3DIMPVT
#13

An Evaluation of Volumetric Interest Points

Anonymous 3DIMPVT submission

Paper ID 13

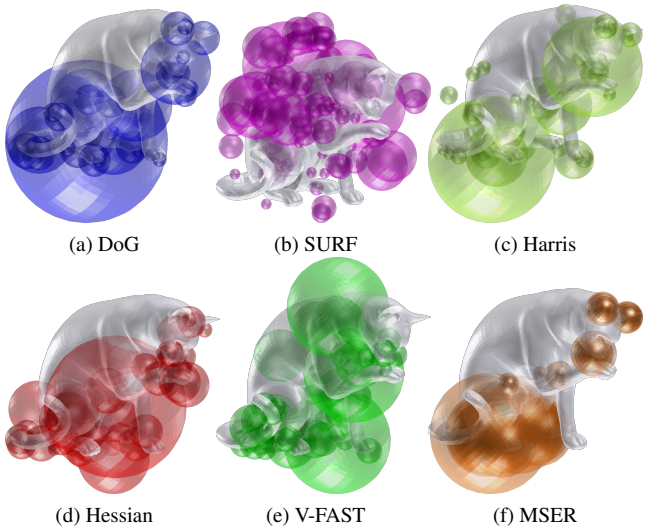
Abstract

*Shape and texture are two fundamental properties of objects. The use of texture (*i.e. *2D surface colouration), or appearance, for object detection, recognition and registration has been well studied, shape (in 3D) less so. However, the increasing prevalence of depth sensors and the diminishing returns to be had from appearance alone have seen a surge in shape-based methods. In this work we investigate the performance of several detectors of interest points 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, enables readers to easily select a method suitable to their application.*

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, helped by the advent of image descriptors such as SIFT and competitions such as the PASCAL challenge, and these methods are now reaching maturity. However, advancing geometry capture techniques, in the form of stereo, structured light, structure-from-motion and also 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.

Shape-based techniques for object recognition range from the very local, with small, indistinct features, *e.g.* single points, combined in a geometrical consistency framework, to the global, with a single descriptor for an entire object. While the former methods offer robustness to occlusion and clutter and provide pose, they do not cope well with shape variability, nor scale well with the number of object classes. The global methods' pros and cons are tend to be the inverse. 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. 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 repeatedly locatable in different data sets, and whose localities are distinctive. A common and crucial stage of such approaches is therefore the detection of *interest points* to be described.



(a) DoG (b) SURF (c) Harris

(d) Hessian (e) V-FAST (f) MSER

Figure 1: Different types of volumetric interest points detected on a testing shape.

This paper aims to provide the first performance evaluation of interest point detectors on scalar volumetric data. Such data not only comes directly from volumetric sensors, *e.g.* MRIs, but can also be generated from other 3D data such as point clouds or meshes, making the evaluation as

widely applicable as possible. The representation allows interest points to be located off an object's surface, *e.g.* at the centre of a limb. In addition, 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 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 also a scale, and distances are computed in this space.

The remainder of this section reviews previous work relevant to interest point detectors and our evaluation framework. The following section introduces the interest point detectors we compared, while section 3 describes our evaluation methodology. Section 4 describes our experiments and presents the results, before we conclude in section 5.

1.1. Previous work

We will now discuss relevant prior work on interest point detectors and evaluation methodologies.

1.1.1 Interest point detectors

Interest points are typically used to match corresponding points across two or more similar sets of data. The data has predominantly been images, though some detectors have recently been applied to volumes. This review will be brief, focusing on the detectors we evaluate and previous applications to 3D data, as the field is well studied (at least in respect to images); we refer the reader to a recent survey [23] for more details.

Differential approaches to interest point detection find optima in the data response to differential operators such as the second moment matrix (first order derivatives), *e.g.* the Harris corner detector [10], and the Hessian matrix (second order derivatives). Lindeberg [14] studied scale-covariant[1] points using the Laplacian-of-Gaussian kernel (equivalent to the trace of the Hessian), as well as the determinant of the Hessian. Lowe [15] approximated the former with a Difference-of-Gaussians (DoG) operator for efficiency, a method recently applied to CT scan volumes [9]. The SURF detector of Bay *et al.* [1] accelerates computation of the Hessian determinant through the use of integral images, since applied to integral volumes of space-time video [24] and binary volumes generated from 3D meshes [11]. Mikolajczyk & Schmid [17] developed the scale-covariant Harris-Laplace detector by finding Harris corners in the spatial domain which are maxima of the Laplacian in the scale domain, an approach which Laptev [13] applied to classification of video volumes.

In a different approach, the FAST detector of Rosten *et al.* [20] uses the largest number of contiguous pixels on a circle of fixed radius which are significantly darker, or brighter, than the centre pixel. They learn a decision tree classifier that makes the approach extremely efficient, leading it to be applied to space-time video volumes also [12, 25]. The MSER detector of Matas *et al.* [16] 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. MSER has already been applied to volumetric data, firstly in the context of segmentation of MRIs [8], then on spatio-temporal data [19].

Image-based detectors have also been made affine-covariant (*e.g.* [17]), in order to approximate the perspective distortion caused by projection of 3D surfaces onto the 2D image plane. 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.

1.1.2 Methodologies

Empirical performance evaluation is a popular pastime in Computer Vision, and the topic of interest point[2] detection is no exception. Some methods evaluate performance in the context of a particular task, lacking generality to other applications. Most 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.* [7] measure accuracy by computing projective invariants and comparing these with the actual values, measured from the scene. Brand & Mohr [4] introduce three further measures, including 2D Euclidean distance from detected points to ground truth corner locations (given by line fitting to a grid). This has been extended to using the distance to the nearest point detected in, and transformed from, another image (*e.g.* [21]). Matching scores for interest points found over location and *scale* [17] , and also *affine* transformations [17], have since been proposed.

When used for object detection and recognition, two other important characteristics of the interest points are their repeatability (*i.e.* is the same feature found in multiple datasets) and their distinctiveness (*i.e.* can a point, or region of interest be easily recognized among many). Repeatability was proposed by Schmid *et al.* [21] as the ratio of repeated points to detected points. Rosten *et al.* [20]

[1] 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.

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

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 [3], which takes into account false matches. Schmid *et al.* [21] also introduce a quantitative measure of distinctiveness—entropy, or "information content". Alternatively, qualitative visual comparison is used (*e.g.* [14, 13]), on test datasets containing a variety of different interest points.

In the context of texture-based interest points, performance of detectors are often measured over variations in image rotation, scale, viewpoint angle, illumination and noise level (*e.g.* [21] covers all these factors), as well as corner properties. Efficiency may be a further consideration [20].

2. Detectors

This section briefly describes the interest point detectors we evaluate. Our work aims to evaluate some of these volumetric interest points, these include difference of Gaussians (DoG) [9], speeded up robust features (SURF) [24, 11], Harris-based [13] and Hessian-based interest points, V-FAST [25] and maximally stable extremal regions (MSER) [8, 19]. This section briefly recapitulates the approaches of these candidate interest points.

DoG Difference of Gaussians is a popular blob detection technique since it has been used by SIFT algorithm [15] for feature localization. Its volumetric extension has been applied in both shape recognition [9]. DoG approximates the Laplacian of Gaussian filter, which enhances features at a particular size. The DoG saliency response S_{DoG} is computed by subtracting two Gaussian smoothed volumes, which are often adjacent scale-space representations, of the same signal. Interest points are detected at the Interest point are detected at the *4D local maxima* (both 3D space and scale) in S_{DoG} within each octave of $\mathbf{V}(\mathbf{x}, \sigma_s)$.

Harris Extending the Harris corner detector [10], Laptev introduces the space-time interest point for video categorization [13]. The volumetric Harris corner detector is largely identical to Laptev's work, except a spherical local window is used instead of a ellipsoidal window. The second-moment matrix 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 $g(\cdot; \sigma_{harris})$:

$$\mathcal{M} = g(\cdot; \sigma^2_{harris}) * \begin{bmatrix} \mathbf{V}_x^2 & \mathbf{V}_x\mathbf{V}_y & \mathbf{V}_x\mathbf{V}_z \\ \mathbf{V}_x\mathbf{V}_y & \mathbf{V}_y^2 & \mathbf{V}_y\mathbf{V}_z \\ \mathbf{V}_x\mathbf{V}_z & \mathbf{V}_y\mathbf{V}_z & \mathbf{V}_z^2 \end{bmatrix}, \quad (1)$$

where $\mathbf{V}_x$,$\mathbf{V}_y$,$\mathbf{V}_z$ denote the partial derivatives of the volume in scale-space $\mathbf{V}(\mathbf{x}; \sigma_s)$ along x, y and z axes respectively. The saliency S_{Harris} is computed from the determinant and trace of $\mathcal{M}$:

$$S_{\text{Harris}} = \sigma_s^3 \det(\mathcal{M}) - k\,\text{trace}(\mathcal{M})^3. \quad (2)$$

The parameter k controls the rejection of edge points, which ranges between 0.003 to 0.006. Each saliency response S is normalized by its scale σ_s. The window size σ_{harris} is proportional to expected feature scales σ_s by a factor of 0.7 as suggested in [17]. Interest point are the 4D local maxima in the scale-space of S_{Harris}.

Hessian The Hessian interest point is similar to the Harris detector. However, instead of a second-moment matrix $\mathcal{M}$, it is based on Hessian matrix $\mathcal{H}$.

$$\mathcal{H} = \begin{bmatrix} \mathbf{V}_{xx} & \mathbf{V}_{xy} & \mathbf{V}_{xz} \\ \mathbf{V}_{yx} & \mathbf{V}_{yy} & \mathbf{V}_{yz} \\ \mathbf{V}_{zx} & \mathbf{V}_{zy} & \mathbf{V}_{zz} \end{bmatrix}, \quad (3)$$

where $\mathbf{V}_{xy}$ denotes the second derivative of the volume at scale σ_s, along x and y axes: $\frac{\partial \mathbf{V}(\mathbf{x};\sigma_s)}{\partial x \partial y}$. The Hessian saliency is the determinant of the $\mathcal{H}$ normalized by scale σ_s. Similarly, the interest points are located at the 4D scale-space maxima of S_{Hessian}.

$$S_{\text{Hessian}} = \sigma_s^3 \det(\mathcal{H}) \quad (4)$$

SURF Speeded up robust features (SURF) is a feature extraction algorithm optimized for efficiency [1]. SURF is similar to Hessian interest point in terms of the feature model it is based on. Haar wavelets are leveraged to approximate derivatives of Gaussian functions, hence the computation of Hessian is accelerated. The first 3D version of SURF was introduced in by Willems *et al.* [24] for video classification. Recently, it was utilized in [11] for 3D shape object recognition tasks.

V-FAST Building on the success of FAST corner detector [20], Yu *et al.* [25] propose V-FAST for video recognition. Koelstra & Patras [12] also introduce a similar FAST-3D interest point. The V-FAST algorithm performs accelerated segment tests (AST) on three orthogonal Bresenham circles. Along each plane, 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 in equation 5.

$$AST_{xy}(n,t) \begin{cases} t & ||v_{nucleus} > \mathbf{c}_{xy} + t|| \geq n \\ t & ||v_{nucleus} < \mathbf{c}_{xy} - t|| \geq n \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

$$S^{xy}_{vfast} = \max(AST_{xy}(n,t)) \quad (6)$$

$$S_{\text{vfast}} = \sqrt{(S^{xy}_{vfast})^2 + (S^{xz}_{vfast})^2 + (S^{yz}_{vfast})^2} \quad (7)$$

where $\mathbf{c}_{xy}$ denotes the voxels on a xy-circle centred at $c_{nucleus}$. The overall saliency S_{vfast} is the Euclidean norm of saliency scores on the three planes in equation 7. Interest points are local maxima in S_{vfast} along both space and scale, with at least two non-zero values in $AST_{xy}(n,t)$, $AST_{xz}(n,t)$ or $AST_{yz}(n,t)$.

MSER Maximally stable extremal region (MSER) is a blob detection technique proposed by Matas *et al.* [16]. MSER is inherently advantageous for volumetric interest point detection, therefore MSER has already been applied for detecting volumetric salient regions [8, 19].

Subvoxel refinement We apply the subpixel refinement process of Lowe [15] to all the candidate detectors (except MSER). Interest points are localized at the subvoxel level by fitting a 4D quadratic function around the local scale-space maxima. More accurate maxima are determined by interpolating the quadratic function. On the contrary, MSER locates interest points by fitting an ellipsoid to the detected salient region [16].

3. Methodology

This section describes our framework for evaluating the volumetric interest points.

3.1. Test data

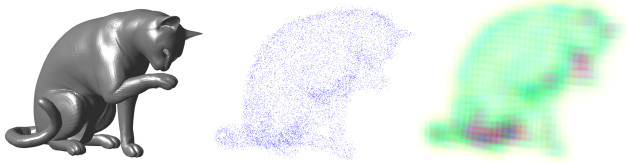


Figure 2: **Mesh to volume conversion**: *Left to right* – mesh, point cloud and voxel array.

We use three sources of data in our evaluation. Two of these are synthetic, as large sets of real, registered, 3D data are not commonly available. The first set, **Mesh**, contains 25 shapes (surface meshes) chosen from the *Princeton shape database* [22] and *TOSCA* [5] set. The selection contains a wide range of geometric features, from coarse structures to fine details. The meshes are converted to scalar volumetric data by first randomly sampling a number of points from a uniform distribution over the area of the mesh, and adding homogeneous Gaussian noise to these points (in 3D space). The point clouds are then *voxelized* to volumetric data using kernel density estimation with a Gaussian kernel $g(\cdot, \sigma_{KDE})$. The conversion process, illustrated in figure 2, allows sampling density and noise to be varied when generating the volumes. The second data set, **MRI**, is two synthetic MRI scans of a brain, generated from BrainWeb [6], again with known transformation between the two. The third data set, **Stereo**, is a pair of dense 3D point clouds of an object, captured using a multi-view stereo system[3]. The point clouds were then registered to the known mesh of the object using ICP [2]. The point clouds are voxelized as before.

3.2. Performance Measure

Whilst previous evaluations have focused on either localization accuracy or repeatability, we combine the two in a single score.

Localization accuracy A key component of the score is the distance metric used to measure the closeness of two interest points.

$$D(\mathbf{P}, \mathbf{Q}') = ||\mathbf{P} - \mathbf{Q}'||_2 \quad (8)$$

$$\mathbf{P} = [P_x,\ P_y,\ P_z,\ f \log P_s]^{\mathrm{T}} \quad (9)$$

where P_x, P_y and P_z are the spatial components and P_s is the scale component of an interest point's location. The log of scale is used to remove multiplicative bias across detectors, and since spatial location and scale are not fully commensurable, a parameter f is introduced to balance the importance of scale to the distance function.

In our evaluation the interest point $\mathbf{P}$ is in the coordinate frame of the volume, V_P, it was found in, whilst $\mathbf{Q}'$ is the point $\mathbf{Q}$ found in volume V_Q of a given pair of volumes, transformed into the coordinate frame of V_P. Our score is based on the distance of an interest point to the nearest transformed interest point, thus:

$$D(\mathbf{P}, \mathcal{Q}') = \min_{\mathbf{Q}'_j \in \mathcal{Q}'} D(\mathbf{P}, \mathbf{Q}'_j) \quad (10)$$

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

Repeatability Schmid *et al.* [21] defined repeatability as the ratio of correspondences to points:

$$R_{\text{ratio}}(\mathcal{P}, \mathcal{Q}', d) = \frac{\sum_{i=1}^{p} [D(\mathbf{P}_i, \mathcal{Q}') < d]}{\min(p, q)} \quad (11)$$

where $\mathcal{P} = \{\mathbf{P}_i\}_{i=1}^{p}$, the set of p interest points found in V_P, and d is a user provided distance threshold. The function $[expr]$ returns 1 if $expr$ is true, 0 otherwise.

Combined score Rosten *et al.* [20] computed the area under R_{ratio} as a function of the number of interest points (varied using a contrast threshold on the detector). We use the

[3] Source withheld for anonymity.

same idea, but computing R_{ratio} as a function of the distance threshold, d. The score therefore increases both if more points are matched, and also if matches are more accurate. We also compute a symmetric score, by averaging the scores with V_P and V_Q as the reference frames respectively. This score is given as

$$R_{\text{area}} = \frac{1}{2D} \int_0^D R_{\text{ratio}}(\mathcal{P}, \mathcal{Q}', d) + R_{\text{ratio}}(\mathcal{Q}, \mathcal{P}', d)\, \mathrm{d}d \tag{12}$$

4. Experiments

A comprehensive evaluation is performed on the volumetric interest point detectors in this section. The main objective is to investigate how the interest points perform under different variations. Our R_{area} score is advantageous over traditional repeatability reflects both repeatability and accuracy of interest points in one measurement.

4.1. Experiment set-up

We evaluate the robustness of the detectors under several variations. These include rotation, scale, sampling density and noise. The three datasets in section 3.1 are tested under different variations:

- **Mesh**: Sampling noise and density, scale, rotation.
- **MRI**: Rotation + translation.
- **Stereo**: Scale + rotation + translation.

In the proposed evaluation framework, comparisons are performed between the transform shapes with their corresponding reference shapes. Reference shapes are generated by voxelizing the input point clouds with a set of default parameters. Subsequently, transformed shapes are created from the reference shapes, with increasing magnitudes of one target variation (*e.g.* scale, noise) and others factors kept at default values. The default parameters for the reference shapes and the number of transformed shapes created are listed in table 1.

Some default parameters for reference shapes are defined in terms of L, which is equal to the largest dimension of the voxelized reference shapes. In this paper, the maximum value of L is set to 200 voxels.

4.2. Experiments on synthetic meshes

4.2.1 Sampling noise

Sampling noise and density are of crucial importance to shape-based interest point detection. However, existing shape acquisition techniques (*e.g.* multi-view stereo) often produce data with sampling noise. In this test, different levels of Gaussian white noise, with standard deviations from $0.25\%L$ to $3\%L$, are applied to the **Mesh** dataset. Its magnitude is measured in a percentage of L (*i.e.* longest dimension of the shape).

Parameter	Value
Default parameters for reference shapes	
Default point cloud size	50000 points
Default noise	$0.25\%L$
Default rotation	0°
Maximum L	200 voxels
Default σ_{KDE} in $g(\cdot, \sigma_{\text{KDE}})$	1.5 voxel
Distance threshold d	$3\%L$
Parameter f in equation 8	$\sqrt{8}$
Number of octaves in scale-space	4
Number of transformed shapes compared	
Sampling noise	13
Sampling density	17
Noise	21
Scale	21

Table 1: The reference parameters for the testing shapes.

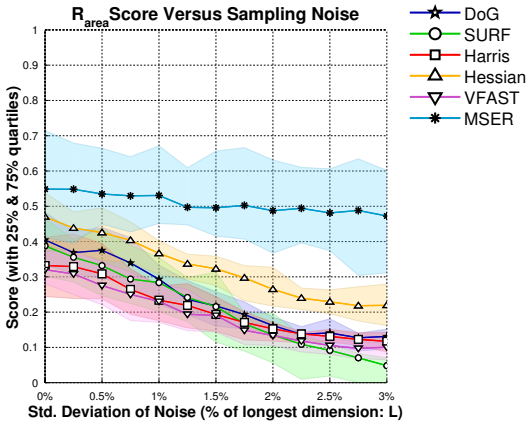


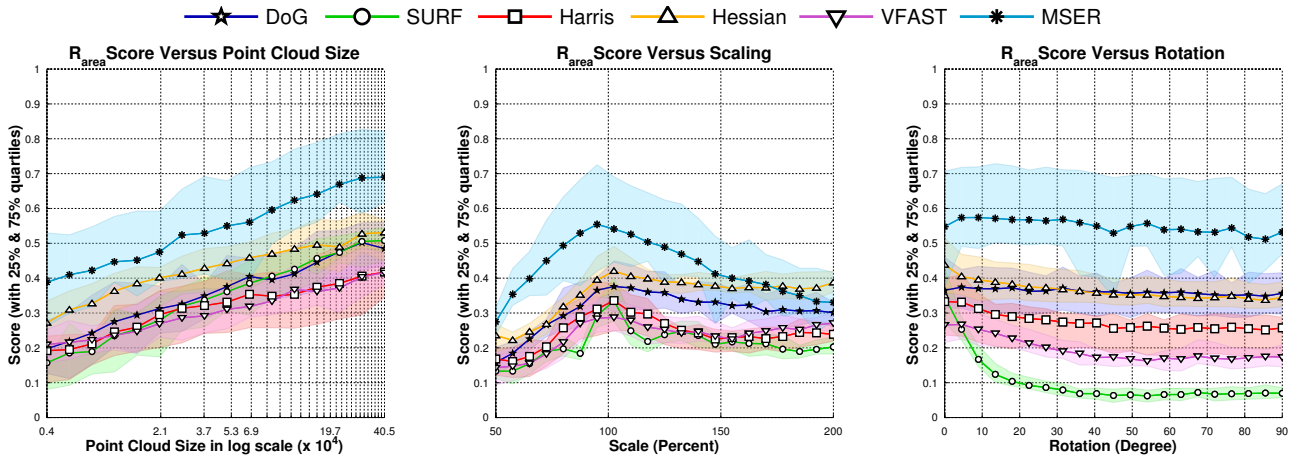
Figure 4: R_{area} Score vs Noise

The result is shown as figure 4. MSER outperforms other interest points by demonstrating high robustness. While other detectors cease working properly due to the noisy input, MSER still obtains a high R_{area} score. The Hessian-based detector shows a relatively stronger tolerance than detectors like SURF, Harris and V-FAST.

4.2.2 Sampling density

The R_{area} score of shapes with various sampling densities are measured. Point clouds are randomly sampled from the input meshes, with sizes ranging from $4K$ points to $405K$ points. Figure 3a presents the change of R_{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. Hessian and DoG produce satisfactory results, with high scores but smaller intra-dataset variance than that of MSER.

3DIMPVT
#13



(a) Sampling Density (b) Scale (c) Rotation

Figure 3: R_{area} scores of testing dataset under changing (a) *sampling density (point cloud size)*, (b) *scaling* and (c) *rotation*. Solid lines indicate the average, and the light colour bands show the 25% and 75% quartiles.

4.2.3 Rotation

In the experiment for rotation transformation, which evaluates susceptibility of the detectors to aliasing effects, eight rotation axes are generated randomly. Rotations of increasing magnitude are applied about these axes, the testing shapes are rotated up to π radians.

The effect of rotation is shown in figure 3c. Most detectors show excellent tolerance to rotation. Most volumetric interest points have inherited strong robustness to rotation from their texture-based counterparts. DoG and MSER performs slightly better than others, with almost unchanged average scores. SURF performs worse than other volumetric interest point because the use of box filters leads to extra quantization errors when the shapes are rotated.

4.2.4 Scale

Dimensions of voxelized input data are scaled from 50% to 200% of their original sizes. The values of R_{area} measured against scale changes are illustrated in figure 3b. DoG and Hessian 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 down-sampled volumes (scale $< 100\%$) than in up-sampled volumes (scale $> 100\%$). It is because information of fine features is lost when the input volume is down-sampled. 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 drop slowly at scale $> 100\%$, when some features become too large to be detected.

Avg.# Pts. Avg.# Corr.Pts. (*Corr. %*)	**Sampling Noise Level** **Low** (0.25L%)	 **Medium** (1.0L%)	 **High** (2L%)
DoG	170.1 30.9 (*18.1%*)	128.5 14.7 (*11.5%*)	95.2 6.1 (*6.4%*)
SURF	160.1 40.5 (*25.3%*)	73.4 15.0 (*20.4%*)	37.0 4.8 (*12.9%*)
Harris	233.3 45.8 (*19.6%*)	180.9 14.1 (*7.8%*)	150.5 5.8 (*3.8%*)
Hessian	**309.7** **62.7** (*20.3%*)	**289.8** **36.9** (*13.9%*)	**201.2** 16.1 (*9.1%*)
VFAST	154.1 22.8 (*14.8%*)	121.7 11.1 (*9.1%*)	111.4 4.0 (*3.6%*)
MSER	94.1 47.9 (**50.9**%)	80.1 36.1 (**45.1**%)	57.5 **24.4** (**41.8**%)

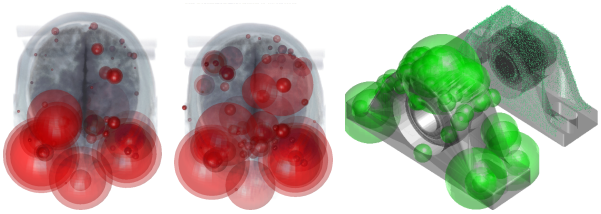
Table 2: *Top to bottom:* The average number of interest points detected, the average number of correspondences ($d \leq 1.5\%L$), percentage of correspondence points.

4.2.5 Number of interest point correspondences

Table 2 presents a quantitative statistics for the number of interest points and correspondences at three noise levels (0.25%, 1% and 2% of L). MSER detector has the best percentage of correspondences, yet it gives a smaller set of

interest points. By contrast, Hessian, SURF and Harris produce larger sets of interest points with good correspondence ratios. The displacement threshold used here ($d = 1.5\%L$) is much lower than the typical value ($d = 3\%L$), hence only highly accurate corresponding points are counted towards the ratios.

4.3. Experiments on MRI and stereo data



(a) MRI scan 1 (b) MRI scan 2 (c) DoG on Stereo data

Figure 5: (a) MRI scan 1 with MSER detected, (b) MRI scan 2 (aligned) with MSER, (c) Point clouds in the Stereo dataset, together with the DoG interest points detected.

In addition to synthetic mesh data, detectors are evaluated on other input datasets (the **MRI** dataset and the **Stereo** dataset) as mentioned in section . As a reference for cross validation, average scores obtained from the **Mesh** dataset are plotted against displacement threshold d in figure 6a.

MRI dataset This dataset contains two MRI scans of a human brain. Each MRI has the longest dimension L of 218 voxels. The first MRI scan is transformed (with 20° rotation and 20 voxels translation) and interpolated to generate the second MRI scan. Figure 5a and 5b shows the MSER interest points detected on 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.

R_{area} scores are measured from the dataset with varying d in figure 6b. The evaluation results obtained are comparable to that of synthetic mesh data — MSER, DoG and Hessian work slightly better in synthetic meshes, while Harris detector is good at detecting complicated internal structures in the MRI scans.

Stereo dataset A pair of point clouds, as shown in figure 5c, are acquired separately from a multi-view stereo system. Volumetric data with maximum length of 132 voxels are converted from these point. The R_{area} scores obtained from the **Stereo** dataset are shown in figure 6c. Although the order of detectors is similar to that of **MRI** and **Mesh** data, R_{area} scores of stereo data are much lower its counterpart. Particularly, the performances of Hessian and Harris drop rapidly, as the uneven sampling of data points affects the computation of volume derivatives. Generally, the decrease of performance is due to: *(1)* low sampling frequency and high noise in stereo data, *(2)* uneven surface points which are infeasible for some detection algorithms (*e.g.* MSER, Harris and Hessian) and *(3)* errors in homography obtained from ICP alignment.

4.4. Qualitative analysis: Blobs vs Corners

The volumetric interest points can be roughly classified into three categories: region-based blob detection (*MSER*), derivative-based blob detection (*DoG*, *Hessian* and *SURF*) and corner detection (*Harris*, *V-FAST*). The evaluation results imply that region-based blob detector works better than derivative-based detectors, and blob detectors are better than corner detectors.

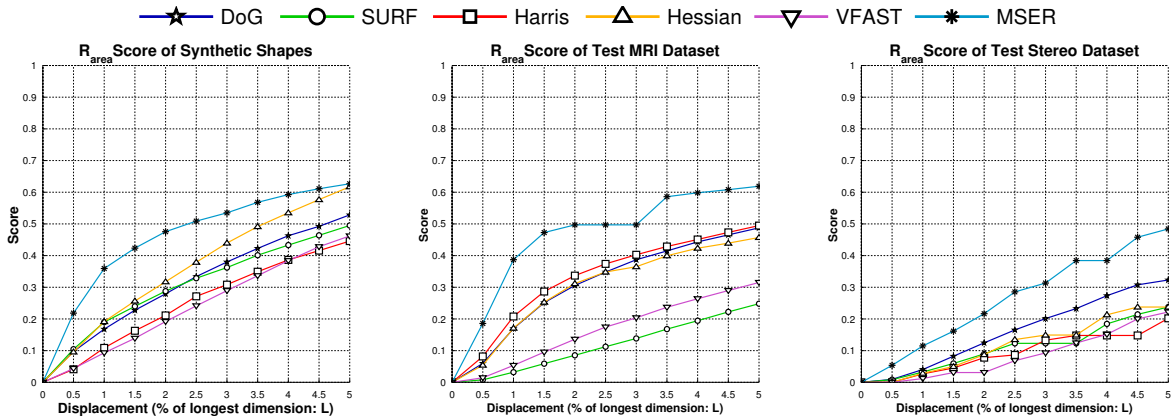
Region-based blob detection (*MSER*) works better in 3D volumes because 3D regions are more stable in volumetric data. Furthermore, 3D shapes are invariant to view point changes (affine tranformations), therefore blob detection in 3D shapes has become more stable than in texture (images) with view point changes.

Derivative-based detection (*DoG*, *Hessian* and *SURF*) also possesses the advantages of blob detection techniques. However, these approaches are more vulnerable to noise because of the derivative calculation involved.

Corner detection is relatively less robust to noise 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. However, they are still suitable for detecting features when blobs cannot be detected correctly at low sampling density (*c.f.* figure 6c).

5. Conclusion

In this paper, we compared the state of the art volumetric interest points. Combining both repeatability and accuracy, we proposed a novel metric to evaluate a feature detector. Three types of input data (meshes, MRI and stereo point clouds) are leveraged in the evaluation framework to give a complete picture of different detectors. Summarising the results with respect to the proposed R_{area} score, there does not exist any interest point that outperforms others in all aspects. MSER achieves the best overall performance. It is robust to both noise and rotation, but does not work well in unevenly sampled shapes. Taking the number of corresponding points into account, Hessian and DoG maintain a balanced performance between repeatability and number of correspondences. From the experiments, blob detectors (*e.g.* Hessian) performs better than corners detectors (*e.g.* Harris) in 3D shapes, and this phenomenon is analysed qualitatively. Interestingly, some parts of our results agree with the evaluation of texture-based detectors in [18]. Nevertheless, the choice of volumetric interest points can also be application dependent. Our work seeks to provide



(a) Repeatability Score: *Mesh* (b) Repeatability Score: *MRI* (c) Repeatability Score: *Stereo*

Figure 6: R_{area} scores versus displacement threshold d: *Left to Right* — **Mesh**, **MRI** and **Stereo** dataset.

guidance on selection of interest point algorithm for specific computer vision or machine learning tasks.

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 2008. 2, 3

[2] P. Besl and N. McKay. A method for registration of 3-D shapes. *Trans. on Pattern Analysis and Machine Intelligence*, 1992. 4

[3] K. Bowyer, C. Kranenburg, and S. Dougherty. Edge detector evaluation using empirical ROC curves. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 1999. 3

[4] P. Brand and R. Mohr. Accuracy in image measure. In S. F. El-Hakim, editor, *Videometrics III*, volume 2350, pages 218–228. SPIE, 1994. 2

[5] A. Bronstein, M. Bronstein, and R. Kimmel. *Numerical Geometry of Non-Rigid Shapes*. Springer, 1 edition, 2008. 4

[6] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, G. B. Pike, and A. C. Evans. BrainWeb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 1997. 4

[7] C. Coelho, A. Heller, J. L. Mundy, D. A. Forsyth, and A. Zisserman. An experimental evaluation of projective invariants. In J. L. Mundy and A. Zisserman, editors, *Geometric invariance in computer vision*, pages 87–104. MIT Press, Cambridge, MA, USA, 1992. 2

[8] M. Donoser and H. Bischof. 3D segmentation by maximally stable volumes (MSVs). In *Intl. Conf. on Pattern Recognition*, 2006. 2, 3, 4

[9] G. Flitton, T. Breckon, and N. Megherbi Bouallagu. Object recognition using 3D sift in complex CT volumes. In *British Machine Vision Conf.*, 2010. 2, 3

[10] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc., 4th Alvey Vision Conference*, 1988. 2, 3

[11] J. Knopp, M. Prasad, G. Willems, R. Timofte, and L. Van Gool. Hough transform and 3D SURF for robust three dimensional classification. In *European Conf. on Computer Vision*, 2010. 2, 3

[12] S. Koelstra and I. Patras. The FAST-3D spatio-temporal interest region detector. In *Proc., WIAMIS '09*, 2009. 2, 3

[13] I. Laptev. On space-time interest points. *Intl. Jour. of Computer Vision*, 2005. 2, 3

[14] T. Lindeberg. Feature detection with automatic scale selection. *Intl. Jour. of Computer Vision*, 1998. 2, 3

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Intl. Jour. of Computer Vision*, 2004. 2, 3, 4

[16] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 2004. 2, 4

[17] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *Intl. Jour. of Computer Vision*, 2004. 2, 3

[18] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *Intl. Jour. of Computer Vision*, 2005. 7

[19] H. Riemenschneider, M. Donoser, and H. Bischof. Bag of optical flow volumes for image sequence recognition. In *British Machine Vision Conf.*, 2009. 2, 3, 4

[20] E. Rosten, R. Porter, and T. Drummond. Faster and better: A machine learning approach to corner detection. *Trans. on Pattern Analysis and Machine Intelligence*, 2010. 2, 3, 4

[21] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *Intl. Jour. of Computer Vision*, 2000. 2, 3, 4

[22] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser. The Princeton shape benchmark. In *Shape Modeling International*, 2004. 4

[23] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: a survey. *Found. Trends. Comput. Graph. Vis.*, pages 177–280, 2008. 2

[24] G. Willems, T. Tuytelaars, and L. Van Gool. An efficient dense and scale-invariant spatio-temporal interest point detector. In *European Conf. on Computer Vision*, 2008. 2, 3

[25] T. H. Yu, T. K. Kim, and R. Cipolla. Real-time action recognition by spatiotemporal semantic and structural forest. In *British Machine Vision Conf.*, 2010. 2, 3