvalues of the second moment matrix (first order derivative). Its 3D adaptation has been applied to registration of volumetric CT scans (Ruiz-Alzola et al. 2001; Dalvi et al. 2010). Building on the success of the traditional Harris detector, Mikolajczyk (2004) developed the scale-covariant<sup>1</sup> Harris-Laplace detector by finding Harris corners in the spatial domain which are maxima of the Laplacian in the scale domain. This approach has been extended to space-time interest points for video classification (Laptev 2005). The SUSAN detector (Smith and Brady 1997) uses the proportion of pixels in a neighbourhood which are dissimilar to the central pixel to classify corners. The FAST keypoint detector (Rosten et al. 2010) uses the *accelerated segment test* (AST), a relaxed version of SUSAN, for stable corner detection. FAST measures the largest number of contiguous pixels on a circle which are significantly darker, or brighter, than the centre pixel. Without computing the derivative at each pixel, the speed of FAST can be further improved by learning a decision tree classifier for feature detection. Thanks to its efficient run-time performance, several volumetric feature detectors have been applied to space-time volumes classification based on FAST interest points (Koelstra and Patras 2009; Yu et al. 2010).

### 2.1.2 Blob Detection

Lindeberg (1998) studied scale-covariant interest points using the Laplacian-of-Gaussian kernel (equivalent to the trace of the Hessian), as well as the determinant of the Hessian (DoH). Lowe (2004) approximated the former with a Difference-of-Gaussians (DoG) operator for efficiency. Recently the DoG approach has been applied in 3D, to object

detection and recognition of synthetic meshes (Wessel et al. 2006), volumetric scans (Flitton et al. 2010) and multi-view stereo data (Pham et al. 2011).

The Hessian-Laplace detector is similar to Harris-Laplace detector; interest points are detected by computing the Hessian matrix from the input data (Mikolajczyk 2004). The SURF detector (Bay et al. 2008) accelerates computation of the determinant of Hessian through the use of integral images and box filters, since applied to integral volumes of videos (Willems et al. 2008) and binary volumes generated from synthetic 3D mesh models (Knopp et al. 2010).

While both DoG and SURF are grounded on the approximation of the Laplacian-of-Gaussian kernel, the *Maximally Stable Extremal Regions* (MSER) interest point (Matas et al. 2004) finds thresholded regions whose areas are maximally stable as the threshold changes. It is therefore inherently multi-scale, as well as invariant to affine intensity variations and covariant with affine transformations. Three dimensional MSER has already been applied to volumetric data, firstly in the context of segmentation of MRIs (Donoser and Bischof 2006), then on spatio-temporal data (Riemenschneider et al. 2009).

### 2.1.3 Covariant Characteristics

Image-based detectors have been made affine-covariant, in order to approximate the perspective distortion caused by projection of 3D world onto the 2D image plane (Mikolajczyk and Schmid 2002). Such covariance is *not necessary* with 3D shape data because most shape acquisition techniques are invariant to view point changes, thus affine transformations are not common among datasets. However, objects might have varying poses during data acquisition (i.e. translation, rotation and scaling), thus rotation and scale covariance are still essential for processing 3D shape data. In addition, 3D shape data are generally not affected by illumination and lighting conditions, but the quality of shape data is instead determined by the amount of noise and sampling artifacts (e.g. holes and occlusions) of the reconstruction process.

## 2.2 Methodologies

Empirical performance evaluation is a popular pastime in computer vision, and the topic of interest point<sup>2</sup> detection is no exception. Different approaches of performance evaluations can be categorized according to the evaluation criteria and the source of ground truth interest point locations.

<sup>1</sup>Covariant characteristics, often (inaccurately) referred to as *invariant* characteristics, undergo the same transformation as the data. We prefer “covariant” in order to distinguish truly invariant characteristics.

<sup>2</sup>When referring to interest points in the context of methodology, we include image features such as corners, lines, edges and blobs.

### 2.2.1 Evaluation Criteria

Some methods evaluate performance in the context of a particular task, e.g. object recognition (Shin et al. 1999; Dutagaci et al. 2011), lacking generality to other applications. Most evaluation frameworks investigate one or more interest point characteristics. One such characteristic, important for registration applications, e.g. camera calibration, scene reconstruction and object registration, is the accuracy of interest point localization. Coelho et al. (1992) measure accuracy by computing projective invariants and comparing these with the actual values measured from the scene. Three further measures, including 2D Euclidean distance from detected points to ground truth corner locations (given by line fitting to a grid), are introduced in Brand et al. (1994). This approach has been extended to using the distance to the nearest point detected in, and transformed from, another image, e.g. Schmid et al. (2000). Matching scores for interest points found over location and *scale* (Laptev and Lindeberg 2003) and also *affine* transformations (Mikolajczyk 2004), have since been proposed.

When used for object detection and recognition, two other important characteristics of interest points are their repeatability and distinctiveness. Repeatability is the geometrical stability of the corresponding interest points among multiple input data taken under varying conditions. It was proposed and defined by Schmid et al. (2000) as the ratio of repeated points to detected points. Rosten et al. (2010) used the area under the repeatability curve as a function of number of interest points, varied using a threshold on the detector response, in their evaluation. When ground truth locations of interest points are given, e.g. hand labelled, an alternative measure is the ROC curve (Bowyer et al. 1999), which takes into account false matches. Schmid et al. (2000) also introduce a quantitative measure of distinctiveness—entropy, or “information content”. Alternatively, qualitative visual comparison is used, on test datasets containing a variety of different interest points (Lindeberg 1998; Laptev 2005).

In the context of image-based interest points, performance of detectors are often measured over variations in image rotation, scale, viewpoint angle, illumination and noise level, e.g. Schmid et al. (2000) covers all these factors, as well as corner properties (Rajan and Davidson 1989). Efficiency may be a further consideration for applications in which run-time performance is a major concern (Rosten et al. 2010). Evaluations of 3D interest points not only vary on the above-mentioned criteria, but also on the type of data, such as meshes, space-time volumes, point clouds and space volumes.

It is also worth noting that the distinction between interest point *detectors* and *descriptors*. The latter topic, also well evaluated in 2D, e.g. (Mikolajczyk et al. 2005), has a

concept of both correct and incorrect matches, allowing the use of recall-precision as an evaluation criterion.

### 2.2.2 Ground Truth Data

With both localization accuracy and repeatability criteria, the ground truth location of interest points in the scene must be known. The ground truth data can be computed in a variety of ways. Some methods specify the location of interest points in an image, either known by design (Rajan and Davidson 1989), or hand labelled by multiple people (Heath et al. 1997). Other methods match points detected across two or more images. Matching is achieved using planar scenes and computing homographies between images (Schmid et al. 2000), scenes of known geometry manually registered in each image (Rosten et al. 2010), scene geometry captured using structured light (Aanæs et al. 2010), and synthetic data (Laptev 2005). Ground truth data for 3D interest point evaluations are likewise obtained from manual annotation (Dutagaci et al. 2011), known projection homography of stereo point clouds (Aanæs et al. 2010) and synthetic shape data (Salti et al. 2011).

## 3 Detectors

Various volumetric interest points have been proposed in applications such as shape retrieval and classification (Riemenschneider et al. 2009; Flitton et al. 2010; Knopp et al. 2010; Prasad et al. 2011), medical imaging (Criminisi et al. 2010; Ni et al. 2008; Donner et al. 2011) and video-based object recognition (Willems et al. 2009; Laptev 2005; Yu et al. 2010). Whilst interest point detectors for images have already been studied extensively (Mikolajczyk et al. 2005), evaluation of volumetric interest points remains largely unexplored.

This section briefly describes the principles and formulations of the volumetric interest point detectors that we will evaluate. These include DoG (Flitton et al. 2010), DoH and Harris-based interest points (Laptev 2005), SURF (Willems et al. 2008; Knopp et al. 2010), V-FAST (Yu et al. 2010) and MSER (Donoser and Bischof 2006; Riemenschneider et al. 2009).

### 3.1 Scale-Space and Subpixel-Refinement

Scale covariance of interest point detectors is achieved by creating the scale-space of the input volumetric data. An octave of linear scale-space is created by convolving the input volume with a Gaussian smoothing kernel. Such smoothing kernel is applied on the volume recursively to suppress fine-scale structures. In addition, a new octave is created by down-sampling the input volumes from the previous octave.

Hence, a series of volumes, with multiple levels of details, is created. The detailed implementation of scale-space, with respect to interest point detection, can be found in Lindeberg (1998).

Scale-space representation is not necessary for MSER because it detects salient regions in different scales. MSER locates interest points by fitting an ellipsoid to the detected salient region (Matas et al. 2004). For other interest point detectors, saliency responses are computed in all volumes within the scale-space. In addition, the subpixel refinement process of Lowe (2004) is applied on these detectors; interest points are localized at the subvoxel level by fitting 4D quadratic functions around the local scale-space maxima, and selecting the maxima of those functions instead.

### 3.2 Difference-of-Gaussians (DoG)

The DoG operator is a blob detection technique for feature localization popularized by the SIFT algorithm (Lowe 2004). DoG approximates the Laplacian of Gaussian filter, which detects features of a particular size. The saliency response of DoG detector  $S_{DoG}$  is computed by subtracting two Gaussian smoothed volumes, usually adjacent scale-space representations, of the same signal and taking the absolute values of this. Interest point are detected at the *4D local maxima* (both 3D space and scale) in  $S_{DoG}$  within each octave of  $\mathbf{V}(\mathbf{x}, \sigma_s)$ :

$$S_{DoG}(x, y, z; \sigma_s) = |V(x, y, z; \sigma_s) - V(x, y, z; \sigma_{s-1})| \quad (1)$$

where  $V(x, y, z; \sigma_s)$  is the scale-space representation of the input volumetric data at scale  $\sigma_s$ .

### 3.3 Harris

Harris corner detector examines changes of intensity due to shift in a local window, interest points are detected at positions where large changes are observed in all directions (Harris and Stephens 1988). While the first 3D extension (Laptev 2005) of the traditional Harris corner detector uses separate scale parameters for the heterogeneous space and time axes, here one scale parameter  $\sigma_s$  is shared among three homogeneous spatial axes. The second-moment matrix  $M$  is computed by smoothing the first derivatives of the volume in scale-space  $\mathbf{V}(\mathbf{x}; \sigma_s)$  by a spherical Gaussian weight function  $g(\cdot; \sigma_{Harris})$ , thus:

$$\begin{aligned} V_x(\mathbf{x}; \sigma_s^2) &= \frac{\partial \mathbf{V}(\mathbf{x}; \sigma_s^2)}{\partial x} \\ V_y(\mathbf{x}; \sigma_s^2) &= \frac{\partial \mathbf{V}(\mathbf{x}; \sigma_s^2)}{\partial y} \\ V_z(\mathbf{x}; \sigma_s^2) &= \frac{\partial \mathbf{V}(\mathbf{x}; \sigma_s^2)}{\partial z} \end{aligned} \quad (2)$$

$$M = g(\cdot; \sigma_{Harris}) * \begin{bmatrix} V_x^2 & V_x V_y & V_x V_z \\ V_x V_y & V_y^2 & V_y V_z \\ V_x V_z & V_y V_z & V_z^2 \end{bmatrix} \quad (3)$$

where  $V_x, V_y, V_z$  denote the partial derivatives of the volume in scale-space  $\mathbf{V}(\mathbf{x}; \sigma_s)$  along  $x, y$  and  $z$  axes respectively. The matrix  $M$  describes the autocorrelation along different directions in a local neighbourhood of size  $\sigma_s$ .

The saliency  $S_{Harris}$  is computed from the determinant and trace of  $M$ , as follows:

$$S_{Harris} = \sigma_s^3 \det(M) - k \text{trace}(M)^3 \quad (4)$$

A user defined threshold  $k$  controls the rejection of edge points. Each saliency response  $S_{Harris}$  is normalized by its scale  $\sigma_s$ . The window size  $\sigma_{Harris}$  is proportional to expected feature scales  $\sigma_s$  by a factor of 0.7 as suggested in (Mikolajczyk 2004). Candidate interest points are located at coordinates  $(x, y, z)$  where the second-moment matrix  $M(x, y, z, \sigma_{Harris}; \sigma_s)$  has large eigenvalues  $\lambda_1, \lambda_2, \lambda_3$ . Interest point are hence the 4D local maxima in the scale-space of  $S_{Harris}$ . Locations of interest points are refined using the sub-voxel refinement method described in Sect. 3.1.

### 3.4 Determinant of Hessian (DoH)

The DoH interest point is similar to the Harris detector with respect to formulation (Lindeberg 1998); instead of computing the second-moment matrix  $M$ , it is based on the Hessian matrix  $H$  in (5):

$$H = \begin{bmatrix} V_{xx} & V_{xy} & V_{xz} \\ V_{yx} & V_{yy} & V_{yz} \\ V_{zx} & V_{zy} & V_{zz} \end{bmatrix} \quad (5)$$

where  $V_{xy}$  denotes the second derivative of the volume at scale  $\sigma_s$ , along  $x$  and  $y$  axes:

$$V_{xy} = \frac{\partial^2 \mathbf{V}(\mathbf{x}; \sigma_s)}{\partial x \partial y} \quad (6)$$

The saliency response is the scale-normalized determinant of Hessian matrix  $H$ :

$$S_{Hessian} = \sigma_s^3 \det(H) \quad (7)$$

Similar to Harris and DoG, the interest points are located at the 4D scale-space local maxima of  $S_{Hessian}$ .

### 3.5 SURF

Speeded up robust features (SURF) is a feature extraction algorithm optimized for efficiency (Bay et al. 2008). The 3D, volumetric version of SURF was first introduced in Willems et al. (2008) for video classification. Recently, it was used in a 3D shape object recognition task (Knopp et al. 2010).

SURF is an efficient approximation of the DoH detector. Second-order derivatives of Gaussians in the DoH detector are approximated by six Haar wavelets (i.e. box filters). Convolutions of the Haar wavelets can be greatly accelerated using integral videos/volumes. The saliency response of 3D SURF is similar to the aforementioned DoH detector.

### 3.6 V-FAST

Building on the success of the FAST corner detector (Rosten et al. 2010), V-FAST (Yu et al. 2010) and FAST-3D (Koelstra and Patras 2009) have been proposed for video-based object classification. The V-FAST algorithm performs accelerated segment tests on three orthogonal circles along  $xy$ ,  $xz$  and  $yz$  planes. The saliency score is computed by maximizing the threshold  $t$  that makes at least  $n$  contiguous voxels brighter or darker than the nucleus voxel by  $t$ , thus:

$$AST_{xy}(n, t) = \begin{cases} t & \text{if } \|v_{\text{nucleus}} > \mathbf{c}_{xy} + t\| \geq n \\ t & \text{if } \|v_{\text{nucleus}} < \mathbf{c}_{xy} - t\| \geq n \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

$$S_{\text{vfast}}^{xy} = \max(AST_{xy}(n, t)) \quad (9)$$

$$S_{\text{vfast}} = \sqrt{(S_{\text{vfast}}^{xy})^2 + (S_{\text{vfast}}^{xz})^2 + (S_{\text{vfast}}^{yz})^2} \quad (10)$$

$\mathbf{c}_{xy}$  denotes the voxels on an  $xy$ -circle centered at  $v_{\text{nucleus}}$ . The combined saliency response  $S_{\text{vfast}}$  is the Euclidean norm of saliency scores on the three planes in (10). Interest points are detected at the local maxima in  $S_{\text{vfast}}$  over both translation and scale, with at least two non-zero responses in  $AST_{xy}(n, t)$ ,  $AST_{xz}(n, t)$  and  $AST_{yz}(n, t)$ .

### 3.7 MSER

The maximally stable extremal regions (MSER) detector is a region-based blob detection technique proposed by Matas et al. (2004). Extremal regions are the connected components of a thresholded input data (image/volume); the maximally stable regions are selected from a set of nested extremal regions obtained using different thresholds. The extremal regions of an input volume can be enumerated efficiently using the union-find algorithm which has a worst case of  $O(N \log \log N)$  (Matas et al. 2004), where  $N$  is the number of pixels/voxels. Being inherently advantageous for volumetric interest point detection (e.g. robust to rotation and scale changes), MSER has been applied to detection of volumetric salient regions (Donoser and Bischof 2006; Riemenschneider et al. 2009).

Normally an ellipsoid is fitted to each maximally stable region from the input data (Matas et al. 2004), the position and scale of MSER interest points being represented by the centres and radii of such ellipsoids respectively. In this work a sphere is fitted to the stable regions instead, making the interest points compatible with the proposed evaluation framework.

## 4 Methodology

The traditional repeatability ratio measures the repeatability of interest point detectors at a single, predefined accuracy; it is therefore not only sensitive to the choice of matching distance threshold, but also gives little indication of localization accuracy, i.e. the closeness of corresponding interest points, for correspondences, other than that they fall within the threshold. As such, a single repeatability ratio is not sufficient to describe the performance of interest point detectors for various applications with different accuracy requirements.

On the other hand, real matching accuracy, e.g. ROC curves in Bowyer et al. (1999), requires hand crafted point-to-point groundtruth correspondences. This approach is therefore difficult to generalize to large, realistic evaluation datasets such as point clouds from multi-view stereo systems.

Whilst previous evaluations have focused on either localization accuracy or repeatability, we combine the two performance metrics into a single score. The proposed combined score is computed based on repeatability ratio with respect to varying accuracy requirements. This section explains the combined score used in our evaluation.

### 4.1 Localization Accuracy

An interest point is considered as a sphere at coordinates  $[P_x, P_y, P_z]$  with radius  $P_s$  given by its scale. A vector  $\mathbf{P}$  is used to describe this interest point by combining its spatial location and scale, thus:

$$\mathbf{P} = [P_x, P_y, P_z, f \log(P_s)]^T \quad (11)$$

The logarithm of scale is used to remove multiplicative bias across detectors, and since spatial location and scale are not fully commensurate, a parameter  $f$  is introduced to weight the importance of scale to the distance function.

A key component of the proposed evaluation score is the distance metric for measuring the closeness of two corresponding interest points with respect to their locations and scales. In this work the Euclidean norm is used:

$$D(\mathbf{P}, \mathbf{Q}') = \|\mathbf{P} - \mathbf{Q}'\| \quad (12)$$

where  $\mathbf{P}$  is the coordinates of an interest point found in the volume,  $V_P$ , and  $\mathbf{Q}'$  is the point  $\mathbf{Q}$  found in volume  $V_Q$ , transformed into the coordinate frame of  $V_P$  using a known ground truth homography. The evaluation score is based on the distance of an interest point to the nearest transformed interest point in (13):

$$D(\mathbf{P}, \mathcal{Q}') = \min_{\mathbf{Q}'_j \in \mathcal{Q}'} D(\mathbf{P}, \mathbf{Q}'_j) \quad (13)$$

where  $\mathcal{Q}' = \{\mathbf{Q}'_j\}_{j=1}^q$ , the set of  $q$  transformed interest points found in  $V_Q$ .

## 4.2 Repeatability

Schmid et al. (2000) defined repeatability as the ratio of correspondences to points:

$$R_{\text{ratio}}(\mathcal{P}, \mathcal{Q}', \delta) = \frac{\sum_{i=1}^p H(D(\mathbf{P}_i, \mathcal{Q}') - \delta)}{\min(p, q)} \quad (14)$$

where  $\mathcal{P} = \{\mathbf{P}_i\}_{i=1}^p$ , the set of  $p$  interest points found in  $V_P$ , and  $\delta$  is a user provided distance threshold. The Heaviside step function  $H(\cdot)$  returns 1 when the input is positive, 0 otherwise. This repeatability measure favours dense interest points over accurate but sparse interest points (Willis and Sui 2009). However, the fairness of our evaluation is not affected because fully-overlapped object pairs are used in the experiments.

## 4.3 Combined Score

Rosten et al. (2010) computed the area under  $R_{\text{ratio}}$  as a function of the number of interest points, varied using a contrast threshold on the detector. We use the same idea, but computing  $R_{\text{ratio}}$  as a function of the distance threshold,  $\delta$ . The score therefore increases both if a higher proportion of points are matched, and also if matches are more accurate. We also compute a symmetric score, by computing the average score across two matching directions, using  $V_P$  and  $V_Q$  as the reference frames respectively, in order to cancel out the effect of differences in interest point density between the volumes. This score is given as the area under the  $\delta$  vs.  $R_{\text{ratio}}$  curve within a maximum matching distance  $D$ , thus:

$$R_{\text{area}} = \frac{1}{2D} \int_0^D R_{\text{ratio}}(\mathcal{P}, \mathcal{Q}', \delta) + R_{\text{ratio}}(\mathcal{Q}, \mathcal{P}', \delta) d\delta \quad (15)$$

Our  $R_{\text{area}}$  score is advantageous over traditional repeatability, as it reflects both repeatability and accuracy of interest points in one measurement.

## 5 Evaluation

In this section we perform a comprehensive evaluation of the volumetric interest point detectors, investigating their performance under different variations of input data.

### 5.1 Test Data

Three different datasets are used in our evaluation. Two of these are synthetic, as large sets of real, registered, 3D data are not commonly available. Synthetic data are used because we can generate new test data with varying noise levels, transformations and sampling density with accurate ground-truths for evaluation. It is shown in Fig. 7 that synthetic and real testing data are comparable in our evaluation.

The first set, *Mesh*, contains 25 shapes (surface meshes) chosen from the Princeton Shape Benchmark (Shilane et al.

2004) and *TOSCA* (Bronstein et al. 2008) dataset. This selection contains a wide range of geometric features, from coarse structures to fine details, as illustrated in Fig. 2. Point clouds are created by sampling 3D points, with a uniform distribution, over the surfaces of the meshes. Gaussian white noise is added to the points to simulate measurement errors introduced during 3D shape acquisition. The point clouds are then *voxelized* to volumetric data using kernel density estimation with a Gaussian kernel  $g(\cdot, \sigma_{\text{KDE}})$ , as illustrated in Fig. 3. Finally, a linear scale-space  $\mathbf{V}(\mathbf{x}; \sigma_s)$  is created from each volume, in order to detect shape features at different scales. All shapes in the dataset undergo this conversion process.

The *MRI* dataset consists of two synthetic MRI scans of a human brain, generated from BrainWeb simulated brain database (Cocosco et al. 1997), with given ground truth homography (20° rotation and 20 voxel translation) between the two scans

The third dataset, *Stereo*, is a series of 16 point clouds of 8 objects from the Toshiba CAD model point clouds dataset (Pham et al. 2011), which is captured using a multi-view stereo system (Vogiatzis and Hernández 2011). Relative transformations are computed by aligning each point cloud with a reference model using the *iterative closest point* algorithm (Besl and McKay 1992). The same voxelization technique is used to convert stereo point clouds to volumetric data.

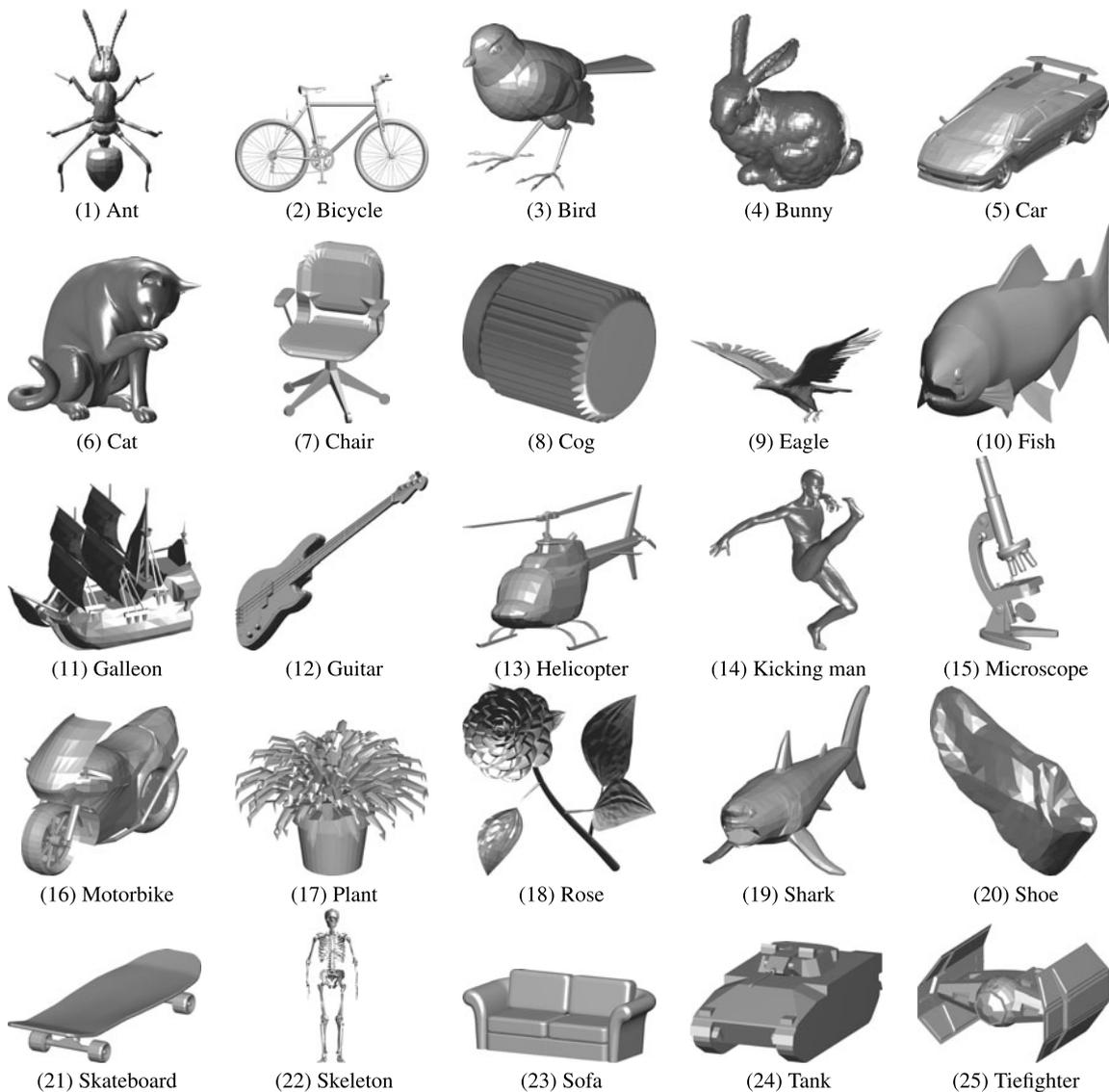
In *Mesh* and *Stereo* datasets, high-intensity voxels are located at the object surface, leaving the interior of the shapes hollow as low-intensity voxels. In contrast, the interior of the *MRI* data is filled with voxels of differing intensity. The experimental results demonstrate the detector behaviours in these two voxelization scenarios.

While the synthetic shape instances of the same object completely overlap one another, avoiding bias to the repeatability score (Willis and Sui 2009), the real stereo data contains occlusions (the underside of each object, which varied across instances, was not captured), as well as uneven sampling density and generally more sampling noise. The applicability to real applications of our performance evaluation using synthetic data will therefore be tested by comparing the results on the *Mesh* dataset with those on the *Stereo* dataset.

### 5.2 Experimental Setup

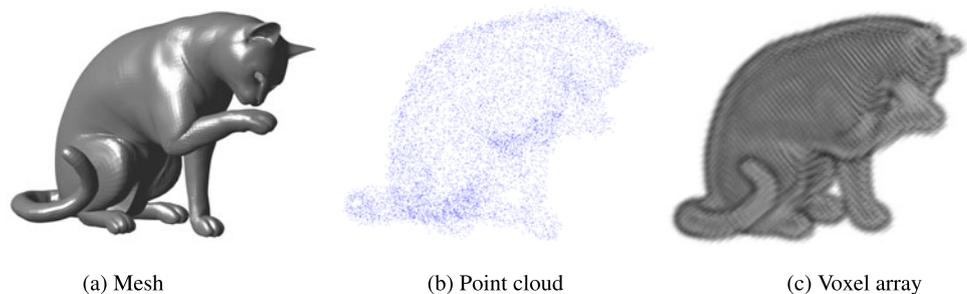
In the evaluation experiments, a series of transformed shapes are created from the reference shape with different magnitudes of a test parameter. A repeatability score is computed by matching the two sets of interest points to each other, according to (15). The overall performance is measured by averaging the  $R_{\text{area}}$  scores across the evaluation dataset.

The characteristics of the interest point detectors are evaluated under several variations. These include rotation, translation, scale, sampling density and noise. Such variations



**Fig. 2** The twenty five 3D shapes in the *Mesh* dataset

**Fig. 3** *Mesh to volume conversion*. Left to right: mesh, point cloud and voxel array



are either introduced during shape acquisition (*MRI* and *Stereo* datasets), or generated synthetically (*Mesh* dataset). The variations observed in the evaluation datasets are described in Table 1. Although image compression rate and

lighting change are also evaluated for image-based detectors (Mikolajczyk et al. 2005), similar experiments are not necessary for 3D shape data because they are unaffected by such changes.

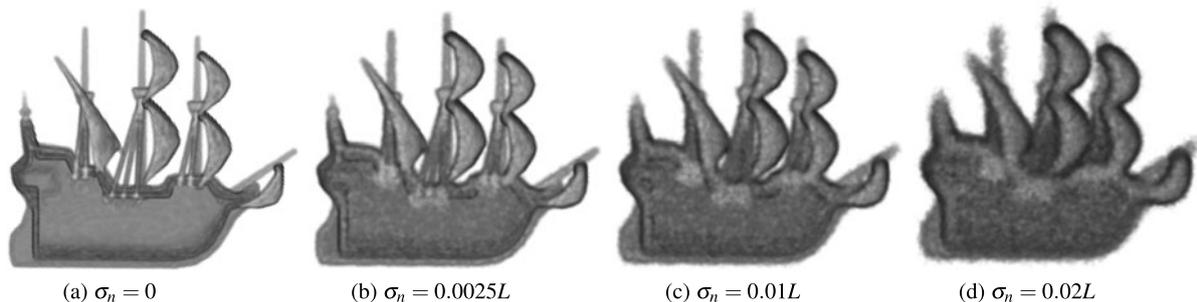
Performances of the candidate detectors are measured as each test parameter is varied individually, keeping all the other parameters at their default values. Sampling parameters (i.e. noise level and sampling density) are applied to all shape instances, whilst pose parameters (i.e. rotation, translation and scale) are applied to only one instance in each matching pair. Some parameters are defined in terms of  $L$ , the largest dimension of the voxelized reference shapes. We set the maximum value of  $L$  to 200 voxels. The default parameters for the reference shapes and the number of transformed shapes created are listed in Table 2.

**Table 1** Variations observed in the evaluation datasets

Variation/ dataset	Noise	Density	Scale	Rotation	Translation
<i>Mesh</i>	✓	✓	✓	✓	
<i>MRI</i>	✓			✓	✓
<i>Stereo</i>	✓	✓	✓	✓	✓

**Table 2** The reference parameters for the testing shapes

Parameter	Value
Default parameters for reference shapes	
Default point cloud size	50000 points
Default noise $\sigma_n$	0.0025 $L$
Default rotation	0°
Maximum $L$	200 voxels
Default $\sigma_{\text{KDE}}$ in $g(\cdot, \sigma_{\text{KDE}})$	1.5 voxels
Distance threshold $D$	0.03 $L$
Parameter $f$ in (11)	$\sqrt{8}$
Number of octaves in scale-space	4
Number of transformed shapes compared	
Sampling noise	13
Sampling density	17
Noise	21
Scale	21



**Fig. 4** The “galleon” shape from the *Mesh* dataset, with different levels of sampling noise. Note that the shape details disappear gradually as the noise level increases.

## 5.3 Experiments on Synthetic Meshes

### 5.3.1 Sampling Noise

Sampling noise and density are crucial factors in 3D interest point detection. As most of the 3D data acquisition techniques rely on shape reconstruction from point clouds or tomograms, existing shape acquisition techniques (e.g. multi-view stereo, 3D ultrasound) often produce data with sampling noise. In this test, different levels of Gaussian white noise, with standard deviations  $\sigma_n$  from  $0L$  to  $0.03L$ , are applied to the *Mesh* dataset.

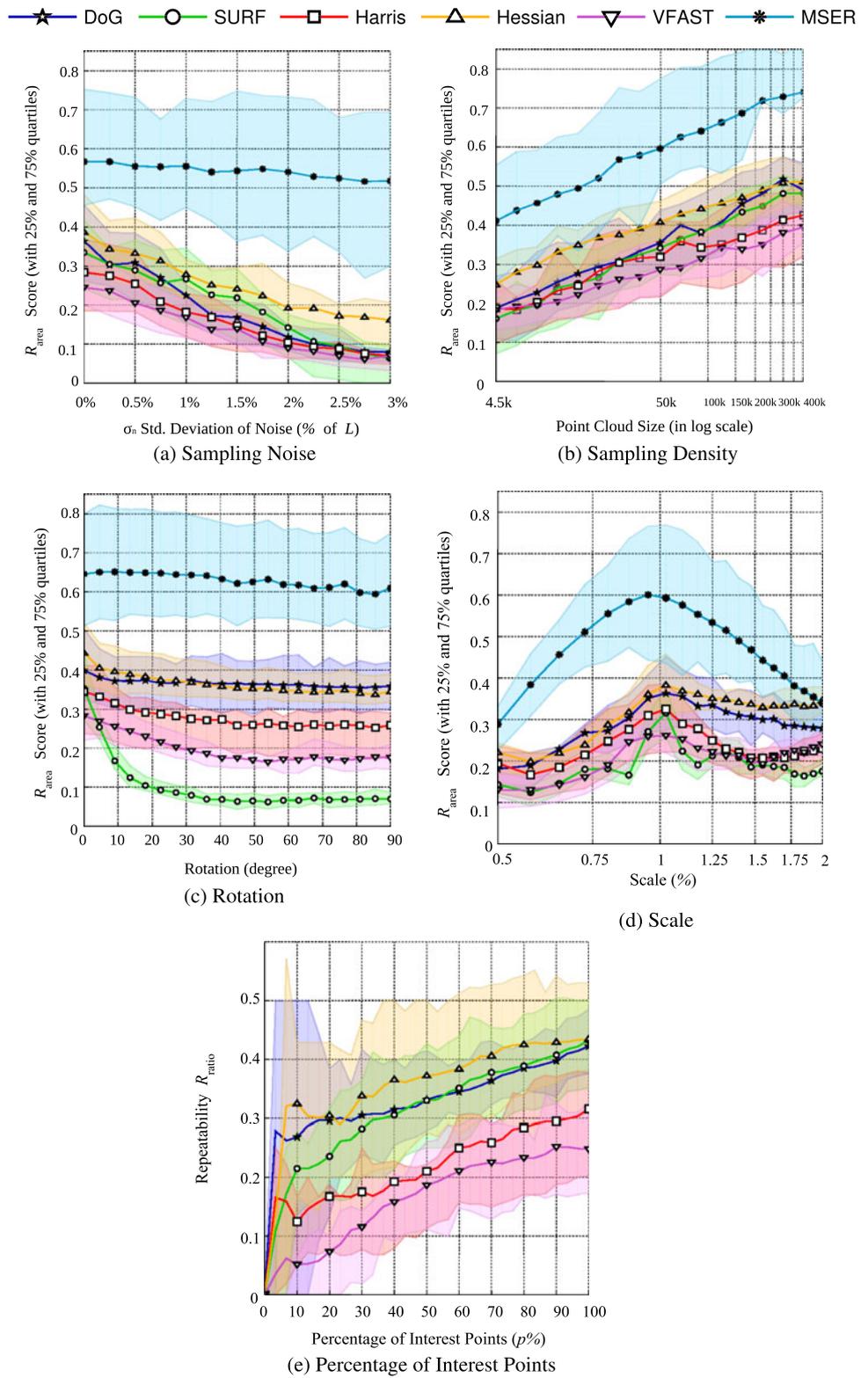
Figure 4 visualizes the effect of different sampling noise levels on the “galleon” shape in the *Mesh* dataset. The result is shown in Fig. 5a. MSER outperforms other interest points, demonstrating high robustness. While the  $R_{\text{area}}$  scores of other detectors decline rapidly, MSER still achieves a high  $R_{\text{area}}$  score. The DoH detector shows a relatively stronger tolerance than detectors like SURF, Harris and V-FAST. In contrast, the SURF detector has almost zero points matched when sampling noise is more than 6.5 % $L$ .

### 5.3.2 Sampling Density

The  $R_{\text{area}}$  score of shapes with various sampling densities are measured. Point clouds are randomly sampled from the input meshes, with point cloud sizes ranging from 4K points to 405K points. The sampling density of a point cloud directly affects its voxelization process, loss of details and holes are usually observed in point clouds in low sampling densities.

Figure 5b presents the change of  $R_{\text{area}}$  scores versus point cloud size. The scores vary linearly in log scale, therefore a diminishing return is observed with increasing sampling density. MSER achieves the best average performance but it also has the largest variance across different shapes. DoH and DoG produce satisfactory results, with high scores but smaller intra-dataset variance than that of MSER.

**Fig. 5**  $R_{area}$  scores of *Mesh* dataset under changing (a) sampling noise, (b) sampling density (point cloud size), (c) rotation, (d) scale and (e) percentage of detected interest points with highest saliency. Solid lines indicate the average  $R_{area}$  score



### 5.3.3 Rotation

This experiment evaluates susceptibility of the detectors to rotational aliasing effects. For each magnitude of rota-

tion angle, an average  $R_{area}$  score is computed by matching the testing shapes multiple times using different rotation axes, making the evaluation results unbiased. Eight rotation axes are generated randomly for each shape, and ro-

tations of increasing magnitude, up to  $90^\circ$ , applied about them.

The effect of rotation is shown in Fig. 5c. Most detectors show excellent tolerance to rotation, inheriting this from their image-based counterparts. DoG and MSER perform slightly better than others, with very stable average score over a broad range of rotation angles. SURF performs worse than other volumetric interest point because the use of box filters introduces quantization errors when the shapes are rotated.

### 5.3.4 Scale

Dimensions of voxelized input data are scaled from 50 % to 200 % of their original sizes. For fairness of evaluation, the transformed shapes are not directly interpolated from their voxelized reference shapes. Rather, input point clouds are re-voxelized with varying volume dimensions  $L$ , whilst other parameters remain unchanged.

The values of  $R_{\text{area}}$  measured against scale changes are illustrated in Fig. 5d. DoG and DoH detectors are comparatively more robust to scale. SURF only works well at 100 % and drops outside the original scale, because of the approximated scale-space used. MSER achieves the best result at its original size, yet its performance decreases steadily when the shape is scaled. Repeatability scores of all detectors drop faster in downsampled volumes (scale  $< 100\%$ ) than in upsampled volumes (scale  $> 100\%$ ). This is due to the information of smaller shape features being lost when the input volume is downsampled. In addition, the scale-space does not cover any feature with size smaller than the first octave, therefore fine details are undetected. Similarly, the performance of most detectors drops slowly at scale  $> 100\%$ , when some features become too large to be detected.

### 5.3.5 Number of Corresponding Interest Points

Table 3 presents quantitative statistics for the number of interest points and correspondences at three noise levels (0.0025, 0.01 and 0.02 of  $L$ ). The MSER detector has the highest percentage of correspondences, yet it gives a smaller set of interest points. By contrast, DoH, SURF and Harris produce larger sets of interest points with good correspondence ratios. The displacement threshold used here ( $D = 0.015L$ ) is about half the typical value, hence only accurate correspondences are counted towards the values in the table.

### 5.3.6 Saliency

Figure 5e shows the repeatability,  $R_{\text{ratio}}$ , with varying percentages of interest points. For each detector, the detected

**Table 3** For each entry, top to bottom: The average number of interest points detected, the average number of correspondences ( $d \leq 0.015L$ ), percentage of points with correspondences

Avg. # Pts. Avg. # Corr. Pts. (Corr. %)	Sampling Noise Level		
	Low (0.0025 $L$ )	Medium (0.01 $L$ )	High (0.02 $L$ )
DoG	122.0	118.2	73.1
	48.8	35.1	9.3
	(39.8 %)	(29.7 %)	(12.7 %)
SURF	154.7	70.4	28.7
	54.7	18.4	3.84
	(35.3 %)	(26.2 %)	(13.4 %)
Harris	303.3	142.2	123.8
	78.6	33.2	13.4
	(25.9 %)	(23.3 %)	(10.83 %)
DoH	<b>330.8</b>	<b>272.0</b>	<b>201.2</b>
	<b>117.1</b>	<b>72.2</b>	<b>30.2</b>
	(35.4 %)	(26.6 %)	(18.2 %)
V-FAST	115.9	85.5	74.6
	33.5	15.5	7.4
	(28.8 %)	(18.10 %)	(9.85 %)
MSER	99.0	74.4	52.5
	59.9	44.7	28.8
	<b>(60.5 %)</b>	<b>(60.2 %)</b>	<b>(54.9 %)</b>

interest points are sorted by their corresponding saliency responses in descending order, then the first  $p\%$  of interest points are used to calculate  $R_{\text{ratio}}$ . However, since no saliency measure is defined in the MSER detector, the number of detected interest points cannot be controlled directly; MSER is therefore not included in Fig. 5e. For analyzing the accuracies of the candidate detectors,  $R_{\text{ratio}}$  is computed using a smaller displacement threshold ( $D = 0.015L$ ) in this experiment.

The performance of DoG and Harris detectors tend to be stable (though Harris performs notably worse than DoG) with increasing numbers of interest points (i.e. decreasing saliency threshold), indicating that a saliency threshold is not necessary for these detectors. DoH's performance, which is initially the best, decreases very slowly and converges with DoG and SURF, indicating that the lower saliency points are less reliable. A saliency threshold for DoH might benefit applications requiring more accurate point localization. By contrast, the repeatability scores

of SURF and V-FAST increase steadily before leveling off, suggesting that some of the high saliency points are unreliable; this poses more of a problem in terms of interest point selection.

### 5.4 Experiments on MRI and Stereo Data

Interest points detectors are also evaluated on the *MRI* and *Stereo* datasets. As a reference for cross validation, average scores obtained from the *Mesh* dataset are plotted against displacement threshold  $d$  in Fig. 7a.

#### 5.4.1 MRI Dataset

This dataset contains two MRI scans of a human brain; each MRI has the longest dimension  $L$  of 218 voxels. Figure 6 shows the MSER interest points detected in the data. It is worth noting that some points detected on the MRIs can be matched easily, such as those on the nose-tips, eye sockets and foreheads, but that there are fewer detections within the brain area.

$R_{\text{area}}$  scores are measured from the dataset with varying  $d$  in Fig. 7b. The evaluation results obtained are comparable to that of synthetic mesh data. MSER, DoG and DoH perform slightly better in synthetic meshes, while the Harris detector is good at detecting complicated internal structures in the MRI scans.

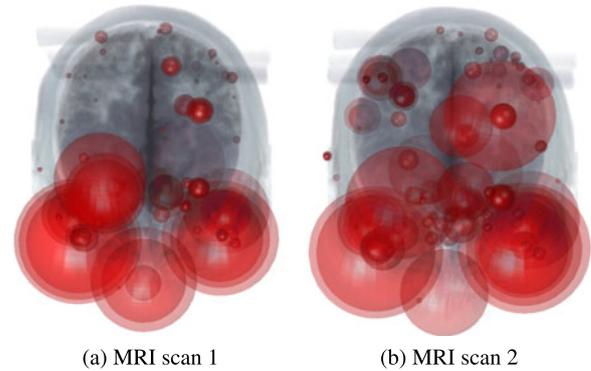
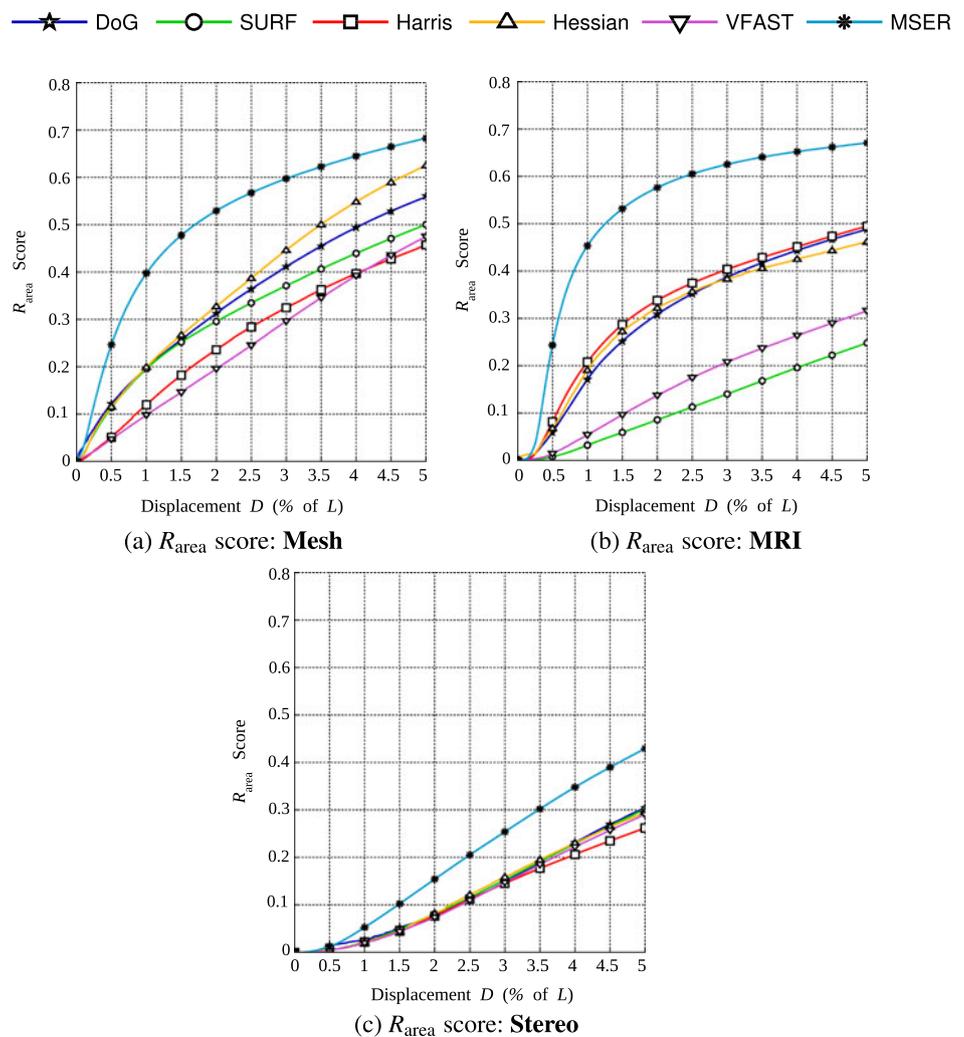


Fig. 6 Two volumetric MRI scans of a human brain, with detected MSER features

Fig. 7  $R_{\text{area}}$  scores versus displacement threshold  $d$ . Left to right: (a) *Mesh*, (b) *MRI* and (c) *Stereo* datasets



**Table 4** From left to right: the algorithmic complexities, the average time (in microseconds per voxel) to detect interest points on the *Mesh* dataset, the potential for heavily parallel implementation (with theircorresponding 2D implementations as references), and the potential for machine learning based acceleration.  $N$  is the number of voxels and  $S$  is the number of octaves in the scale-space

Interest point	Complexity	Speed ( $\mu$ s/voxel)	Parallelizable	Machine learning acceleration
DoG	$O(SN)$	1.325	✓ (Sinha et al. 2006)	
SURF	$O(SN)$	1.035	✓ (Cornelis and Gool 2008)	
Harris	$O(SN)$	1.112	✓ (Teixeira et al. 2009)	
DoH	$O(SN)$	1.325	✓ (Bhatia et al. 2007)	
V-FAST	$O(SN)$	1.670	✓ (Dohi et al. 2011)	✓ (Rosten et al. 2010)
MSER	$O(N \log \log N)$	2.070	✓ (Kristensen and MacLean 2007)	

#### 5.4.2 Stereo Dataset

Sixteen point clouds, generated using a multi-view stereo technique (Vogiatzis and Hernández 2011), are converted into volumes with maximum length of 132 voxels. Figure 9 shows some sample point clouds from the *Stereo* dataset and different interest points detected on their corresponding voxelized shapes. The  $R_{\text{area}}$  scores obtained from the *Stereo* dataset, shown in Fig. 7c, are lower compared with *Mesh* and *MRI* datasets, especially at small  $D$ . Nonetheless, in terms of overall rankings and relative scores of the detectors, our synthetic and real data demonstrate similar behaviour. The decrease in performance for our stereo data could be due to its: (a) low sampling frequency and high noise, (b) uneven object surfaces, which are infeasible for blob detection algorithms (e.g. MSER, SURF and DoH) and (c) small errors in the estimated ground truth poses from ICP alignment. The differences in  $R_{\text{area}}$  scores are due to *occlusions* due to viewpoint changes and *uneven sampling density* of the *Stereo* data.

Some objects (e.g. “mini” in Fig. 9) exhibit a much sparser reconstruction due to a lack of texture on the object’s surface, but it is interesting to note that the distribution of detected features is no less dense across any of the detectors, suggesting that our synthetic results are representative of sparse point clouds as well as dense.

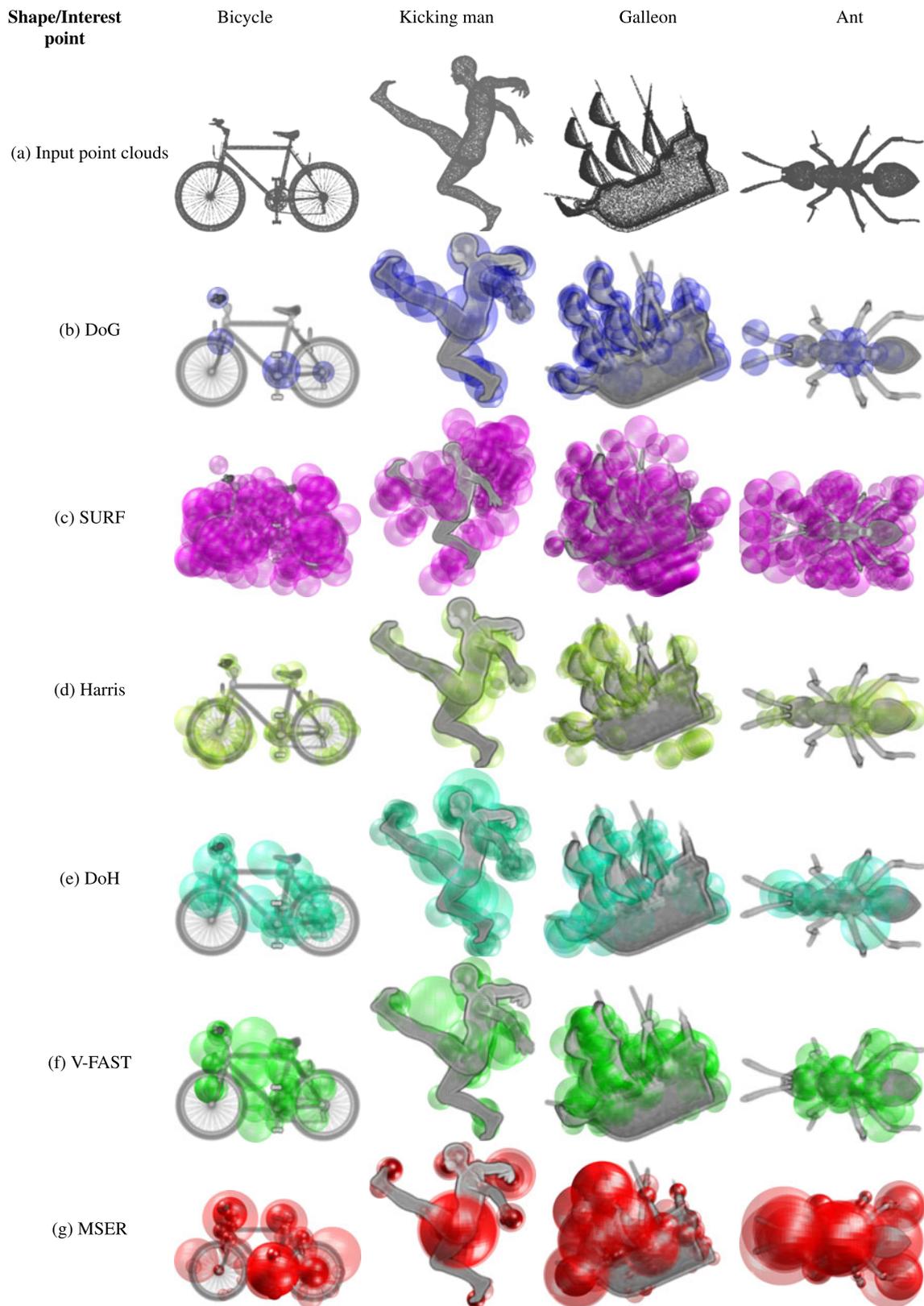
#### 5.5 Computation Time

Computation efficiency is also a crucial factor for choosing a suitable interest point detector for a particular computer vision application. Since the storage size of volumetric data is usually much larger than 2D images, the time required to compute the interest points increases accordingly. Operations which are considered efficient for 2D images may become much slower for 3D volumetric data. For non-volumetric data such as point clouds, extra computation time is required to voxelize the shape data. On the other hand, some interest point detectors can be accelerated

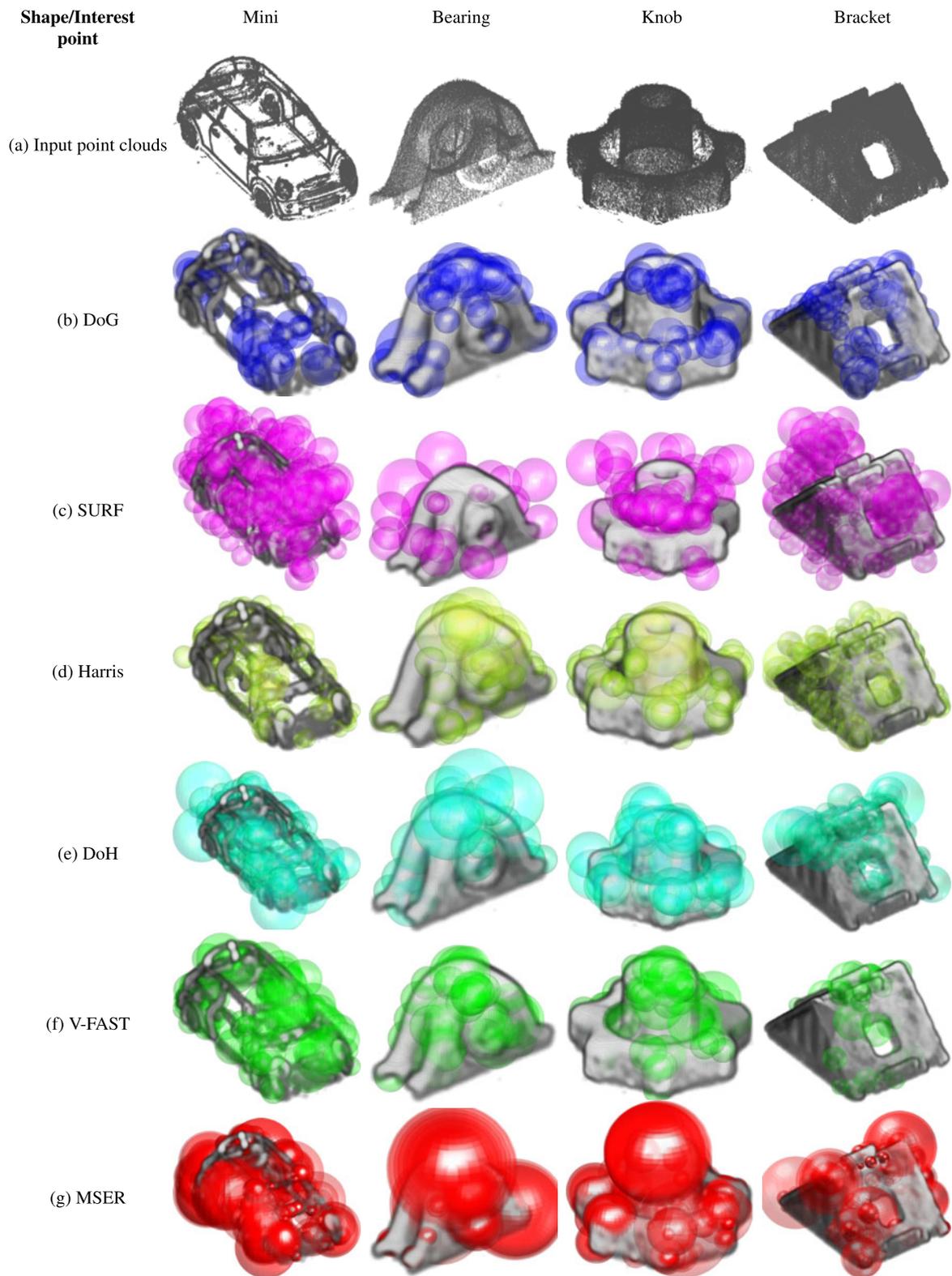
by parallel processing techniques, using specialized hardware (e.g. GPUs and FPGAs). Table 4 summarizes the algorithmic complexities of the candidate detectors and the time required for them to compute interest points on the *Mesh* dataset. In order to compare the speed performance in a common evaluation framework, we implemented the all candidate detectors in MATLAB, without any hardware acceleration. The experiments were performed on the same hardware platform (Intel Core i7, 12 GB RAM). Theoretically all detectors have a similar time complexity, except for MSER, which does not implement a Gaussian scale-space. The SURF detector is the fastest among the candidate detectors, as it uses Haar wavelets to accelerate the computation of Gaussian derivatives. Harris also shows high efficiency due to its relatively simple algorithm. MSER is the slowest detector, as the search algorithm for stable regions is less efficient in 3D volumes than in 2D images. Surprisingly, the V-FAST detector is the second slowest detector in the experiment. Since FAST is a pixel/voxel-wise rule based algorithm, the accelerated segment test in (8) cannot utilize the high performance linear algebra routines used by other detectors such as DoG and SURF. On the other hand, the FAST detector can be accelerated by learning a decision tree classifier from training data (Rosten et al. 2010), which our implementation does not use.

#### 5.6 Qualitative Analysis: Blobs Versus Corners

Volumetric interest points can be roughly classified into three categories: region-based blob detection (MSER), derivative-based blob detection (DoG, DoH and SURF) and corner detection (Harris, V-FAST). The quantitative evaluation results imply that region-based blob detectors work better than derivative-based blob detectors, and blob detectors are better than corner detectors, but this is not the whole story. The candidate detectors demonstrate different behaviours in terms of locations and scales of the detected interest points. Therefore, besides repeatability, it is also important to analyze the characteristics of detectors qualitatively. Figure 8 visualizes interest points detected by the



**Fig. 8** (a) Sample point clouds obtained from the *Mesh* dataset, (b) DoG, (c) SURF (d) Harris, (e) DoH, (f) V-FAST and (g) MSER, visualized on the voxelized data. The color spheres represent the positions and relative scales of the detected interest points (Color figure online)



**Fig. 9** (a) Sample point clouds obtained from the *Stereo* dataset. (b) DoG, (c) SURF (d) Harris, (e) DoH, (f) V-FAST and (g) MSER interest points visualized on the voxelized data (Color figure online)

six candidate detectors on the “cat” object from the *Mesh* dataset.

### 5.6.1 Region-Based Blob Detection

MSER detects contiguous regions of any shape (i.e. not limited to spherical blobs) allowing it to select more robust regions from a greater selection, and hence perform better. MSER performs well in 3D shape data because the shapes of salient regions are inherently less susceptible to viewpoint changes. It can be seen in Figs. 8g and 9g that MSER finds features at fewer locations, but over multiple scales. The locations tend to center on regions of high surface curvature.

### 5.6.2 Derivative-Based Blob Detection

DoG, DoH and SURF detectors theoretically find, in order of preference, spherical blobs, corners and planes. DoG and DoH have qualitatively similar output, as shown in Figs. 8b, 9b, 8e and 9e, finding features such as limb extremities in *Mesh* dataset or sharp corners in *Stereo* dataset, as well as inside shape such as the hull of “ship” and the hole in “knob”. By contrast, SURF (Fig. 1b), despite being an approximation of the DoH detector, produces features off the surface (both inside and outside), often over regions of low surface curvature. Instead of highly repeatable features, derivative-based approaches produce qualitatively more diverse features than region-based detectors. Finally, their repeatability scores degrade more rapidly than MSER for noisy input data (see Fig. 1a), mainly due to higher rates of false positive detections.

### 5.6.3 Corner Detection

Harris and V-FAST both aim to find areas of high curvature. However, their outputs (Figs. 1c and 1e) vary qualitatively, with the former tending to find fewer features and more sharp corners than the latter, which finds an even distribution of features over both scale and location. In general, corner detection approaches are relatively less robust to noise and transformations than blob detection techniques. Whilst blobs remain stable in noisy or transformed volumetric data, corners are more easily affected by quantization errors or sampling noise. The performance of blob detectors drops faster than that of edge detectors at low (See Fig. 5b) and uneven (compare Figs. 7a and 7c) sampling densities. Holes form on shape surfaces in these situations, creating false positives for both corner and blob detectors, but the detection of true positives appears to be more robust for corner detectors in these cases.

## 6 Conclusion

In this paper, the state of the art in volumetric interest point detection is evaluated. The purpose of this work is to provide comprehensive guidance on the selection of interest point detector for any computer vision or machine learning task that uses 3D input data. Six interest point detectors, from existing 3D computer vision applications, are evaluated on three different datasets (meshes, MRI scans and 3D point clouds) under varying noise levels and transformations. A novel evaluation metric is introduced by combining two existing performance metrics, repeatability and accuracy, into a single measurement. The acquired experimental results are analyzed both quantitatively and qualitatively.

Summarizing the quantitative results with respect to the proposed  $R_{\text{area}}$  score, MSER achieves the best overall performance, being robust to both noise and rotation. Taking the number of corresponding points into account, DoH and, to a lesser extent, DoG maintain a balanced performance between this and repeatability. Generally speaking, blob detectors (e.g. DoH and DoG) appear to perform better than corner detectors (e.g. Harris and V-FAST) in 3D shapes, a result that agrees with an evaluation of image-based detectors (Mikolajczyk et al. 2005). Consistent results are obtained from different evaluation datasets, which indicate that both the detectors and evaluation framework are applicable to multiple types of input data. In the context of efficiency, MSER is the slowest interest point detector in terms of computation time for 3D volumetric data. Evaluation results show that fast detectors for 2D images, e.g. FAST, may not obtain consistent performances for 3D volumes. Hardware acceleration is essential when real-time performance is needed. While parallelized, hardware-accelerated 2D interest points are available, volumetric implementations of such detectors are still uncommon.

In the qualitative analysis we discussed the nature of features found by the candidate interest point detectors. They exhibit unique characteristics with respect to their locations, sizes and number of interest points. The analysis suggests that the repeatability of a detector is also affected by the nature of input data, such as the number of distinct corners, edges or blobs on the 3D shapes. In addition, the suitability of an interest point depends on the qualitative requirements of the target applications (e.g. object classification, correspondence matching, segmentation). Hence, repeatability is not the sole factor in determining the performance of applications; the choice of volumetric interest points is actually application dependent.

## References

- Aanaes, H., Dahl, A. L., & Pedersen, K. S. (2010). On recall rate of interest point detectors. In *Proceedings of the fifth international*

- symposium on 3D data processing, visualization and transmission.*
- Bay, H., Ess, A., Tuytelaars, T., & Gool, L. V. (2008). Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3), 346–359.
- Besl, P. J., & McKay, N. D. (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 239–256.
- Bhatia, A., Laganiere, R., & Gerhard Roth, G. (2007). Performance evaluation of scale-interpolated Hessian-Laplace and Haar descriptors for feature matching. In *The 14th international conference on image analysis and processing* (pp. 61–66).
- Bowyer, K., Kranenburg, C., & Dougherty, S. (1999). Edge detector evaluation using empirical ROC curves. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (Vol. 1, p. 2). vol. (xxiii+637+663).
- Brand, P., Mohr, R., & Rhones-Alpes, L. I. (1994). Accuracy in image measure. In *Proceedings of the SPIE conference on videometrics III* (Vol. 2350, pp. 218–228).
- Bronstein, A., Bronstein, M., & Kimmel, R. (2008). *Numerical geometry of non-rigid shapes* (1st ed.). Berlin: Springer.
- Brown, M., & Lowe, D. G. (2005). Unsupervised 3D object recognition and reconstruction in unordered datasets. In *Proceedings of the fifth international conference on 3-D digital imaging and modeling* (pp. 56–63). Washington: IEEE Comput. Soc.
- Cocosco, C. A., Kollokian, V., Kwan, R. K. S., Pike, G. B., & Evans, A. C. (1997). BrainWeb: online interface to a 3D MRI simulated brain database. *NeuroImage*, 5, 425.
- Coelho, C., Heller, A., Mundy, J. L., Forsyth, D. A., & Zisserman, A. (1992). An experimental evaluation of projective invariants. In J. L. Mundy & A. Zisserman (Eds.), *Geometric invariance in computer vision* (pp. 87–104). Cambridge: MIT Press.
- Cornelis, N., & Gool, L. V. (2008). Fast scale invariant feature detection and matching on programmable graphics hardware. In *The IEEE conference on computer vision and pattern recognition workshops* (pp. 1–8).
- Criminisi, A., Shotton, J., Robertson, D. P., & Konukoglu, E. (2010). Regression forests for efficient anatomy detection and localization in CT studies. In B. H. Menze, G. Langs, Z. Tu, & A. Criminisi (Eds.), *Lecture notes in computer science: Vol. 6533. The MICCAI workshop of medical computer vision 2010: recognition techniques and applications in medical imaging* (pp. 106–117). Berlin: Springer.
- Dalvi, R., Hacihaliloglu, I., & Abugharbieh, R. (2010). 3D ultrasound volume stitching using phase symmetry and Harris corner detection for orthopaedic applications. In *SPIE medical imaging* (p. 762330).
- Dohi, K., Yorita, Y., Shibata, Y., & Oguri, K. (2011). Pattern compression of fast corner detection for efficient hardware implementation. In *The international conference on field programmable logic and applications (FPL)* (pp. 478–481).
- Donner, R., Birmgruber, E., Steiner, H., Bischof, H., & Langs, G. (2011). Localization of 3d anatomical structures using random forests and discrete optimization. In B. Menze, G. Langs, Z. Tu, & A. Criminisi (Eds.), *Lecture notes in computer science: Vol. 6533. Medical computer vision. Recognition techniques and applications in medical imaging* (pp. 86–95). Berlin: Springer.
- Donoser, M., & Bischof, H. (2006). 3d segmentation by maximally stable volumes (MSVS). In *Proceedings of the international conference on pattern recognition* (Vol. 1, pp. 63–66).
- Dutagaci, H., Cheung, C. P., & Godil, A. (2011). Evaluation of 3D interest point detection techniques. In *Proceedings of Eurographics workshop on 3D object retrieval*, Llandudno, UK (pp. 57–64).
- Fisher, R. B. (1987). Modelling second order volumetric features. In *Proceedings of the 3rd Alvey vision conference* (pp. 79–86).
- Flitton, G., Breckon, T., & Bouallagu, N. M. (2010). Object recognition using 3D SIFT in complex CT volumes. In *Proceedings of the British machine vision conference*. Guildford: BMVA Press.
- Gomb, P. (2009). Detection of interest points on 3d data: extending the Harris operator. In M. Kurzynski & M. Wozniak (Eds.), *Computer recognition systems 3, advances in intelligent and soft computing* (Vol. 57, pp. 103–111). Berlin: Springer.
- Harris, C., & Stephens, M. (1988). A combined corner and edge detection. In *Proceedings of the 4th Alvey vision conference* (pp. 147–151).
- Heath, M. D., Sarkar, S., Sanocki, T., & Bowyer, K. W. (1997). A robust visual method for assessing the relative performance of edge-detection algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(12), 1338–1359.
- Janoch, A., Karayev, S., Jia, Y., Barron, J., Fritz, M., Saenko, K., & Darrell, T. (2011). A category-level 3-d object dataset: putting the kinect to work. In *The IEEE international conference on computer vision workshops* (pp. 1168–1174).
- Knopp, J., Prasad, M., Willems, G., Timofte, R., & Gool, L. V. (2010). Hough transform and 3D SURF for robust three dimensional classification. In *Proceedings of the European conference on computer vision* (pp. 589–602). Berlin: Springer.
- Koelstra, S., & Patras, I. (2009). The FAST-3D spatio-temporal interest region detector. In *Workshop on image analysis for multimedia interactive services* (pp. 242–245).
- Kristensen, F., & MacLean, W. J. (2007). Real-time extraction of maximally stable extremal regions on an fpga. In *The IEEE international symposium on circuits and systems 2007* (pp. 165–168).
- Lai, K., & Fox, D. (2010). Object recognition in 3d point clouds using web data and domain adaptation. *The International Journal of Robotics Research*, 29(8), 1019–1037.
- Laptev, I. (2005). On space-time interest points. *International Journal of Computer Vision*, 64(2–3), 107–123.
- Laptev, I., & Lindeberg, T. (2003). A distance measure and a feature likelihood map concept for scale-invariant model matching. *International Journal of Computer Vision*, 52(2–3), 97–120.
- Lindeberg, T. (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2), 79–116.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.
- Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10), 761–767.
- Mikolajczyk, K. (2004). Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1), 63–86.
- Mikolajczyk, K., & Schmid, C. (2002). An affine invariant interest point detector. In *ECCV '02: Vol. 1. Proceedings of the European conference on computer vision* (pp. 128–142). London: Springer.
- Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., & Gool, L. V. (2005). A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1/2), 43–72.
- Ni, D., Qu, Y., Yang, X., Chui, Y. P., Wong, T. T., Ho, S. S., & Heng, P. A. (2008). Volumetric ultrasound panorama based on 3d sift. In *Proceedings of the 11th international conference on medical image computing and computer-assisted intervention MICCAI'08* (Vol. 2, pp. 52–60). Berlin: Springer.
- Papazov, C., & Burschka, D. (2011). An efficient RANSAC for 3D object recognition in noisy and occluded scenes. In *Proceedings of the 10th Asian conference on computer vision* (pp. 135–148). Berlin: Springer.
- Pham, M. T., Woodford, O. J., Perbet, F., Maki, A., Stenger, B., & Cipolla, R. (2011). A new distance for scale-invariant 3d shape recognition and registration. In *Proceedings of the IEEE international conference on computer vision* (pp. 145–152).

- Prasad, M., Knopp, J., & Gool, L. V. (2011). Class-specific 3D localization using constellations of object parts. In *Proceedings of the British machine vision conference*. British Machine Vision Association.
- Rajan, P. K., & Davidson, J. M. (1989). Evaluation of corner detection algorithms. In *Proceedings of the 21st Southeastern symposium on system theory* (pp. 29–33).
- Riemenschneider, H., Donoser, M., & Bischof, H. (2009). Bag of optical flow volumes for image sequence recognition. In *Proceedings of the British machine vision conference*. British Machine Vision Association.
- Rosten, E., Porter, R., & Drummond, T. (2010). Faster and better: a machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32, 105–119.
- Ruiz-Alzola, J., Kikinis, R., & Westin, C. F. (2001). Detection of landmarks in multidimensional tensor data. *Signal Processing*, 81, 2243–2247.
- Salti, S., Tombari, F., & Stefano, L. D. (2011). A performance evaluation of 3d keypoint detectors. In *Proceedings of the 2011 international conference on 3D imaging, modeling, processing, visualization and transmission*, 3DIMPVT'11 (pp. 236–243). Washington: IEEE Comput. Soc.
- Schmid, C., Mohr, R., & Bauckhage, T. (2000). Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2), 151–172.
- Shilane, P., Min, P., Kazhdan, M., & Funkhouser, T. (2004). The Princeton shape benchmark. In *Proceedings of the shape modeling international 2004*, SMI'04 (pp. 167–178). Washington: IEEE Comput. Soc.
- Shin, M. C., Goldgof, D., & Bowyer, K. W. (1999). Comparison of edge detectors using an object recognition task. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (Vol. 1, p. 1360).
- Sinha, S. N., Frahm, J. M., Pollefeys, M., & Genc, Y. (2006). Gpu-based video feature tracking and matching. In *Workshop on edge computing using new commodity architectures* (Vol. 278).
- Sipiran, I., & Bustos, B. (2011). Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes. *The Visual Computer*, 27(11), 963–976. Special Issue on 3DOR 2010.
- Smith, S. M., & Brady, J. M. (1997). SUSANa new approach to low level image processing. *International Journal of Computer Vision*, 23(1), 45–78.
- Teixeira, L., Celes, W., & Gattass, M. (2009). *Accelerated corner-detector algorithms*.
- Tuytelaars, T., & Mikolajczyk, K. (2008). Local invariant feature detectors: a survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3), 177–280.
- Unnikrishnan, R., & Hebert, M. (2008). Multi-scale interest regions from unorganized point clouds. In *Workshop on search in 3D (S3D)* (pp. 1–8). IEEE conference on computer vision and pattern recognition.
- Vikstén, F., Nordberg, K., & Kalms, M. (2008). Point-of-interest detection for range data. In *Proceedings of the international conference on pattern recognition* (pp. 1–4).
- Vogiatzis, G., & Hernández, C. (2011). Video-based, real-time multi view stereo. *Image and Vision Computing*, 29(7), 434–441.
- Wessel, R., Novotni, M., & Klein, R. (2006). Correspondences between salient points on 3D shapes. In *Proceedings of vision, modeling, and visualization workshop 2006 (VMV 2006)* (pp. 365–372). Berlin: Akad. Verlagsgesellschaft.
- Willems, G., Tuytelaars, T., & Gool, L. (2008). An efficient dense and scale-invariant spatio-temporal interest point detector. In *Proceedings of the European conference on computer vision*, ECCV'08 (Vol. 2, pp. 650–663). Berlin: Springer.
- Willems, G., Becker, J. H., Tuytelaars, T., & Gool, L. J. V. (2009). Exemplar-based action recognition in video. In *Proceedings of the British machine vision conference*. British Machine Vision Association.
- Willis, A., & Sui, Y. (2009). An algebraic model for fast corner detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2296–2302).
- Yu, T. H., Kim, T. K., & Cipolla, R. (2010). Real-time action recognition by spatiotemporal semantic and structural forest. In *Proceedings of the British machine vision conference*. British Machine Vision Association.
- Zaharescu, A., Boyer, E., Varanasi, K., & Horaud, R. P. (2009). Surface feature detection and description with applications to mesh matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 373–380).